

RESEARCH

Open Access



# The influence of the gut microbiome on BCG-induced trained immunity

Martin Stražar<sup>1†</sup>, Vera P. Mourits<sup>2†</sup>, Valerie A. C. M. Koeken<sup>2,3</sup>, L. Charlotte J. de Bree<sup>2,4,5</sup>, Simone J. C. F. M. Moorlag<sup>2</sup>, Leo A. B. Joosten<sup>2,6</sup>, Reinout van Crevel<sup>2</sup>, Hera Vlamakis<sup>1</sup>, Mihai G. Netea<sup>2,7\*</sup> and Ramnik J. Xavier<sup>1,8,9\*</sup> 

\* Correspondence: [mihai.netea@radboudumc.nl](mailto:mihai.netea@radboudumc.nl); [xavier@molbio.mgh.harvard.edu](mailto:xavier@molbio.mgh.harvard.edu)

<sup>†</sup>Martin Stražar and Vera P. Mourits contributed equally to this work.

<sup>2</sup>Department of Internal Medicine, Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA

Full list of author information is available at the end of the article

## Abstract

**Background:** The bacillus Calmette-Guérin (BCG) vaccine protects against tuberculosis and heterologous infections but elicits high inter-individual variation in specific and nonspecific, or trained, immune responses. While the gut microbiome is increasingly recognized as an important modulator of vaccine responses and immunity in general, its potential role in BCG-induced protection is largely unknown.

**Results:** Stool and blood were collected from 321 healthy adults before BCG vaccination, followed by blood sampling after 2 weeks and 3 months. Metagenomics based on de novo genome assembly reveals 43 immunomodulatory taxa. The nonspecific, trained immune response is detected by altered production of cytokines IL-6, IL-1 $\beta$ , and TNF- $\alpha$  upon ex vivo blood restimulation with *Staphylococcus aureus* and negatively correlates with abundance of *Roseburia*. The specific response, measured by IFN- $\gamma$  production upon *Mycobacterium tuberculosis* stimulation, is associated positively with *Ruminococcus* and *Eggerthella lenta*. The identified immunomodulatory taxa also have the strongest effects on circulating metabolites, with *Roseburia* affecting phenylalanine metabolism. This is corroborated by abundances of relevant enzymes, suggesting alternate phenylalanine metabolism modules are activated in a *Roseburia* species-dependent manner.

**Conclusions:** Variability in cytokine production after BCG vaccination is associated with the abundance of microbial genomes, which in turn affect or produce metabolites in circulation. *Roseburia* is found to alter both trained immune responses and phenylalanine metabolism, revealing microbes and microbial products that may alter BCG-induced immunity. Together, our findings contribute to the understanding of specific and trained immune responses after BCG vaccination.

**Keywords:** Nonspecific immune responses, Trained immunity, BCG, Microbiome, Phenylalanine metabolism, *Roseburia*, *Eggerthella lenta*

## Background

The bacillus Calmette-Guérin (BCG) vaccine consists of the live attenuated microorganism *Mycobacterium bovis* and partially protects against tuberculosis caused by *Mycobacterium tuberculosis* in humans. Epidemiological and experimental data provide evidence that BCG vaccination also protects against heterologous infections [1–3]. This



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

nonspecific protection is at least partially explained by trained immunity, the process in which certain infections and vaccinations result in epigenetic and metabolic rewiring of innate immune cells that increases responsiveness upon subsequent restimulation with unrelated pathogens [4, 5]. In monocytes and natural killer cells, trained immunity induced by BCG vaccination leads to increased production of pro-inflammatory cytokines upon restimulation with nonspecific pathogens, such as *Staphylococcus aureus* and *Candida albicans* [6, 7]. These effects are enabled by modulation of hematopoietic stem and progenitor cells in the bone marrow [8, 9] and last for at least 1 year [10]. However, BCG is not effective in all individuals, and host and environmental factors underlying inter-individual variability of specific and trained immune responses to BCG vaccination are currently incompletely understood [6, 11].

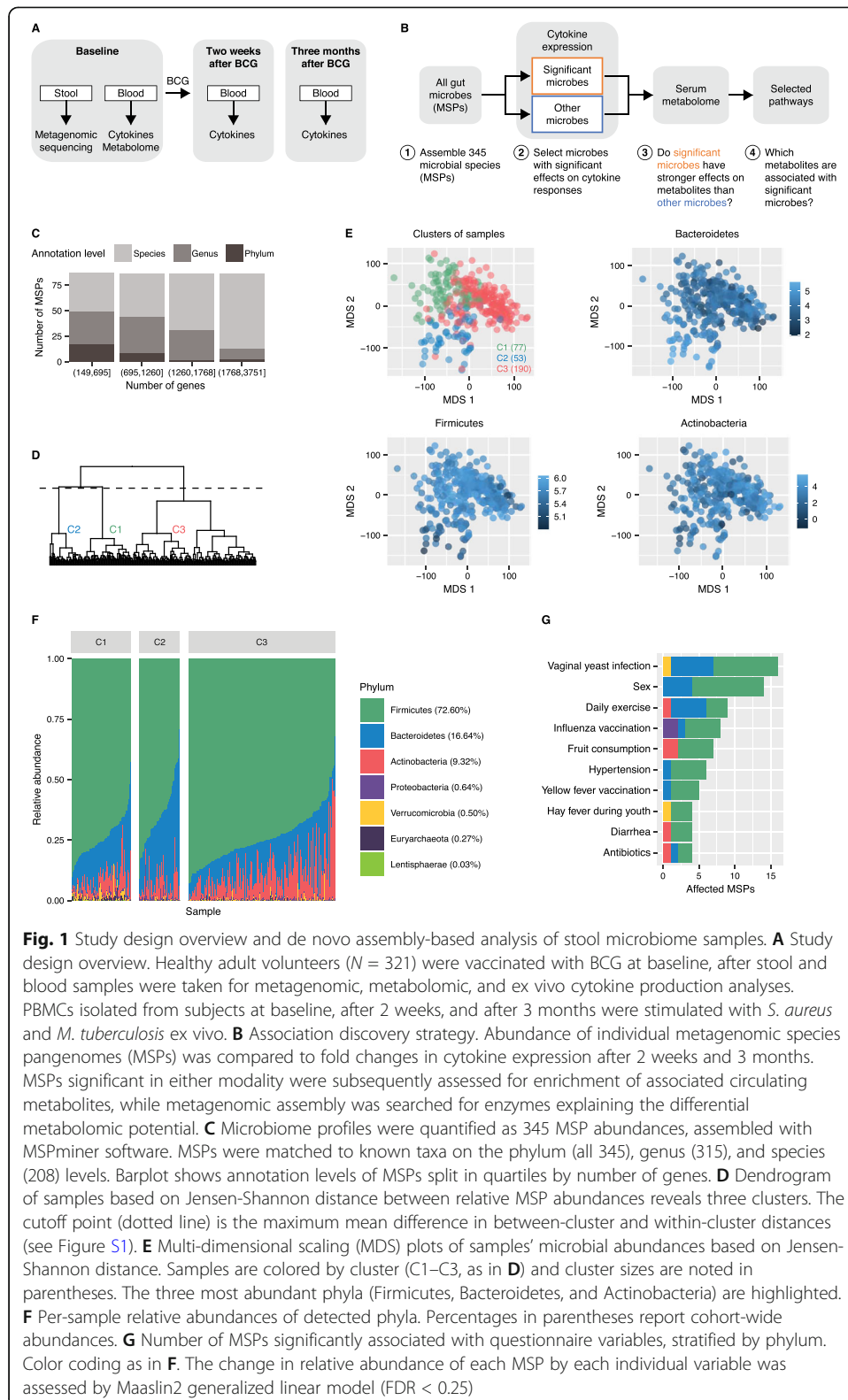
The microbiome plays an essential role in regulating immune responses [12, 13], and numerous mechanisms have been identified through which the microbiome affects immune cell functions, both locally and systemically. For example, microbiome-produced metabolites such as short-chain fatty acids (SCFAs) directly affect myeloid cells [14], while the attachment of segmented filamentous bacteria to the intestinal epithelium improves the function of T helper 17 cells [13, 15]. Moreover, while the microbiome has been proposed to influence the efficacy of vaccines, only a few studies have investigated this [16, 17]. In BCG vaccine responses, two studies observed a role for gut microbiota: mice treated with antibiotics early in life showed impaired antigen-specific IgG responses and T cell generation upon BCG vaccination as well as an increased amount of colony forming units in the lungs upon subsequent *M. tuberculosis* infection [18, 19]. Furthermore, a positive association between *Actinobacteria* and BCG vaccine responsiveness in terms of T cell proliferation was observed in samples from 48 infants at 15 weeks of age [20].

Studies investigating the role of the gut microbiome on BCG-induced trained immune responses, however, have not been performed to date. Here, we investigate the effect of the gut microbiota on the induction of specific and nonspecific immune memory responses in a cohort of 321 healthy, BCG-vaccinated individuals. Through associations with the increase in cytokine production after vaccination, we identify potential immunomodulatory species that are further supported and filtered by associations with plasma metabolites. This approach yielded a *Roseburia* genome that showed consistent associations with trained immune responses at two post-vaccination time points and enzyme-mediated effects on phenylalanine metabolites.

## Results

### Microbial profile of study cohort

To study the influence of the microbiome on BCG-induced trained immunity, we vaccinated 321 healthy individuals of Western European background with BCG [18–75 years old, median age 23 years, 57% female, 83% body mass index (BMI) between 18.5 and 25]. Stool and blood samples were collected at baseline, prior to participants receiving the BCG vaccine. Blood was subsequently collected after 2 weeks and 3 months, as depicted in Fig. 1A. DNA was isolated from the 321 stool samples (“Methods”), amplified, and sequenced using whole genome sequencing. Peripheral blood mononuclear cells (PBMCs) isolated from participants at each time point were subsequently



stimulated ex vivo with *S. aureus* or *M. tuberculosis*; IL-6, IL-1 $\beta$ , and TNF- $\alpha$  were measured after 24 h and IFN- $\gamma$  was measured after 7 days. Additionally, metabolites were analyzed in plasma from peripheral blood at baseline (Koeken et al. in preparation).

Our approach consisted of (1) assembling all microbial species, (2) testing the microbes for modulatory effects on cytokine responses, (3) assessing whether immunomodulatory microbes have stronger effects on circulating metabolites compared to the other microbes, and (4) identifying which metabolites are associated with immunomodulatory microbial species (Fig. 1B).

Reference-based profiling with MetaPhlAn 2.2 detected 202 microbial species, while de novo genome assembly using MSPminer resulted in more than 1.7 million microbial genes. Three hundred and forty-five core metagenomic species pangenomes (MSPs) matched to a phylum, of which 315 (91.3%) and 208 (60.3%) matched to the genus and species level, respectively, with more genes leading to a more precise annotation (Fig. 1C). De novo assembly was subsequently used in all downstream analyses except where explicitly mentioned; besides retrieving all 202 initial species, this method yielded a 1.7-fold increase in the size of the profiled metagenomic space. We hereafter use MSP to refer to a taxonomic unit as a gene set matched to a phylum, genus, or species.

Community-level profiling using core MSP abundances yielded three distinct clusters of samples (Fig. 1D; Additional file 1: Fig. S1A). Average cluster compositions and sample positioning in multi-dimensional scaling plots reflect abundances of the three most common phyla (Fig. 1E). Firmicutes prevailed in cluster 1, with an average relative abundance of 72.6%, followed by Bacteroidetes (16.64%, cluster 2) and Actinobacteria (9.32%, cluster 3) (Fig. 1F). Cluster 1 also showed increased relative abundances of Verrucomicrobia and Euryarchaeota (Additional file 1: Fig. S1B-E). The overall composition reflected that of developed, Western-type societies that tend to be dominated by Firmicutes, while the presence of Bacteroidetes often reflects diet-driven variation [21]. This was corroborated by the species-level profiles being indistinguishable from the independent Dutch 500FG cohort of 473 healthy participants (Additional file 1: Fig. S1F-G) [12].

All the participants in our study were asked to complete a questionnaire that included general health-related questions. Eighteen out of 185 variables associated with the three microbiome clusters ( $P < 0.05$ , unadjusted Cauchy test), including diet (fruit intake), sex, smoking status, vaccination history, and gastrointestinal status-related variables (Additional file 1: Fig. S2; Additional file 2: Table S1).

Shifts in the relative abundance of individual species correlated with 25 variables (MaAsLin2 generalized linear model,  $P_{\text{FDR}} < 0.25$ ), of which 10 correlated with more species (Fig. 1G; Additional file 2: Table S1). In particular, vaginal yeast infection, sex, and daily exercise associated with most microbial species (Fig. 1G). Other species-specific variables that were missed by cluster associations included hypertension, hay fever during youth, and antibiotic use.

### **Microbiome species are associated with the capacity to modulate cytokine production after BCG vaccination**

Cytokine production in the supernatant of stimulated PBMCs was measured to assess the trained immune response (IL-6, IL-1 $\beta$ , TNF- $\alpha$  after 24-h stimulation) and the specific immune response (IFN- $\gamma$  after 7-day stimulation). IL-1 $\beta$  and TNF- $\alpha$  production upon *S. aureus* stimulation and IFN- $\gamma$  production upon *M. tuberculosis* stimulation gradually increased 2 weeks and 3 months after BCG vaccination compared to baseline



(Fig. 2A; Additional file 1: Fig. S3A). Trained immunity (indicated by increases in TNF- $\alpha$ , IL-1 $\beta$ , IL-6 production) as well as specific (*M. tuberculosis*-induced IFN- $\gamma$ ) and heterologous lymphocyte-derived (*S. aureus*-induced IFN- $\gamma$ ) cytokine responses correlated with each other (Fig. 2B, C). The microbiome clusters did not show a significant divergence in any of the cytokine responses measured upon ex vivo stimulation of PBMCs (maximum  $P = 0.16$  for IFN- $\gamma$  with *S. aureus*, Additional file 1: Fig. S3B), prompting the search for a potential individual taxon or functional immunomodulatory factors.

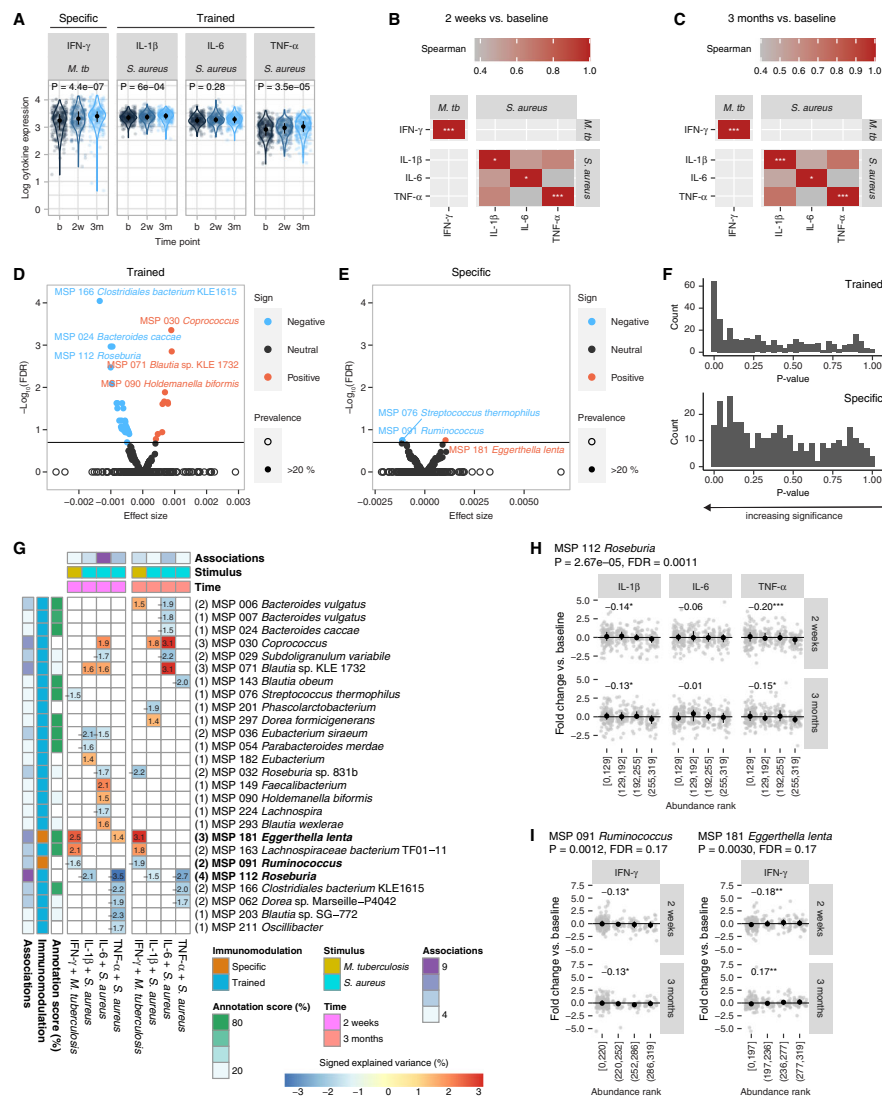
The associations between microbes and both immune system modalities were assessed with two linear models, adjusted for age and sex, testing for effects on trained immunity and specific T cell responses (“Methods”). Trained immunity was associated with 40 significant species (prevalence > 20%, FDR < 0.2, Fig. 2D; Additional file 3: Table S2), while specific immunity was associated with three species: *Ruminococcus* (MSP 091), *Eggerthella lenta* (MSP 181), and *Streptococcus thermophilus* (MSP 076) (Fig. 2E; Additional file 3: Table S2). At FDR < 0.1, 27 species were associated with trained immunity. Because the model for the trained immunity response combined more biological markers (three cytokines) than that for the specific response (one cytokine), it retrieved more species at the same significance threshold. Nevertheless, both modalities elicited a non-uniform distribution of p-values (Fig. 2F), supporting an influence of the microbiome on BCG vaccine responses.

We next examined the correlation between species relative abundance and individual cytokine production at both time points after vaccination (Fig. 2G). A *Roseburia* genome (MSP 112), only identifiable on the genus level, was negatively correlated with the increase in IL-1 $\beta$  and TNF- $\alpha$  production at both time points ( $P = 2.67e-05$ , FDR = 0.0011, Fig. 2H). In the specific response model, *Ruminococcus* (MSP 091,  $P = 0.0012$ , FDR = 0.17) and *E. lenta* (MSP 181,  $P = 0.0030$ , FDR = 0.17) showed consistent negative and positive effects, respectively, on IFN- $\gamma$  at both time points (Fig. 2I). The specific response model revealed the most associations between a single cytokine and microbial species, although for all the cytokine–stimuli combinations, the number of unadjusted associations ranged between 10 and 23 microbes (Additional file 4: Table S3).

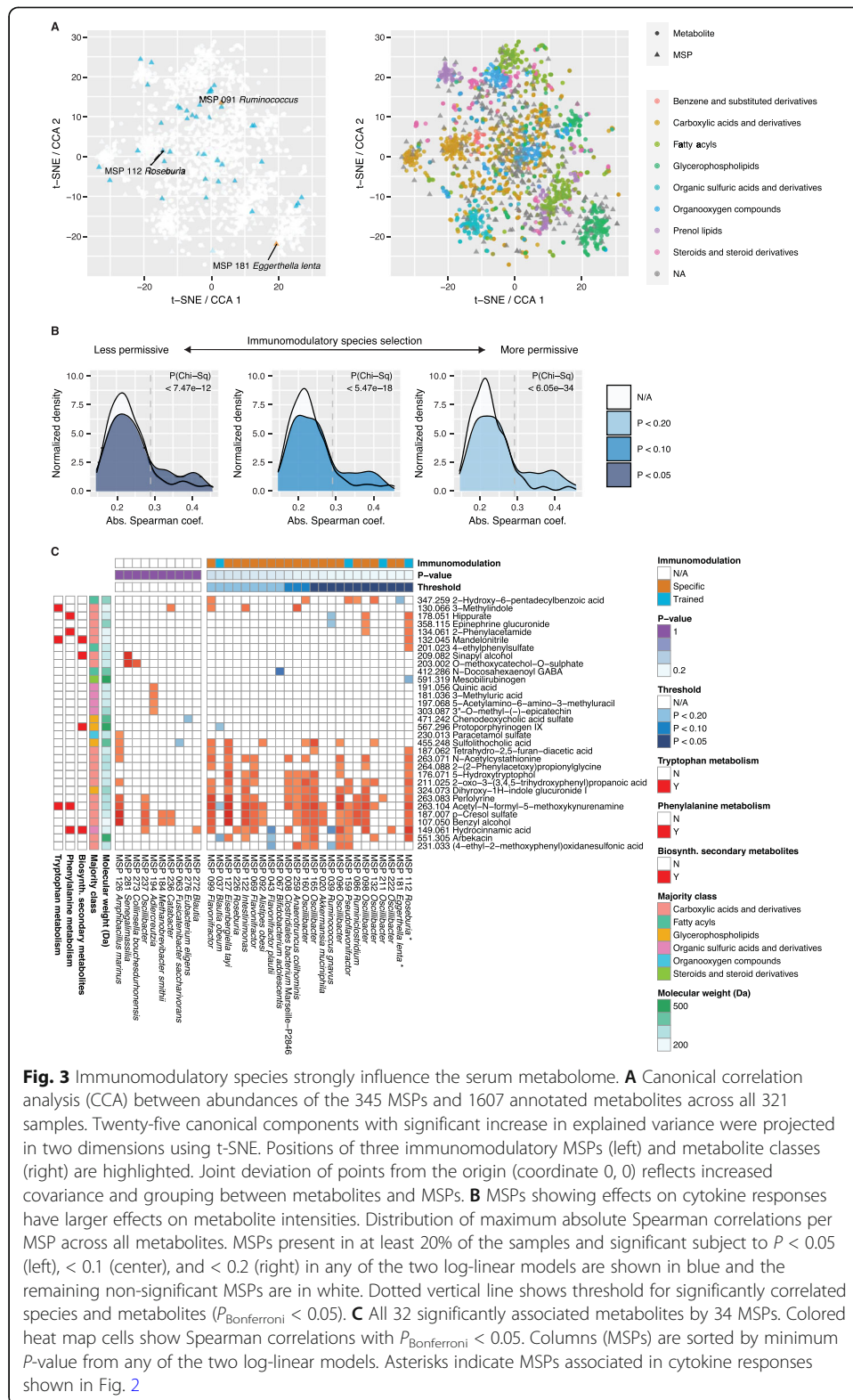
### Immunomodulatory microbes strongly influence circulating metabolites

Detectable changes in cytokine production yielded potential immunomodulatory species, whose exact numbers varied with selected statistical thresholds. To corroborate cytokine associations and narrow down the potential causal mechanisms, we compared the effects of immunomodulatory microbes on the plasma metabolome, measured as previously described [59]. A total of 1607 mass/charge (m/z) peaks were matched to known compound identities and clustered into 20 groups, enriched for eight major molecular classes (Additional file 1: Fig. S4A-C, “Methods”).

Canonical correlation analysis confirmed a strong microbiome effect on compound intensities, as 15 of the top 25 canonical components (CCs) showed significant explained variance (Additional file 1: Fig. S4D-E). A t-SNE projection of the top 25 CCs showed strong grouping by molecular class, with evident co-occurrence of microbes in clusters harboring carboxylic acids and derivatives, glycerophospholipids and organic sulfuric acids (Fig. 3A). The three immunomodulatory species *Roseburia* (MSP 112),



**Fig. 2** Microbial species abundance changes with differences in trained and stimulus-specific immune responses. **A** Ex vivo cytokine expression upon stimulation with *M. tuberculosis* (*M. tb*) or *S. aureus* at baseline (b) and at 2 weeks (2w) and 3 months (3m) after BCG vaccination. *P*-values were determined by paired Kruskal-Wallis tests. Significant correlations between changes in cytokine expression at 2 weeks (**B**) and 3 months (**C**), measured as fold change against baseline levels (Spearman correlation,  $P < 0.05$ ). Diagonal cells show significance of changes for each individual cytokine versus baseline (*T*-test,  $*P < 0.05$ ,  $***P < 0.001$ ). **D** Significant MSPs associated with trained immune responses subject to  $FDR < 0.2$  and prevalence  $> 20\%$ . The effect of each MSP was estimated by a linear model of the fold change in TNF- $\alpha$ , IL-1 $\beta$ , and IL-6 responses versus the baseline, adjusted for age and sex (“Methods”). **E** Significant MSPs associated with specific response (IFN- $\gamma$  stimulated by *M. tuberculosis*) and subject to  $FDR < 0.2$  and prevalence  $> 20\%$ , using a model of the same form as in **D**. **F** Distribution of unadjusted *P*-values for each MSP in the trained and specific response models. Skew from uniform distribution suggests more pronounced effects of MSPs on cytokine expression fold change. **G** Individual effects on cytokine fold changes by each MSP and time point, quantified as explained variance multiplied by effect sign. Number of associations per MSP and confidence in species-level annotation are reported on the left. **H** MSP 112 (*Roseburia*) abundance decreases two of the three trained immunity phenotypes in both subsequent time points. Relative abundance of the MSP was converted to ranks, and ranks were sorted into bins by equal-frequency binning. *P*-values in titles refer to the trained immunity linear model. Coefficients in panels are Spearman correlations ( $*P < 0.1$ ,  $***P < 0.001$ ). **I** MSP 091 (*Ruminococcus*) and MSP 181 (*Eggerthella lenta*) ranked abundance versus specific response phenotype. Relative abundance of the MSP was converted to ranks, and ranks were sorted into bins by equal-frequency binning. *P*-values in titles refer to the specific response linear model. Coefficients in panels are Spearman correlations ( $*P < 0.1$ ,  $**P < 0.01$ )



**Fig. 3** Immunomodulatory species strongly influence the serum metabolome. **A** Canonical correlation analysis (CCA) between abundances of the 345 MSPs and 1607 annotated metabolites across all 321 samples. Twenty-five canonical components with significant increase in explained variance were projected in two dimensions using t-SNE. Positions of three immunomodulatory MSPs (left) and metabolite classes (right) are highlighted. Joint deviation of points from the origin (coordinate 0, 0) reflects increased covariance and grouping between metabolites and MSPs. **B** MSPs showing effects on cytokine responses have larger effects on metabolite intensities. Distribution of maximum absolute Spearman correlations per MSP across all metabolites. MSPs present in at least 20% of the samples and significant subject to  $P < 0.05$  (left),  $< 0.1$  (center), and  $< 0.2$  (right) in any of the two log-linear models are shown in blue and the remaining non-significant MSPs are in white. Dotted vertical line shows threshold for significantly correlated species and metabolites ( $P_{Bonferroni} < 0.05$ ). **C** All 32 significantly associated metabolites by 34 MSPs. Colored heat map cells show Spearman correlations with  $P_{Bonferroni} < 0.05$ . Columns (MSPs) are sorted by minimum  $P$ -value from any of the two log-linear models. Asterisks indicate MSPs associated in cytokine responses shown in Fig. 2

*Ruminococcus* (MSP 091), and *E. lenta* (MSP 181) strongly influenced the composition of the metabolite space, as seen by their polarized placement in the joint projection (Fig. 3A).

Next, we investigated whether immunomodulatory microbes have stronger effects on individual metabolites compared to the other microbes. All pairwise Spearman correlations yielded 198 significant associations ( $P_{\text{Bonferroni}} < 0.05$ ) between 34 MSPs and 32 metabolite intensities (Additional file 5: Table S4). While most associations were positive (183/198), the resulting increase in metabolite intensities might result both from microbe-mediated anabolism or compound breakdown. To test whether species with large effects in either of the two linear models are more likely to elicit significant effects on the metabolome, we used three different thresholds that selected at least 40 species and computed the largest absolute effect on any one metabolite. All three settings yielded an enrichment of significant effects among immunomodulatory microbial species (maximum  $P < 10^{-11}$ , Chi-square test, Fig. 3B). This confirmed that microbial species with effects on cytokine production also have strong effects on circulating metabolites.

Of the 198 significant metabolite associations, 18 were associated with the presence of *Roseburia* (MSP 112), the most of any MSP, and one with *E. lenta* (MSP 181) (Fig. 3C). Due to uncertainty in annotation and pathway-mediated dependence on metabolite abundance, we attempted to identify general mechanisms by grouping the metabolites using KEGG pathways. Phenylalanine metabolism was influenced most strongly by *Roseburia* (MSP 112), with four metabolite associations in this pathway. *Roseburia* (MSP 112) additionally influenced three metabolites each in tryptophan metabolism and biosynthesis of secondary metabolites (Fig. 3C).

Untargeted plasma metabolomics detected two SCFAs, acetate and propionate, which are microbial products of dietary fiber metabolism and candidate mediators of host-microbiome signaling [14]. Both molecules showed significant positive associations with trained immunity but not with the specific response (Additional file 1: Fig. S5). While immunomodulatory microbes were not significantly correlated with either acetate or propionate, *E. lenta* had the highest positive correlation with propionate.

Butyrate, an SCFA previously shown to decrease IL-6 production [22], was not detected in circulation, possibly due to its absorption in the gut [23]. Nevertheless, butyrate kinase, which catalyzes the last step of butyrate biosynthesis [24], was detected in a single *Roseburia* (MSP 209) and five *Coprococcus* species, where *Coprococcus* (MSP 030) had the strongest positive association with trained immunity, in particular an increase in IL-6, after vaccination (Additional file 1: Fig. S6A-C). Conversely, *Coprococcus comes* (MSP 213) correlated with a decreased trained immunity response, manifested in reduced TNF- $\alpha$  expression after 3 months (Additional file 1: Fig. S6D). The absence of butyrate kinase in immunomodulatory *Roseburia* (MSP 112) may stem from gaps in assembled genomes or allow for metabolic modalities other than SCFAs in mediating BCG-induced immunity.

Combined evidence from the (i) modeled effects of microbes on cytokine production, (ii) associations between species relative abundances and the production of two trained immunity-related cytokines at both time points after vaccination, and (iii) effects of the immunomodulatory microbiome species on the plasma metabolome yielded the prediction that *Roseburia*-mediated phenylalanine metabolism influences trained immunity.

### ***Roseburia* encodes phenylalanine pathway enzymes that are prevalent and abundant**

The effects on phenylalanine metabolites prompted us to investigate the *Roseburia* genomes and confirm the presence of relevant enzymes. We compared the differential

enzymatic and metabolic potential of *Roseburia* to phylogenetically and biochemically related taxa by a deeper profiling of the stool metagenomes. All 345 identified MSPs were grouped by phylogenetic divergence using core genes (PhyloPhlAn 3.0, Additional file 1: Fig. S7), revealing a clade harboring most *Roseburia* species (Fig. 4A). The phylogenetically closest Firmicutes included five *Ruminococcus*, three *Coprococcus* and two *Eubacterium rectale* MSPs. While *E. rectale* is related to *Roseburia* in terms of butyrate production, we also included a biochemically distinct but abundant Actinobacteria *Bifidobacterium* (MSP 111). The assembled panel consisted of 59 MSPs derived from 85,208 assembled genes of *Bifidobacterium* (7 MSPs), *Roseburia* (13 MSPs), *Ruminococcus* (17 MSPs), *Eubacterium* (15 MSPs), and *Coprococcus* (7 MSPs), with each genus having more than 95% prevalence in the cohort (Fig. 4B).

A total of 330 genes mapped to 20 out of 77 enzymes in the phenylalanine pathway (KEGG ko00360). *Roseburia* contributed the most genes (92, 45.1% average prevalence; Fig. 4C; Additional file 6: Table S5), whereas other genera in the panel contributed fewer [*Ruminococcus* (77, 19.2%), *Eubacterium* (66, 36.3%), *Coprococcus* (50, 35.5%) and *Bifidobacterium* (45, 46.1%)]. *Roseburia* also encoded the most phenylalanine enzymes (13 out of 20). Of these, *hisC*, *mhpE*, *yhDR*, *paaH*, *paaF*, *paaI*, and *enr* were significantly more abundant in *Roseburia* than in all of the other four genera (Fig. 4D).

A total of 15 m/z peaks matched to compounds from the phenylalanine pathway (KEGG ko00360), potentially spanning up to 19 compounds due to ambiguous matches. Species from the *Roseburia* genus significantly correlated with 11 of the 15 compounds in the pathway, of which MSP 112 contributed to six of these compounds (Spearman correlation,  $P < 0.05$ , Fig. 4E). The genera *Ruminococcus*, *Coprococcus*, *Bifidobacterium*, and *Eubacterium* each showed fewer associations (Fig. 4E). Taken together, *Roseburia*'s increased influence on the phenylalanine metabolism pathway compared to closely related species is confirmed through increased enzymatic potential and metabolite intensity.

### Comparison of *Roseburia* genomes reveals alternate routes in phenylalanine metabolism

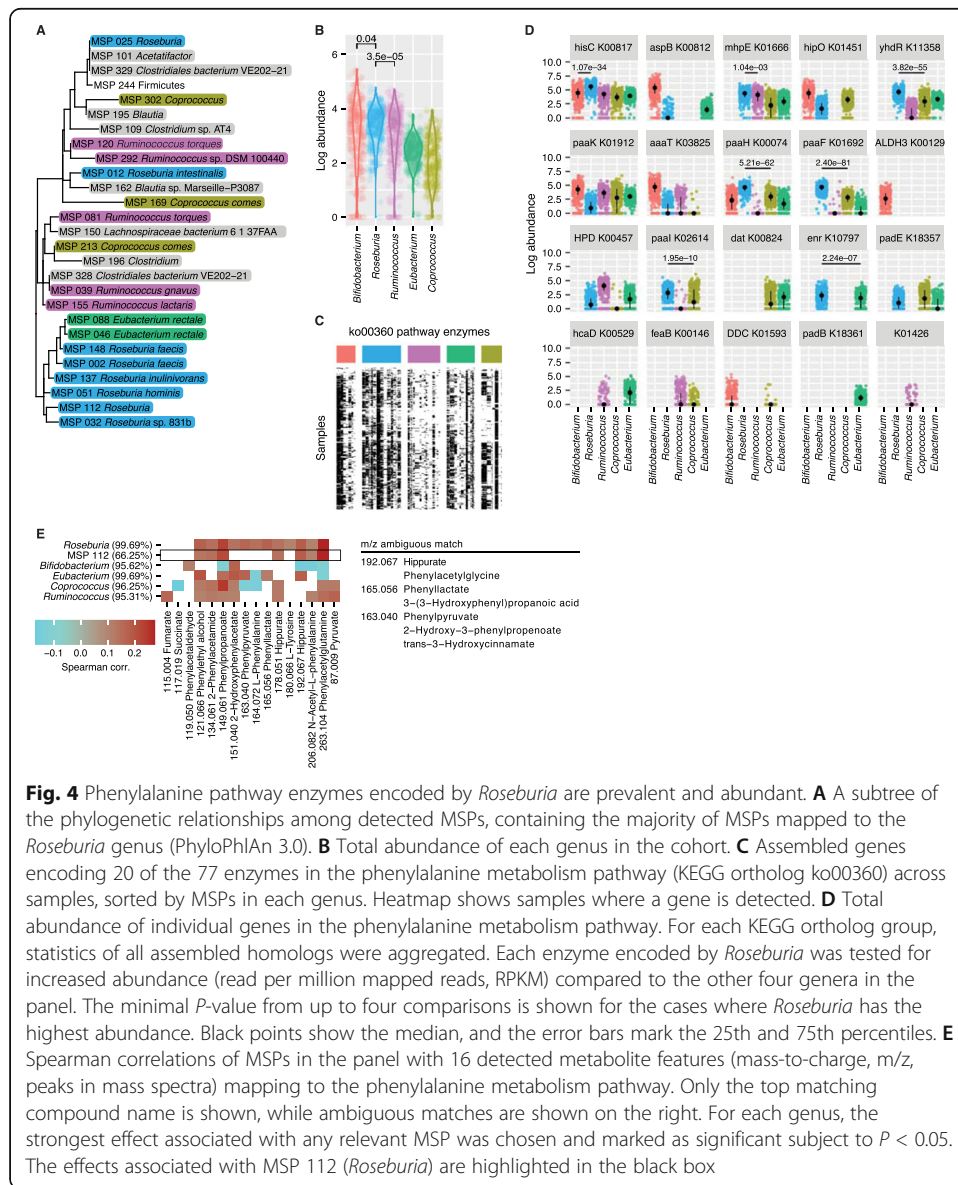
Finally, we more closely investigated the 13 phenylalanine metabolism enzymes found within the *Roseburia* genus and their differential effects on circulating metabolites. Five of the enzymes (*hisC*, *mhpE*, *paaH*, *yhDR*, and *paaF*) were common to over half of the MSPs, while the remaining eight were more species-specific (Fig. 5A). The co-occurrence of *aaaT* (L-phenylalanine N-acetyltransferase) and *paaI* (acyl-CoA thioesterase) was specific to the immunomodulatory MSP 112, as they were not found together in any other *Roseburia* MSP. These enzymes act in the path from phenylalanine to N-acetyl-L-phenylalanine and phenylacetylglutamine, whose plasma concentrations were associated with increased *Roseburia* (MSP 112) abundance (Fig. 4E).

Correlations between individual enzyme abundances and compound intensities revealed two distinct pathway modalities. First, an increase in L-tyrosine, L-phenylalanine, and phenylpyruvate (m/z ambiguous matches: 2-hydroxy-3-phenylpropenoate, trans-3-hydroxycinnamate) was accompanied by the five commonly shared enzymes (*hisC*, *mhpE*, *paaH*, *yhDR*, and *paaF*) as well as *hipO* (grouped by blue bars in Fig. 5B). Second, *HPD*, *paaK*, *aaaT*, and *padE* were associated with an increase in phenylpropanoate, hippurate (m/z ambiguous match: phenylacetylglutamine), and



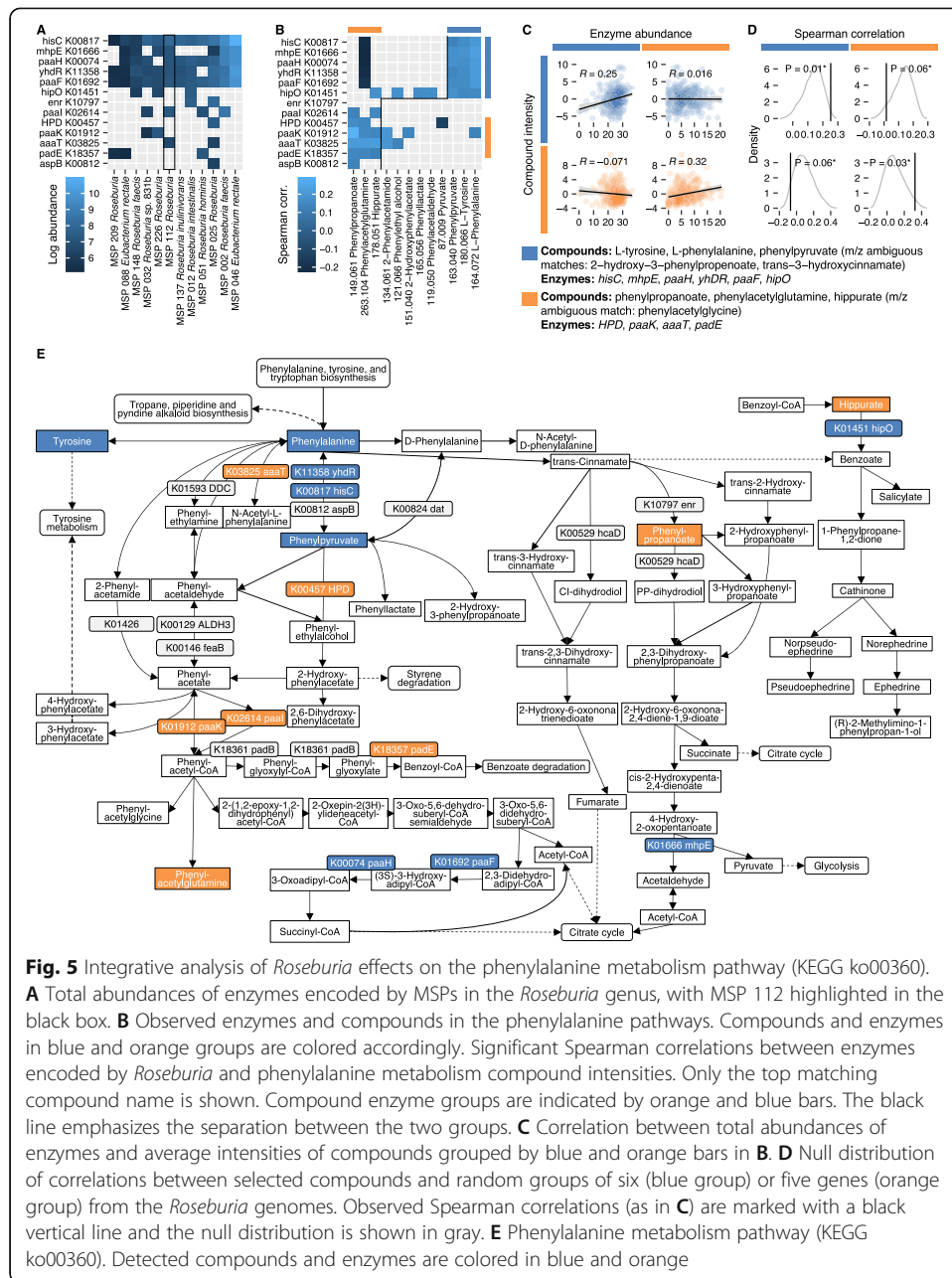
phenylacetylglutamine (grouped by orange bars in Fig. 5B). The latter compounds also strongly decreased with the presence of the five shared enzymes, suggesting an opposing mechanism between the two groups of compounds and enzymes.

We assessed the significance of the observed correlations to verify that the differences in compound intensities in each group were indeed more likely due to the associated



enzymes than any other genes. The Spearman correlation between sums of enzyme abundances and compound intensities within each group revealed even stronger effects than any individual compound–enzyme association (Fig. 5C, top-left and bottom-right panels). Conversely, correlations between groups showed weak to negative effects (Fig. 5C, top-right and bottom-left panels). Finally, the observed correlations were compared to a null distribution of correlations from 10,000 random draws of six (blue group) or





**Fig. 5** Integrative analysis of *Roseburia* effects on the phenylalanine metabolism pathway (KEGG ko00360). **A** Total abundances of enzymes encoded by MSPs in the *Roseburia* genus, with MSP 112 highlighted in the black box. **B** Observed enzymes and compounds in the phenylalanine pathways. Compounds and enzymes in blue and orange groups are colored accordingly. Significant Spearman correlations between enzymes encoded by *Roseburia* and phenylalanine metabolism compound intensities. Only the top matching compound name is shown. Compound enzyme groups are indicated by orange and blue bars. The black line emphasizes the separation between the two groups. **C** Correlation between total abundances of enzymes and average intensities of compounds grouped by blue and orange bars in **B**. **D** Null distribution of correlations between selected compounds and random groups of six (blue group) or five genes (orange group) from the *Roseburia* genomes. Observed Spearman correlations (as in **C**) are marked with a black vertical line and the null distribution is shown in gray. **E** Phenylalanine metabolism pathway (KEGG ko00360). Detected compounds and enzymes are colored in blue and orange

four (orange group) genes from all 22,832 *Roseburia* genes. Indeed, all four observed effects were extreme when compared to random sets of genes (Fig. 5D), confirming that the opposing compound–enzyme associations are unlikely to be due to other *Roseburia* genes. The complete sets of detected compounds and enzymes are shown in Fig. 5E, with group separation indicated. These results reveal two mechanisms, each characterized by a mutually exclusive set of enzymes and compounds, outlining how species-level differences in *Roseburia* may affect phenylalanine metabolites in circulation.

Additional evidence for the capability of *Roseburia* to alter compounds in the phenylalanine metabolism pathway is provided by a study that performed targeted metabolomics using reference standards on supernatants of individual microbial species grown

in defined media [25], including *Roseburia inulinivorans*, a close relative of *Roseburia* (MSP 112) with species-level annotation (Fig. 4A). A change of at least twofold in *R. inulinivorans* supernatant was observed for two compounds: an increase in 2-hydroxyacetate, which is encoded downstream of *yhDR* and *hisC*, and a decrease in succinate, which is encoded upstream of *mphE* (Additional file 1: Fig. S8, Fig. 5D). Although these compounds were not detected using our plasma metabolomics methods, the direction of changes is aligned with the relevant genes and supports an active role of *Roseburia* in phenylalanine metabolism pathways.

## Discussion

Growing evidence suggests that the gut microbiome mediates effects on vaccine efficacy through microbial ligands, circulating metabolites or both. Here, we identified microbial species associated with BCG-induced specific (adaptive immune memory) and heterologous (trained immunity) responses. De novo assembly methods and associations with circulating metabolites and cytokines at multiple time points enabled us to obtain statistically robust microbial signals that would otherwise be missed by reference-based methods.

We uncovered a total of 43 immunomodulatory microbes, including two species that showed consistent effects through metagenomic and metabolomic analyses: *Roseburia* (MSP 112) negatively associated with cytokine responses upon BCG-induced nonspecific immunity, while *Eggerthella lenta* positively associated with BCG-induced specific T cell-mediated memory responses. To date, the only other study investigating the BCG vaccination and human gut microbiota was performed with a cohort of 48 infants from Bangladesh [20]. Positive associations were observed between Actinobacteria (including *Bifidobacterium*) and immune responses, as measured by T cell proliferation and delayed-type hypersensitivity skin test, 15 weeks after BCG vaccination at birth. Although it is difficult to extrapolate these results to our cohort, we also observed that a genus of Actinobacteria (*Eggerthella*) was associated with specific T cell immune responses upon BCG vaccination.

Microbial products and cell wall components such as low-dose lipopolysaccharides (LPS) are known to modulate immune response and tolerance through engagement of host pattern recognition receptors [26]. Trained immunity responses in particular were shown to be dependent on engagement of the NOD2 receptor by muramyl dipeptide (MDP), the active component of mycobacterial peptidoglycans [27], which positively correlated with changes in IL-6, IL-1 $\beta$ , and TNF- $\alpha$  upon *S. aureus* stimulation 3 months after BCG vaccination [28]. As MDP can originate from the microbiome [29], it is unsurprising to detect numerous microbial associations with BCG-induced trained immunity (Fig. 2D, E).

*Roseburia* species produce SCFAs such as butyrate, which are known to possess immunomodulatory properties [30]. Interestingly, butyrate supplementation reduces induction of trained immunity [31], which could be speculated to be an additional mechanism through which *Roseburia* modulates trained immunity. Associations between *Roseburia* and atherosclerosis [32] — a process in which trained immunity also plays an important role [33] — as well as between *Eggerthella* and rheumatoid arthritis [34], multiple sclerosis [35], and inflammatory bowel disease [36] further suggest immunomodulatory roles for these species. Moreover, *E. lenta* is often associated with

serious gastrointestinal infections [37] and has been shown to increase intestinal T helper 17 cell numbers and worsen murine models of colitis.

Upon analyzing microbiome composition in conjunction with health-related variables from a questionnaire, we identified correlations between variations in the intestinal microbiome and adult vaccinations against influenza, yellow fever, and HPV. However, we are unable to determine whether these vaccinations altered microbiome composition or whether different lifestyles associated with additional vaccinations contributed to differences in the microbiome. In line with this, changes in the microbiome associated with yellow fever and influenza vaccination also coincided with participants traveling outside Europe. Variables including history of vaginal fungal infections, sex, diet, diarrhea, antibiotic use, and daily exercise also correlated with the number of changes in species relative abundance. The latter five factors are known to impact the gut microbiome [38–42]. Interestingly, hay fever during youth was associated with *Verrucomicrobia*, which has been observed at higher abundance in infants with allergy [43]. While gut microbiome profiles correlated with a number of health-related variables, overall metabolic health (e.g., BMI) did not correlate with cytokine responses upon stimulation, likely since our cohort is composed of healthy young adults (median age 23 and BMI = 25 at the 85th percentile).

The immunomodulatory microbes we identified, in particular *Roseburia* (MSP 112), had a stronger impact on circulating metabolites than the other microorganisms, suggesting a mechanism through which immune responses are influenced. Of the metabolites associated with metagenomic species, several were already known to be produced or modified by the gut microbiome, including acetyl-N-formyl-5-methoxykynurenamine (AFMK) [44], p-cresol sulfate [45], 5-hydroxytryptophol [46], hydrocinnamic acid [47], 4-ethylphenyl sulfate [48], hippurate [49], mandelonitrile [50], and 3-methylindole [51].

Most of the metabolites corresponding with microbiome changes were related to phenylalanine, an essential amino acid involved in many processes and that is also converted into tyrosine. Tyrosine comprises catecholamines and is involved in fumarate production, both of which are known to induce trained immunity responses [52]. Phenylalanine to tyrosine synthesis is facilitated by cofactor BH<sub>4</sub>, whose availability is influenced by pro- and anti-inflammatory stimuli in human macrophages. In general, increased serum phenylalanine concentrations are found during immune activation in humans, most likely due to the reduced converting rate. In phenylketonuria, a human congenital disorder in which a mutation in phenylalanine hydroxylase leads to high phenylalanine concentrations, a deficit of tyrosine and catecholamines is observed [53]. Microbial derivatives of phenylalanine and tyrosine, including microbiota-dependent metabolites p-cresol sulfate and phenylacetylglutamine [54], are known to contribute to chronic kidney disease pathogenesis and disease progression by inducing renal damage and inflammation [55, 56]. Furthermore, these metabolites have been associated with overall mortality and cardiovascular disease in individuals with kidney disease [54], and phenylacetylglutamine is a risk factor for major adverse cardiovascular events [57].

The *Roseburia* genus was associated with 11 out of 15 compounds in the phenylalanine pathway, with MSP 112 having the strongest effect. Moreover, the *Roseburia* genus encodes enzymes in the phenylalanine metabolism pathway itself. Related Firmicutes *Ruminococcus*, *Eubacterium* and *Coprococcus*, as well as *Bifidobacterium* displayed

decreased enzyme presence compared to *Roseburia*. The enzymes detected in MSP 112 and absent in the majority of other *Roseburia* genomes were L-phenylalanine N-acetyltransferase (*aaaT*), acyl-CoA thioesterase (*paaI*), and 4-hydroxy 2-oxovalerate aldolase (*mhpE*), which are found at different steps in the phenylalanine pathway. Notably, *mhpE* is involved in the production of pyruvate [58], and although pyruvate was not associated with *Roseburia* itself, it is used during trained immunity to increase metabolic activity [21].

Limited sensitivity and coverage of current metagenomic methods do not exclude the existence of additional, undetected immunomodulatory microbes. Nevertheless, associations with cytokine expression, circulating metabolites, and differential presence of phenylalanine metabolism enzymes increase the confidence that the presence of *Roseburia* in the gut microbiome dampens trained immune responses after BCG vaccination. Further investigation of the role of phenylalanine metabolism in these responses is warranted in future studies, as a possible causal role between *Roseburia* species and phenylalanine metabolism is supported by recently published in vitro experiments and should continue to be investigated in similar controlled settings. Furthermore, since individuals included in this study were all from Western European ancestry, of adult age, and living in similar environments, these findings should be validated in additional cohorts of different age groups and of different genetic backgrounds. This study nevertheless provides a detailed insight into associations between assembled microbial genomes, plasma metabolites and subject-specific immune responses.

## Conclusions

The data presented here details for the first time microbiome-associated effects on BCG vaccine responsiveness in a large human cohort. The pathways and microbial species that mediate BCG-induced trained and specific immune responses contribute to the understanding of subject-specific immunity and may inform personalized strategies to enhance vaccine efficacy.

## Methods

### Experimental model and subject details

In the 300BCG study, 321 healthy Dutch individuals were included from April 2017 until June 2018. Individuals were not included if they did not have a Western European ancestry, used systemic medication other than oral contraceptives and acetaminophen, used antibiotics 3 months before inclusion, received previous BCG vaccination, had a history of tuberculosis, any febrile illness 4 weeks before participation, vaccination 3 months before participation, or a medical history of immunodeficiency. Individuals were excluded for not meeting inclusion criteria ( $n = 2$ ), medication use ( $n = 4$ ), different time of vaccination ( $n = 18$ ), or a missing stool sample ( $n = 1$ ). At the hospital, written informed consent was obtained, stool and blood samples were collected, and participants were vaccinated with a standard dose of 0.1 mL BCG Bulgaria (InterVax) intradermally in the left upper arm. At 2 weeks and 3 months after vaccination, blood was again collected and a questionnaire was completed. This study was approved by the Arnhem-Nijmegen Medical Ethical Committee (NL58553.091.16) and performed in accordance with the Declaration of Helsinki.

## Method details

### *PBMC isolation and stimulation*

Ficoll-Paque (GE Healthcare, UK) density gradient separation was used to isolate peripheral blood mononuclear cells (PBMCs) from EDTA whole blood. After washing twice with phosphate buffered saline (PBS), cells were counted with the Sysmex hematology analyzer (XN-450) and resuspended in culture medium consisting of Dutch modified RPMI 1640 (Invitrogen, USA) supplemented with 50 µg/mL gentamicin, 2 mM Glutamax (Gibco Life Technologies), and 1 mM pyruvate (Gibco Life Technologies). Five hundred thousand PBMCs per well were added to a round bottom 96-well plate (Greiner) and stimulated with culture medium (control), heat-killed *Mycobacterium tuberculosis* HR37v (5 µg/mL), or heat-killed *Staphylococcus aureus* (10<sup>6</sup> CFU/mL, clinical isolate), and incubated at 37 °C, 5% CO<sub>2</sub>, for 24 h or 7 days. Medium of 7 days stimulation was supplemented with 10% human pooled serum. After 24 h and 7 days, supernatants were collected and stored at -20 °C until further analysis.

### *Cytokine measurements*

IL-1β, IL-6, and TNF-α (R&D Systems) concentrations were measured in 24-h supernatants with ELISA, according to the manufacturer's protocol. IFN-γ was measured in 7-day supernatants with Luminex (ProcartaPlex, Thermo Fisher), according to the manufacturer's protocol. The cytokine concentrations were measured in pg/mL, adjusted for plate variation using linear model residuals, and log-transformed.

### *Untargeted plasma metabolomics*

Plasma was obtained from participants by centrifugation of EDTA blood and was immediately stored at -80°C until further analysis. Flow injection-time-of-flight mass (flow injection TOF-M) spectrometry was performed by General Metabolomics (Boston, USA), as previously described [59]. A list of putative metabolites was annotated with a series of analysis strategies including deisotoping, decluttering, adduct detection, and library matching in KEGG, HMDB and CHEBI databases, resulting in 1,607 annotated features. All metabolomic measurements were performed in duplicate, and the average value was calculated for each sample per feature.

### *Stool samples collection and sequencing*

Stool samples were collected at home the day before or on the day of blood collection, time was recorded, and samples were stored in the refrigerator at home. At the hospital visit before BCG vaccination, samples were immediately aliquoted and stored at -80 °C.

Nucleic acids were extracted the AllPrep 96 PowerFecal DNA/RNA kit from QIAGEN (custom product # 1114341). This method pairs bead-beating on a Tissuelyser II (QIAGEN) with a 96-well AllPrep protocol and is available through QIAGEN. Bead-beating is performed twice at 20 Hz for 5 minutes each round with a rotation of the plate in between rounds. Purified DNA was stored at -20 °C. Metagenomic sequencing libraries were prepared from 2 ng of input DNA using the Nextera XT DNA Library Preparation kit (Illumina) according to the manufacturer's recommended protocol. Prior to sequencing, libraries were pooled by collecting equal volumes of each library. Insert sizes and concentrations for each pooled library were determined using an

Agilent Bioanalyzer DNA 1000 kit (Agilent Technologies) prior to sequencing on an Illumina NovaSeq 6000 with 151 bp paired-end reads to yield ~ 10 million paired-end reads per sample. Data was analyzed using the Broad Picard Pipeline which includes de-multiplexing and data aggregation (<https://broadinstitute.github.io/picard>).

## **Quantification and statistical analysis**

### ***Metagenomics data processing***

The quality control for sequencing data was conducted using Trim Galore! to detect and remove sequencing adapters (minimum overlap of 5 bp) and kneadData to remove human DNA contamination and trim low-quality sequences (HEADCROP:15 SLIDINGWINDOW:1:20), retaining reads that were at least 50 bp. Metagenomic reads from all cohorts were assembled individually for each sample into contigs using MegaHIT [60], followed by an open reading frame prediction with Prodigal [61] and retaining only full length genes (containing both start and stop codon).

A non-redundant gene catalog was constructed by clustering predicted genes based on sequence similarity at 95% identity and 90% coverage of the shorter sequence using CD-HIT [62]. Reads were mapped to the gene catalog with BWA [63], filtered to include strong mappings with at least 95% sequence identity over the length of the read, counted (count matrix) and normalized to transcript-per-million (TPM matrix). Count matrix served as an input for binning genes into metagenomic species pan genomes (core and accessory genes) using MSPminer with default settings [64].

The abundance of an MSP was quantified as median gene TPM. The genes forming each MSP were quantified as reads per kilobase (RPK).

We annotated the gene catalog at species, genus, and phylum levels with NCBI RefSeq (version May 2018) as described previously [65]. The KEGG Orthogroups were assigned to each gene in the catalog using eggNOG-Mapper [66].

The PhyloPhlAn 3.0 with default settings [67] was used to organize the phylogenetic tree of MSPs based on similarity in core genes.

For comparison with the 500FG cohort, we used the sample pipeline as described in the associated publication [12]. After the kneadData quality control step, marker-gene based taxonomic annotation tool MetaPhlAn [68] yielded relative abundances of 202 species, of which 199 were shared between the two cohorts.

### ***Clustering of metagenomic species***

The abundances of MSPs were normalized to relative abundance in each sample. The clusters were determined based on the Jensen-Shannon divergence and a consensus hierarchical clustering with pvclust using 100 bootstrap runs. The number of clusters was determined to maximize the difference of average between- and within-cluster divergences.

### ***Variable associations with metagenomic species***

We tested variables known for at least 200 samples and had a minimum 40 samples per categorical variable level. Maaslin2 generalized linear model was used to model differential abundance of each MSP and the recommended threshold of  $P_{\text{FDR}} < 0.25$  was applied.



### **Significance of cytokine expression changes after vaccination**

Significance of increased log cytokine expression over three time points was assessed using paired Kruskal-Wallis test. The  $T$ -test was used to determine significance of fold changes from the baseline time point for 2 weeks and 3 months. Spearman correlation was used to test for significant correlations between cytokine levels.

### **Associations between metagenomic species and changes in cytokine expressions**

To find potential immunomodulatory species affecting the magnitude in the immune response after BCG vaccination, we combined the data on cytokine expression over three time points with stool samples at baseline and subjects' age and sex. We included 303 samples vaccinated in the morning and at least two cytokine measurements ( $i = 1 \dots 303$ ) and 163 metagenomic species (MSPs) with at least 20% prevalence ( $m = 1 \dots 202$ ). Cytokine expressions ( $c$ ) were quantified as fold change against the baseline for measurements after 2 weeks and 3 months ( $t = 2 \text{ weeks}, 3 \text{ months}$ ). To equalize the variance in cytokine expressions, the fold changes were standardized for each cytokine and time point using the z-score transform.

Trained and specific responses were assessed with two separate models. The cytokines TNF- $\alpha$ , IL-1 $\beta$  and IL-6 stimulated with *S. aureus* were used to quantify the trained immune response, while IFN- $\gamma$  stimulated with *M. tuberculosis* was used to detect the specific immune response. Excluding missing values, the trained immunity model comprised 1685 observations, while the specific response model comprised 535 observations. For both response types the z-scored fold changes  $z_{c,t,i}$  were fit using the linear model

$$z_{c,t,i} = \alpha_0 + \alpha_a \text{age}_i + \alpha_s \text{sex}_i + \alpha_m \text{msp}_{m,i} + \epsilon_{c,t,i},$$

with the following parameters:  $\epsilon_{c,t,i}$  is the unexplained variance,  $\alpha_0$  the intercept, and  $\alpha_a$ ,  $\alpha_s$ ,  $\alpha_m$  coefficients for age, sex, and MSP abundance, respectively. The subject-specific variables are  $\text{age}_i$  (continuous),  $\text{sex}_i$  (discrete), and  $\text{msp}_{m,i}$  the relative of the MSP  $m$ , converted to ranks (where no abundance amounts to rank 0 and maximum abundance to rank 302). Note that there is no separate longitudinal or cytokine coefficient, as we sought species that associate with responses in all cytokines at both time points, effectively treating the measurements as replicates in order to decrease model complexity and increase statistical power. Despite the observations not being entirely independent (each subject can be repeated in multiple time points and cytokines), we chose to treat them as such rather than average across cytokines and timepoints, treating them as separate experiments that bear additional information.

To account for co-abundance effects, each of the two models was fit once for each MSP, using the R function `lm`. The MSPs were deemed immunomodulatory if  $\alpha_m$  had a significant contribution after multiple testing adjustment ( $P_{\text{FDR}} < 0.2$ ,  $T$ -test). The explained variance in cytokine expression fold change was computed for each MSP, using the model based on age and sex alone:

$$z_{c,t,i} = \beta_0 + \beta_a \text{age}_i + \beta_s \text{sex}_i + \gamma_{c,t,i},$$

where  $\beta_0$ ,  $\beta_a$ , and  $\beta_s$  are the ordinary least squares coefficients and the explained variance is estimated as squared Pearson correlation between ranked MSP abundance and residuals  $\gamma_{c,t,i}$ . Note that this estimate is conservative in the context of a single isolated

MSP, as the effect of age and sex is explained away using their full least squared coefficients.

### **Metabolomics statistical analysis**

Metabolite intensity was standardized (*z*-scored) across samples. Clusters were obtained using *k*-means, with the number of clusters determined by optimal matching to the molecular classes (from HMDB database) using adjusted random index (ARI) measure. The two dimensional projection was obtained using t-SNE.

Canonical correlation analysis (CCA) between standardized metabolite intensity matrix **M** (321 samples times 1607 metabolite features) and MSP abundance matrix **S** (321 samples times 345 MSPs) was performed by singular value decomposition of the covariance matrix

$$\mathbf{UDV}^T = \mathbf{M}^T \mathbf{S},$$

where **U** and **V** are orthonormal matrices of canonical vectors and **D** is the diagonal matrix of explained variances. The significance of canonical vectors was assessed by comparing diagonal entries in **D** against the null distribution, simulated by 1000 random permutations of **M** and **S** and subject to threshold  $P < 0.05$ . The top 25 canonical vectors in **U**, **V** were then projected to two dimensions using t-SNE.

Metabolite intensities were compared to relative species abundance and cytokine expressions using Spearman correlation subject to  $P_{\text{Bonferroni}} < 0.05$ .

To compare intensity of multiple metabolites with abundance of multiple enzymes, the respective intensities and abundances were averaged and Spearman correlation was calculated.

### **Associations between pathways, genes, and compounds**

The KEGG pathways KGML data formats were used to link metabolite and enzyme identifiers with metabolic pathways. The enzyme and pathway identifiers used the KEGG ortholog (ko) format. The metabolite features mapped to 721 KEGG compound identifiers, contained in 350 metabolic pathways. Out of 10,049 enzymes associated with the 350 pathways, 2726 were detected in the stool metagenomes.

## **Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02482-0>.

**Additional file 1: Supplementary Figures.** This file contains Figures S1-S8.

**Additional file 2: Table S1.** Associations between microbiome clusters 1-3 and subject anthropometric and questionnaire variables. Associations between MSP abundance and subject anthropometric and questionnaire variables (Maaslin 2).

**Additional file 3: Table S2.** Coefficients and significance values for each metagenomic species pangenome (MSP) in the log-linear model of trained and specific immune response.

**Additional file 4: Table S3.** Individual Pearson correlations between MSP abundance ranks and cytokine fold changes, adjusted for age and sex, 2 weeks and 3 months after BCG vaccination.

**Additional file 5: Table S4.** Significant correlations between MSP abundances and metabolite intensities (Spearman correlation,  $P_{\text{Bonferroni}} < 0.05$ ).

**Additional file 6: Table S5.** Phenylalanine metabolism pathway enzyme identity and prevalence detected in the metagenomic assembly encoded by *Bifidobacterium*, *Roseburia*, *Ruminococcus*, *Eubacterium* and *Coprococcus*.

**Additional file 7: Table S6.** List of key resources.

**Additional file 8.** Review history.

### Acknowledgements

The authors thank all volunteers for their participation in this study. In addition, we would like to thank Helga Dijkstra, Heidi Lemmers, and Trees Jansen for their help collecting materials in the 300BCG study. We thank Melanie Schirmer and Damian Plichta for fruitful discussions and Theresa Reimels for manuscript editing and preparation of the figures.

### Peer review information

Kevin Pang was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Review history

The review history is available as Additional file 8.

### Authors' contributions

M.S. performed metagenomic and metabolomic data analysis and modeling. M.S. and V.P.M. performed analysis of immunological data. M.S., R.J.X., V.P.M., M.G.N., and H.V. interpreted the results. V.P.M., V.A.C.M.K., S.J.C.F.M.M., L.C.J.d.B., M.G.N., and R.J.X. designed the study cohort. V.P.M., M.S., M.G.N., and R.J.X. wrote the paper. M.G.N. and R.J.X. served as project leaders. L.A.B.J., R.v.C., H.V., M.G.N., and R.J.X. served as principal investigators. The author(s) read and approved the final manuscript.

### Funding

This study was supported by an ERC Advanced grant (#833247) and a Spinoza grant of the Netherlands Organization for Scientific Research to M.G.N. as well as grants from the National Institutes of Health (P30 DK043351) and the Center for Microbiome Informatics and Therapeutics to R.J.X.

### Availability of data and materials

All the datasets and code generated during this study are available at <https://gitlab.com/xavier-lab-computation/public/bcg300> [69] and <https://doi.org/10.5281/zenodo.5148664> [70], released under the MIT license. The raw metagenomic sequencing reads are available in the NCBI BioProject under identifier PRJNA685797 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA685797>) [71]. This study did not generate new unique reagents; key resources are listed in Additional file 7: Table S6. Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding authors.

### Declarations

#### Ethics approval and consent to participate

This study was approved by the Arnhem-Nijmegen Medical Ethical Committee (NL58553.091.16) and performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants prior to sample collection.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>2</sup>Department of Internal Medicine, Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands. <sup>3</sup>Department of Computational Biology for Individualised Infection Medicine, Centre for Individualised Infection Medicine (CiIM) & TWINCORE, Joint Ventures Between the Helmholtz-Centre for Infection Research (HZI) and the Hannover Medical School (MHH), Hannover, Germany. <sup>4</sup>Research Center for Vitamins and Vaccines, Bandim Health Project, Statens Serum Institut, Copenhagen, Denmark. <sup>5</sup>Odense Patient Data Explorative Network, University of Southern Denmark/Odense University Hospital, Odense, Denmark. <sup>6</sup>Department of Medical Genetics, Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania. <sup>7</sup>Department for Genomics & Immunoregulation, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany. <sup>8</sup>Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>9</sup>Center for Computational and Integrative Biology, Department of Molecular Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA.

Received: 23 December 2020 Accepted: 30 August 2021

Published online: 22 September 2021

### References

1. Garly ML, Martins CL, Bale C, Balde MA, Hedegaard KL, Gustafson P, et al. BCG scar and positive tuberculin reaction associated with reduced child mortality in West Africa. A non-specific beneficial effect of BCG? Vaccine. 2003;21(21-22):2782–90. [https://doi.org/10.1016/S0264-410X\(03\)00181-6](https://doi.org/10.1016/S0264-410X(03)00181-6).
2. van't Wout JW, Poell R, van Furth R. The role of BCG/PPD-activated macrophages in resistance against systemic candidiasis in mice. Scand J Immunol. 1992;36(5):713–9. <https://doi.org/10.1111/j.1365-3083.1992.tb03132.x>.
3. Tribouley J, Tribouley-Duret J, Appriou M. Effect of Bacillus Callmette Guerin (BCG) on the receptivity of nude mice to *Schistosoma mansoni*. C R Seances Soc Biol Fil. 1978;172(5):902–4.
4. Netea MG, Quintin J, van der Meer JW. Trained immunity: a memory for innate host defense. Cell Host Microbe. 2011;9(5):355–61. <https://doi.org/10.1016/j.chom.2011.04.006>.

5. Netea MG, Dominguez-Andres J, Barreiro LB, Chavakis T, Divangahi M, Fuchs E, et al. Defining trained immunity and its role in health and disease. *Nat Rev Immunol*. 2020;20(6):375–88. <https://doi.org/10.1038/s41577-020-0285-6>.
6. Kleinnijenhuis J, Quintin J, Preijers F, Joosten LA, Iffrim DC, Saeed S, et al. Bacille Calmette-Guerin induces NOD2-dependent nonspecific protection from reinfection via epigenetic reprogramming of monocytes. *Proc Natl Acad Sci U S A*. 2012;109(43):17537–42. <https://doi.org/10.1073/pnas.1202870109>.
7. Kleinnijenhuis J, Quintin J, Preijers F, Joosten LA, Jacobs C, Xavier RJ, et al. BCG-induced trained immunity in NK cells: role for non-specific protection to infection. *Clin Immunol*. 2014;155(2):213–9. <https://doi.org/10.1016/j.clim.2014.10.005>.
8. Kaufmann E, Sanz J, Dunn JL, Khan N, Mendonca LE, Pacis A, et al. BCG educates hematopoietic stem cells to generate protective innate immunity against tuberculosis. *Cell*. 2018;172(1–2):176–90.e19. <https://doi.org/10.1016/j.cell.2017.12.031>.
9. Cirovic B, de Bree LCJ, Groh L, Blok BA, Chan J, van der Velden W, et al. BCG vaccination in humans elicits trained immunity via the hematopoietic progenitor compartment. *Cell Host Microbe*. 2020;28(2):322–34.e5. <https://doi.org/10.1016/j.chom.2020.05.014>.
10. Kleinnijenhuis J, Quintin J, Preijers F, Benn CS, Joosten LA, Jacobs C, et al. Long-lasting effects of BCG vaccination on both heterologous Th1/Th17 responses and innate trained immunity. *J Innate Immun*. 2014;6(2):152–8. <https://doi.org/10.1159/000355628>.
11. Koeken VA, de Bree LCJ, Mourits VP, Moorlag SJ, Walk J, Cirovic B, et al. BCG vaccination in humans inhibits systemic inflammation in a sex-dependent manner. *J Clin Invest*. 2020;130(10):5591–602. <https://doi.org/10.1172/JCI133935>.
12. Schirmer M, Smeekens SP, Vlamakis H, Jaeger M, Oosting M, Franzosa EA, et al. Linking the human gut microbiome to inflammatory cytokine production capacity. *Cell*. 2016;167(4):1125–36.e8. <https://doi.org/10.1016/j.cell.2016.10.020>.
13. Levy M, Kolodziejczyk AA, Thaiss CA, Elinav E. Dysbiosis and the immune system. *Nat Rev Immunol*. 2017;17(4):219–32. <https://doi.org/10.1038/nri.2017.7>.
14. Chang PV, Hao L, Offermanns S, Medzhitov R. The microbial metabolite butyrate regulates intestinal macrophage function via histone deacetylase inhibition. *Proc Natl Acad Sci U S A*. 2014;111(6):2247–52. <https://doi.org/10.1073/pnas.1322269111>.
15. Atarashi K, Tanoue T, Ando M, Kamada N, Nagano Y, Narushima S, et al. Th17 cell induction by adhesion of microbes to intestinal epithelial cells. *Cell*. 2015;163(2):367–80. <https://doi.org/10.1016/j.cell.2015.08.058>.
16. Ciabattini A, Olivieri R, Lazzeri E, Medagliani D. Role of the microbiota in the modulation of vaccine immune responses. *Front Microbiol*. 2019;10:1305. <https://doi.org/10.3389/fmicb.2019.01305>.
17. Zimmermann P, Curtis N. The influence of the intestinal microbiome on vaccine responses. *Vaccine*. 2018;36(30):4433–9. <https://doi.org/10.1016/j.vaccine.2018.04.066>.
18. Lynn MA, Tumes DJ, Choo JM, Sribnaia A, Blake SJ, Leong LEX, et al. Early-life antibiotic-driven dysbiosis leads to dysregulated vaccine immune responses in mice. *Cell Host Microbe*. 2018;23(5):653–60.e5. <https://doi.org/10.1016/j.chom.2018.04.009>.
19. Nadeem S, Maurya SK, Das DK, Khan N, Agrewala JN. Gut dysbiosis thwarts the efficacy of vaccine against *Mycobacterium tuberculosis*. *Front Immunol*. 2020;11:726. <https://doi.org/10.3389/fimmu.2020.00726>.
20. Huda MN, Lewis Z, Kalanetra KM, Rashid M, Ahmad SM, Raqib R, et al. Stool microbiota and vaccine responses of infants. *Pediatrics*. 2014;134(2):e362–72. <https://doi.org/10.1542/peds.2013-3937>.
21. Costea PI, Hildebrand F, Arumugam M, Backhed F, Blaser MJ, Bushman FD, et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol*. 2018;3(1):8–16. <https://doi.org/10.1038/s41564-017-0072-8>.
22. Luo W, Shen Z, Deng M, Li X, Tan B, Xiao M, et al. *Roseburia intestinalis* supernatant ameliorates colitis induced in mice by regulating the immune response. *Mol Med Rep*. 2019;20(2):1007–16. <https://doi.org/10.3892/mmr.2019.10327>.
23. Liu H, Wang J, He T, Becker S, Zhang G, Li D, et al. Butyrate: a double-edged sword for health? *Adv Nutr*. 2018;9(1):21–9. <https://doi.org/10.1093/advances/nmx009>.
24. Louis P, Duncan SH, McCrae SI, Millar J, Jackson MS, Flint HJ. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol*. 2004;186(7):2099–106. <https://doi.org/10.1128/JB.186.7.2099-2106.2004>.
25. Han S, Van Treuren W, Fischer CR, Merrill BD, DeFelicis BC, Sanchez JM, et al. A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature*. 2021;595(7867):415–20. <https://doi.org/10.1038/s41586-021-03707-9>.
26. Seeley JJ, Ghosh S. Molecular mechanisms of innate memory and tolerance to LPS. *J Leukoc Biol*. 2017;101(1):107–19. <https://doi.org/10.1189/jlb.3MR0316-118RR>.
27. van der Meer JH, Netea MG, Dinarello CA. Modulation of muramyl dipeptide stimulation of cytokine production by blood components. *Clin Exp Immunol*. 2009;156(3):428–33. <https://doi.org/10.1111/j.1365-2249.2009.03926.x>.
28. Mourits VP, Koeken V, de Bree LCJ, Moorlag S, Chu WC, Xu X, et al. BCG-induced trained immunity in healthy individuals: the effect of plasma muramyl dipeptide concentrations. *J Immunol Res*. 2020;2020:5812743–8. <https://doi.org/10.1155/2020/5812743>.
29. Johnson JW, Fisher JF, Mobashery S. Bacterial cell-wall recycling. *Ann N Y Acad Sci*. 2013;1277(1):54–75. <https://doi.org/10.1111/j.1749-6632.2012.06813.x>.
30. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol*. 2017;19(1):29–41. <https://doi.org/10.1111/1462-2920.13589>.
31. Cleophas MCP, Ratter JM, Bekkering S, Quintin J, Schraa K, Stroes ES, et al. Effects of oral butyrate supplementation on inflammatory potential of circulating peripheral blood mononuclear cells in healthy and obese males. *Sci Rep*. 2019;9(1):775. <https://doi.org/10.1038/s41598-018-37246-7>.
32. Kasahara K, Krautkramer KA, Org E, Romano KA, Kerby RL, Vivas EI, et al. Interactions between *Roseburia intestinalis* and diet modulate atherogenesis in a murine model. *Nat Microbiol*. 2018;3(12):1461–71. <https://doi.org/10.1038/s41564-018-0272-x>.
33. Flores-Gomez D, Bekkering S, Netea MG, Riksen NP. Trained immunity in atherosclerotic cardiovascular disease. *Arterioscler Thromb Vasc Biol*. 2021;41(1):62–9.
34. Chen J, Wright K, Davis JM, Jeraldo P, Marietta EV, Murray J, et al. An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Med*. 2016;8(1):43. <https://doi.org/10.1186/s13073-016-0299-7>.
35. Cekanaviciute E, Yoo BB, Runia TF, Debelius JW, Singh S, Nelson CA, et al. Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. *Proc Natl Acad Sci U S A*. 2017;114(40):10713–8. <https://doi.org/10.1073/pnas.1711235114>.

36. Palm NW, de Zoete MR, Cullen TW, Barry NA, Stefanowski J, Hao L, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell*. 2014;158(5):1000–10. <https://doi.org/10.1016/j.cell.2014.08.006>.
37. Gardiner BJ, Tai AY, Kotsanas D, Francis MJ, Roberts SA, Ballard SA, et al. Clinical and microbiological characteristics of *eggerthella lenta* bacteremia. *J Clin Microbiol*. 2015;53(2):626–35. <https://doi.org/10.1128/JCM.02926-14>.
38. Modi SR, Collins JJ, Relman DA. Antibiotics and the gut microbiota. *J Clin Invest*. 2014;124(10):4212–8. <https://doi.org/10.1172/JCI72333>.
39. Allen JM, Mailing LJ, Niemi GM, Moore R, Cook MD, White BA, et al. Exercise alters gut microbiota composition and function in lean and obese humans. *Med Sci Sports Exerc*. 2018;50(4):747–57. <https://doi.org/10.1249/MSS.0000000000001495>.
40. Clarke SF, Murphy EF, O'Sullivan O, Lucey AJ, Humphreys M, Hogan A, et al. Exercise and associated dietary extremes impact on gut microbial diversity. *Gut*. 2014;63(12):1913–20. <https://doi.org/10.1136/gutjnl-2013-306541>.
41. Kim YS, Unno T, Kim BY, Park MS. Sex differences in gut microbiota. *World J Mens Health*. 2020;38(1):48–60. <https://doi.org/10.5534/wjmh.190009>.
42. Vandeputte D, Falony G, Vieira-Silva S, Tito RY, Joossens M, Raes J. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut*. 2016;65(1):57–62. <https://doi.org/10.1136/gutjnl-2015-309618>.
43. Shen X, Wang M, Zhang X, He M, Li M, Cheng G, et al. Dynamic construction of gut microbiota may influence allergic diseases of infants in Southwest China. *BMC Microbiol*. 2019;19(1):123. <https://doi.org/10.1186/s12866-019-1489-4>.
44. Lee T, Clavel T, Smirnov K, Schmidt A, Lagkouvardos I, Walker A, et al. Oral versus intravenous iron replacement therapy distinctly alters the gut microbiota and metabolome in patients with IBD. *Gut*. 2017;66(5):863–71. <https://doi.org/10.1136/gutjnl-2015-309940>.
45. Saito Y, Sato T, Nomoto K, Tsuji H. Identification of phenol- and p-cresol-producing intestinal bacteria by using media supplemented with tyrosine and its metabolites. *FEMS Microbiol Ecol*. 2018;94(9):fiy125.
46. Yano JM, Yu K, Donaldson GP, Shastri GG, Ann P, Ma L, et al. Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell*. 2015;161(2):264–76. <https://doi.org/10.1016/j.cell.2015.02.047>.
47. Coman V, Vodnar DC. Hydroxycinnamic acids and human health: recent advances. *J Sci Food Agric*. 2020;100(2):483–99. <https://doi.org/10.1002/jsfa.10010>.
48. Sharon G, Garg N, Debelius J, Knight R, Dorrestein PC, Mazmanian SK. Specialized metabolites from the microbiome in health and disease. *Cell Metab*. 2014;20(5):719–30. <https://doi.org/10.1016/j.cmet.2014.10.016>.
49. Pallister T, Jackson MA, Martin TC, Zierer J, Jennings A, Mohny RP, et al. Hippurate as a metabolomic marker of gut microbiome diversity: modulation by diet and relationship to metabolic syndrome. *Sci Rep*. 2017;7(1):13670. <https://doi.org/10.1038/s41598-017-13722-4>.
50. Jaswal V, Palanivelu J, Ramalingam C. Effects of the Gut microbiota on Amygdalin and its use as an anti-cancer therapy: substantial review on the key components involved in altering dose efficacy and toxicity. *Biochem Biophys Rep*. 2018;14:125–32. <https://doi.org/10.1016/j.bbrep.2018.04.008>.
51. Roager HM, Licht TR. Microbial tryptophan catabolites in health and disease. *Nat Commun*. 2018;9(1):3294. <https://doi.org/10.1038/s41467-018-05470-4>.
52. van der Heijden C, Groh L, Keating ST, Kaffa C, Noz MP, Kersten S, et al. Catecholamines induce trained immunity in monocytes in vitro and in vivo. *Circ Res*. 2020;127(2):269–83. <https://doi.org/10.1161/CIRCRESAHA.119.315800>.
53. Williams RA, Mamotte CD, Burnett JR. Phenylketonuria: an inborn error of phenylalanine metabolism. *Clin Biochem Rev*. 2008;29(1):31–41.
54. Krautkramer KA, Fan J, Backhed F. Gut microbial metabolites as multi-kingdom intermediates. *Nat Rev Microbiol*. 2021;19(2):77–94. <https://doi.org/10.1038/s41579-020-0438-4>.
55. Nallu A, Sharma S, Ramezani A, Muralidharan J, Raj D. Gut microbiome in chronic kidney disease: challenges and opportunities. *Transl Res*. 2017;179:24–37. <https://doi.org/10.1016/j.trsl.2016.04.007>.
56. Vanholder R, Schepers E, Pletinck A, Nagler EV, Glorieux G. The uremic toxicity of indoxyl sulfate and p-cresyl sulfate: a systematic review. *J Am Soc Nephrol*. 2014;25(9):1897–907. <https://doi.org/10.1681/ASN.2013101062>.
57. Nemet I, Saha PP, Gupta N, Zhu W, Romano KA, Skye SM, et al. A cardiovascular disease-linked gut microbial metabolite acts via adrenergic receptors. *Cell*. 2020;180(5):862–77.e22. <https://doi.org/10.1016/j.cell.2020.02.016>.
58. Manjasetty BA, Powlowski J, Vrielink A. Crystal structure of a bifunctional aldolase-dehydrogenase: sequestering a reactive and volatile intermediate. *Proc Natl Acad Sci U S A*. 2003;100(12):6992–7. <https://doi.org/10.1073/pnas.1236794100>.
59. Fuhrer T, Heer D, Begemann B, Zamboni N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Anal Chem*. 2011;83(18):7074–80. <https://doi.org/10.1021/ac201267k>.
60. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674–6. <https://doi.org/10.1093/bioinformatics/btv033>.
61. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):119. <https://doi.org/10.1186/1471-2105-11-119>.
62. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28(23):3150–2. <https://doi.org/10.1093/bioinformatics/bts565>.
63. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
64. Plaza Onate F, Le Chatelier E, Almeida M, Cervino ACL, Gauthier F, Magoules F, et al. MSPminer: abundance-based reconstitution of microbial pan-genomes from shotgun metagenomic data. *Bioinformatics*. 2019;35(9):1544–52. <https://doi.org/10.1093/bioinformatics/bty830>.
65. Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, et al. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol*. 2014;32(8):834–41. <https://doi.org/10.1038/nbt.2942>.
66. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47(D1):D309–D14. <https://doi.org/10.1093/nar/gky1085>.

67. Asnicar F, Thomas AM, Beghini F, Mengoni C, Manara S, Manghi P, et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun.* 2020;11(1):2500.
68. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods.* 2012;9(8):811–4. <https://doi.org/10.1038/nmeth.2066>.
69. Strazar M, Mourits VP, Koeken VACM, de Bree LCJ, Moorlag SJCFM, Joosten LAB, et al. The influence of the gut microbiome on BCG-induced trained immunity: Github; 2021. <https://gitlab.com/xavier-lab-computation/public/bcg300>
70. Strazar M, Mourits VP, Koeken VACM, de Bree LCJ, Moorlag SJCFM, Joosten LAB, et al. The influence of the gut microbiome on BCG-induced trained immunity: Zenodo; 2021. <https://zenodo.org/record/5148664#.YSz6I4hKiks>
71. Strazar M, Mourits VP, Koeken VACM, de Bree LCJ, Moorlag SJCFM, Joosten LAB, et al. The influence of the gut microbiome on BCG-induced trained immunity: NCBI SRA; 2020. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA685797>

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

