Title: Newborn T Cell and Microbiota Developmental Trajectories and Their Inter-related Roles in Health and Disease

Authors: Alex Grier[1#], Nathan Laniewski[2#], Ann L. Gill[2], Haeja A. Kessler[2], Heidie Huyck[3], Jeanne Holden-Wiltse[4], Sanjukta Bandyopadhyay[4], Jennifer Carnahan[3], Andrew M. Dylag[3], David J. Topham[2], Ann R. Falsey[5], Mary T. Caserta[3], Gloria S. Pryhuber[3], Steven R. Gill[2], Andrew McDavid[4†], Kristin M. Scheible[3†*]

Affiliations:

[1]University of Rochester, Genomics Research Center

[2]University of Rochester, Department of Microbiology and Immunology

[3]University of Rochester, Department of Pediatrics

[4]University of Rochester, Department of Biostatistics and Computational Biology

[5]University of Rochester, Department of Medicine

* Corresponding Author: Kristin_Scheible@urmc.rochester.edu

†#Authors contributed equally to this manuscript

One Sentence Summary: T cells and microbiota follow predictable, coordinated trajectories in newborns, and their coordination is an important determinant of illness in the first year.

Keywords:

Development, fetus, neonate, newborn, preterm, prematurity, immunity, T cell, microbiome, microbiota, respiratory, intestine, systems immunology, systems biology, respiratory morbidity

Abstract:

Neonatal immune-microbiota co-development is poorly understood, yet appropriate recognition of – and response to – pathogens and commensal microbiota is critical to health. In this longitudinal study of 133 pre- and 79 full-term infants from birth through one year of age we found that postmenstrual age, or weeks from conception, is the dominant factor influencing T cell and mucosal microbiota development. However, numerous features of the T cell and microbiota trajectories remain unexplained by host age, and are better explained by discrete peri-

and post-natal events. Most strikingly, we establish that disruption of the normal developmental program, as is seen with discordant or atypical T cell and microbiota trajectories, increases the risk of illnesses in the first year of life. Altogether, this study presents compelling evidence that newborn health is marked by predictable, coordinated immune and microbiota development, but deviation from this pattern places an infant at significant risk for respiratory morbidity.

Introduction

Early immunity and microbiota in human infants have profound impacts on health and disease, but factors influencing their development are incompletely understood (1-3). Recent studies point to there being a developmental program that determines major shifts in bulk immune cell population distribution (T cells, B cells, granulocytes, monocytes, etc.), over the first 3 months of postnatal life (4). Likewise, there appears to be an age-related trajectory for gut and respiratory microbiota which, when disrupted by peri- and post-natal events, are associated with adverse outcomes such as atopy, stunted growth, and respiratory infection (5-9). Two recent studies demonstrated that the nasopharyngeal microbiome and virome together predict infant respiratory tract infection, but these studies left unresolved the microbiome's impact and dependency on immune development (10, 11).

The observation that the parallel developmental processes of the immune system and the microbiome are likely to be influenced by many of the same extrinsic factors and to have mutually formative effects on one another has given rise to the concept of a neonatal window of opportunity; a critical period of primary exposures and immune maturation which may have lifelong health implications (12) .Whether or not an exposure is differentially "remembered" by the immune system based on its timing during development is not known, but is particularly relevant when considering the long-lived adaptive immune system. Aberrant or mistimed

immune and microbiota trajectories, if found to be associated with disease states, have important implications for any infant-targeted interventions proposing to perturb either of these systems. To date, there are no published studies addressing how T cells and the mucosal microbiome are linked during early human development, and none have explored the degree to which immunity and mucosal bacteria interplay impact an infant's health in the first year of life. The need to understand the complex relationship between microbiota, the immune system and infant development is urgent, given the accelerating use of microbiome-based therapies in humans (13-16). In this work, we seek to directly address this critical knowledge gap using a systems biology approach modeling structured patterns of progression of T cell populations and the gut and respiratory microbiota relative to postmenstrual age. We identified several interdependencies between these systems which could not be explained by host age. By tracking clinical outcomes in our longitudinal cohort, we found that morbidity was increased in infants exhibiting atypical or discordant acquisition of microbiota and T cell populations, implicating aberrant co-development of these systems as an early marker for disease. To our knowledge, this work represents the most extensive assessment to date of the relationship between developing T cell populations and microbiota in humans, and it is the first to demonstrate a link between the co-development of these systems and clinical outcomes in infants.

Results

Study Design and Demographics

Neonatal subjects (n=267) born 23-42 weeks gestational age (GA) were recruited within 7 days of birth at the University of Rochester from 2012-2016, as part of the NIAID-sponsored Prematurity, Respiratory, Immune Systems and Microbiomes study (PRISM) (Fig. 1a). In all, 122 preterm (PT, < 37 0/7 weeks gestation) and 80 full-term (FT, ≥ 37 0/7 weeks gestation)

subjects completed the study to 12 months of age corrected for premature birth and were categorized as having or not having the primary outcome persistent respiratory disease (PRD) using previously published criteria (17). Cohort demographics are shown in Table 1. Sufficient blood to perform T cell phenotyping by flow cytometry at three pre-defined timepoints was collected from 55% of subjects at birth, 61% of subjects at NICU discharge, and 38% at 12 months. For microbiota profiling, inpatient samples were obtained weekly and outpatient samples for PT and FT were obtained monthly, with additional sampling during acute illness. After sample processing, 16S rRNA sequencing, quality control, and removing subjects without any immunophenotyping data, 149 subjects yielded 1748 usable nasal samples and 143 subjects yielded 1899 usable rectal samples. The median subject had 24 samples, with 28 days on average between samples. Finally, 109 and 117 subjects had sufficient combined T cell phenotyping and microbiota data to be included for immunome-nasal microbiota and immunome-rectal microbiota association analyses, respectively (Supplementary Tables 1-2).

Postmenstrual age exerts a greater influence on early T cell and microbiota development than does postnatal age.

We predicted, based on our and other previous studies that T cell and microbiota evolution in the first year of life would proceed in an age-dependent manner. Age can be defined in several useful ways in newborns, and these definitions have different implications. Postmenstrual age (PMA), or weeks from conception, takes into account that T cell maturation begins in utero, prior to antigen exposure, and continues postnatally as the infant grows. This is important when PT infants, who are born before completion of a gestational program, are included in a cohort. Alternatively, models based on postnatal exposures would most aptly

incorporate days of life (DOL), or days since birth, as the key factor affecting immune and microbiota development. In considering the strengths of a combined PT and FT cohort, we therefore tested two competing hypotheses. First, that the microbiome and T cells would be shaped primarily by days of life, i.e., postnatal exposures would be the essential driver of T cell and microbial maturation. Under this hypothesis, we would expect PT and FT infants to display overlapping states at birth that would diverge as PT subjects achieved term equivalent postmenstrual age, because they will be considerably older in days of life compared to FT newborns at that point. Alternatively, these systems could be more heavily influenced by an infant's physiologic maturity, therefore PMA. If this alternative proved true, PT and FT subjects would exhibit distinct profiles at birth, which would converge when PT infants achieved term equivalent age.

We applied unsupervised clustering approaches to reveal fine-grained, biologically interpretable categories of T cell populations and microbiota. The clustering algorithm FlowSOM identified 80 discrete populations of T cells using flow cytometry data (18), 50 from a T cell phenotyping panel (Tphe) and 30 from an intracellular cytokine panel (ICS). For the microbiome data, DADA2 was used to denoise and resolve the 16S rRNA amplicon sequence variants. We compared the effect of PMA and DOL on microbiota and T cell populations at a high level by using multivariate ANOVA to determine the explanatory power of each measure of age across all component microbes or T cell populations (Fig. 1B). Compared to DOL, PMA was superior in predicting T cell, gut and respiratory microbiota composition (R^2_{adj} 0.18, 0.04, 0.08 respectively). Based on these results, we focused on PMA, rather than DOL, as the best predictor of T cell and microbiota maturation.

Anticipating that T cell populations and microbiota composition would show interrelated patterns of variation, we again applied multivariate ANOVA to quantify the amount of total variance the composition of one system could explain in another. All pairs of systems exhibited significant relationships with one another. The R^2_{adj} ranged from 0.05 (nasal microbiota explaining gut microbiota) to 0.15 (nasal microbiota explaining T cells). However, given that all systems exhibited strong associations with PMA, we reasoned that much of the observed effects would be due to the common influence of PMA progression within subjects rather than direct action of one system on the other. Indeed, adjusting for PMA in these models attenuated the variance explained between systems by approximately 50%, though all pairs were still significantly interrelated. These results support PMA as a significant driver of T cell and microbiota maturation, but further suggests a more complicated model in which these systems, albeit to a lesser degree, coordinate independently of host age.

Premature birth transiently alters T cell development

Based on our finding that PMA (a variable that precedes birth and continues postnatally), better explains the temporal progression of T cell populations than DOL, we anticipated that PT and FT infants who were born at different PMA would begin with distinct phenotypes at birth but then converge as they achieved equivalent PMA. Using uniform manifold approximation and projection (UMAP) to visualize a reduced-dimension representation of each sample's combined vector of component T cell population proportions (ICS and Tphe combined), we confirmed that PT and FT T cell phenotype and function clustered separately at birth, began to converge at 40 weeks PMA, and were fully overlapping between GA groups by 12 months (Fig. 2A, 2B).

We next investigated the patterns of individual T cell population abundances over time from birth through one year. Previous studies focusing on a limited number of T cell subsets have suggested a straightforward model in which the fetal T cell developmental program is first biased towards tolerance and hyporesponsiveness, but then acquires, following repeat immune priming postnatally, conventional memory T cells with polarized cytokine functions (19-22). To more comprehensively characterize the T cell populations that follow a PMA-directed program, we examined populations that were selected as statistical predictors of PMA using elasticnet regression (see "*Mistimed Immune and Microbial Development Predict Respiratory Morbidity*"). Tphe populations were grouped according to the following established naming conventions: effector memory (EM, CD45RO+, CCR7-, CD28-), naïve (N, CD45RO-, CCR7+, CD28+), central memory (CM, CD45RO+, CCR7+), virtual memory (Vmem, CD8+, CD45ROlo, CD122hi), terminal effector (TE, CD45ROlo, CCR7-, CD28-), (23-28). ICS populations were first grouped into naïve and memory (CD45RA+ and CD45RA-, respectively), and then named based on predominant cytokine profile. Interestingly, T cells associated with the "youngest" PMA displayed TE marker combinations, including CD8+ T cells positive for cytotoxic granules. Naïve and EM subsets followed TE chronologically, and naïve populations showed considerable heterogeneity across PMA. CM populations were generally associated with older PMA, and several of the CM populations that arose earlier carried a FOXP3+IL7rαlow T-reg cell phenotype (Fig. 2C).

Based on previous studies showing VM T cells arising during rapid homeostatic expansion, we expected to find this population enriched in early PT infants, corresponding to a period of highest growth velocity. Counter to this expectation, VM were enriched at later PMA in FT-born infants. Functionally, CD4+ T cells progressed from naive TNF-α and IL-2 high,

then IL-8 high, then polarized, polyfunctional cells at later PMA. Naïve, cytokine low and TNF-α positive CD8+ T cells were present at low PMA, then progressed through IL-4 and IL-8 positive, then cytotoxic CD45RA+. CD45RA low (memory phenotype) T cells were enriched at later PMA.

Individual T cell subsets with different abundances between PT and FT subjects were most prominent at the birth and discharge timepoints. To isolate the impact of premature birth on T cell development, we performed a multivariate regression based on GA, PMA and interactions thereof, to estimate GA-associated changes in abundance on T cell subsets at 37 weeks PMA. T cell subsets again were grouped based on CCR7, CD45RO, CD45RA, and CD122 expression as described above. Naïve subsets were distributed across gestational ages, though PT naïve subsets were distinct in their lower CD31 expression and IL-8 expression. Memory and effector subsets (CD45RA-, CM, EM, TE) were associated with younger GA, with the exception of VM, which was associated with FT subjects (Fig. 2D). Even within the FOXP3+ populations, there was subtle variation in phenotype between PT and FT, with PT-associated Tregs displaying low CCR7 expression.

Early variation between PT and FT followed by convergence at one year did not necessarily imply that T cell development from early PMA to one year would proceed in a linear fashion. That is to say, the early PMA effector/memory enrichment may not set the PT T cell pool on a linear trajectory across all three sampled timepoints. We therefore sought out cell population subsets with distinctly non-linear patterns of abundance from birth to discharge to one year. Utilizing the three timepoints typically captured for each subject, we identified 10 T cell populations with non-monotone V- or inverted-V trajectories from birth to 12 months in PT samples (Fig. 2E, Supplementary Fig. 1). Most frequently, these population abundances followed

a V-shaped trajectory that decreased sharply from birth (derived from cord blood), to 37 weeks PMA, followed by a slower recovery from 37 weeks to one year. This pattern was seen in several memory CD4+ and CD8+ T cell subsets, indicating a transiently activated T cell phenotype in PT at birth that resolves under more homeostatic conditions. Two CD4+ ICS populations, which were IL-8-positive, had inverted-V trajectories. Together these results reveal a trajectory in which pauci-functional memory, effectors and regulatory T cells are enriched during early development, and these give way to more "typical" naïve populations, followed by a much later gain of fully functional memory T cell subsets.

Inflammatory Exposures Disrupt Typical T Cell Developmental Trajectories

As described, T cell population abundance was robustly associated with PMA. Examining shifts in individual T cell subsets, while informative, may not reveal broader features of the immune system that are affected by either time or exposures. To characterize a more global T cell trajectory during infancy, we partitioned Tphe and ICS samples into immune state types (ISTs) based on the abundances of their respective T cell populations using Dirichlet Multinomial Mixture (DMM) models (Fig 3A-3B). Briefly, this method assumes that there are a relatively small, fixed number of unobserved sample archetypes or typical sample profiles, each distinguished by their characteristic composition, and that each observed sample is an instance of one of the archetypes. The model determines the optimal number of archetypes and their typical composition based on what best fits the data. Each IST, therefore, represented an archetypal profile of T cell composition in terms of the abundance of the various T cell subpopulations, and samples were assigned to the IST which best explains their observed makeup. The seven T cell phenotype immune state types (Tphe ISTs) and eight ICS immune state types (ICS ISTs),

numbered according to their average order of occurrence, exhibited strong associations with PMA (ANOVA, R^2 = 0.86 and 0.69, respectively; Fig. 3C-3D). The phenotypic and functional trajectory revealed by ISTs is consistent with that found using individual T cell populations. With the exception of one 12 month sample in ICS4, ISTs Tphe1-Tphe4 and ICS1-ICS4 were only seen in samples drawn at birth and discharge.

Two Tphe ISTs (Tphe5 and Tphe6) were present in small proportions of samples at all timepoints and GA. Because these state types deviated from the normal IST progression, we hypothesized that their occurrence was associated with clinical exposures common to PT and FT, as opposed to development alone. In support of this, we noted that Tphe5 was most marked by the abundance of Treg subpopulations and atypical early activated (CD31+, CD45RO-, CCR7-, CD28-) CD8+, CD4+ CM subpopulations. Consistent with our hypothesis, chorioamnionitis and/or exposure to antenatal antibiotics raised the odds of a subject ever entering Tphe5 by 7-fold (95% CI 1.0-54, $p<.05$) and 4-fold (95% CI 1.1-13, $p<.03$), respectively, in a joint logistic regression model that adjusted for GA, sex, race, mode of delivery, and premature rupture of membranes (Supplementary Fig. 2). Tphe6 was marked by high abundance of CD57+ and cytotoxic CD8+ T cells, which are associated with T cell exhaustion and chronic viral infection such as CMV, (29, 30). While CMV occurred in less than 7% of Tphe6 negative subjects, 60% of subjects ever entering Tphe6 tested positive for CMV at 6 or 12 months (odds ratio 10.2, $p<.0001$). These results are further evidence that the normal, intrinsic T cell trajectory during infancy is largely determined by PMA, and indicate that deviation from the normal is a result of perturbation to this trajectory by extrinsic clinical factors.

Postmenstrual Age Drives Convergent but Not Identical Microbiota Community Progression In Preterms and Full-Terms

Having characterized the compositional progression of T cell population profiles with respect to PMA, we performed a similar assessment of the microbiota. To summarize and examine the data structure broadly, unweighted Unifrac distances between all samples within each body site were computed as a measure of β-diversity and were used to perform principal coordinate analysis (PCoA). For both body sites, the first principal coordinate (PC1) corresponded to PMA (Supplementary Fig. 3). Samples lower in PC1 tended to be taken prior to 40 weeks PMA, consistent with a unique PT microbiota. Over time, subjects progressed along PC1 and PT and FT subjects converged, exhibiting equal representation on the right side of the PC1 axis. These results establish broad parallels between the development of T cell populations and microbiota with respect to PMA.

To summarize microbiota composition and facilitate subsequent comparative analyses, we applied a similar approach as performed on T cell populations using DMM modeling to partition samples into characteristic community state types (CSTs) based on their compositional profiles. Based on model fit and parsimony, 13 CSTs were defined for both respiratory (nCST) and gut microbiota (gCST) and were enumerated (1-13) according to the average PMA at which samples assigned to each CST were collected. Both gCST and nCST 1 were predominated by *Staphylococcus*, which was replaced over time with more niche-specific taxa in later CSTs, including *Enterobacteriales* and *Clostridiales* in the gut and *Streptococcus* and *Corynebacterium* in the respiratory tract (Fig. 4A-4B). Progression from CST 1 to 13 in the gut and the nose was strongly associated with PMA (ANOVA, R^2 = 0.57 and 0.61, respectively) (Fig. 4C-4D). Several early CSTs with the lowest average PMA were dramatically over-represented by PTs, again

suggesting a unique PT microbiota. Overall though, PT and FT subjects tended to converge to a shared microbiota and most CSTs were represented by equal proportions of FT and PT infants. However, a small number of gCSTs and nCSTs occurring later in the first year of life violated this tendency and were overrepresented by either PT or FT subjects. For example, gCST 9 contained 78% PT samples vs. 22% FT samples ($p<0.05$, two-tailed binomial test) and was notable for diminished levels of *Bifidobacterium* relative to gCSTs 8 and 10, which occurred over similar PMA intervals but which were over-represented by FT subjects ($p<0.01$ and $p<0.05$, respectively). These findings reveal that while the sequence of CST occurrence depends primarily on PMA, maturity at birth biases infants towards or against entering certain CSTs, even months after birth.

T Cells and Microbiota Interact Sparsely After Controlling for PMA

The strong relationship between PMA, T cell and microbiota state type trajectories suggests that the immune system and microbiota are regulated in tandem with the developing infant mucosal ecosystem. We therefore asked if T cell-microbial interactions occur beyond what can be explained by host age and whether such interactions might imperil an infant's health. In order to explore this question, we modeled overall CST *duration*, *occurrence ever*, and *time to first occurrence* as functions of T cell populations and state types at specific time points. In the *duration* model the number of days a subject spent in a given CST, adjusting for the total length of surveillance, was the response. For the *occurrence ever* model the log odds that the CST ever occurred in a subject was the response. Lastly, in the *time to first occurrence* model the response was the DOL of the subject's initial transition into the CST. We modeled each as a function of one of the immunological parameters (IST or T cell population abundance at a particular time

point), and adjusted for GA and mode of delivery. This approach reduced the longitudinal time series down to a sequence of subject-level summaries amenable to typical cross-sectional analyses. We fit models on all pairwise combinations of CSTs and immunological parameters. The significant results of these tests, which corresponded to interactions between the immune system and microbiota present in our cohort, were visualized as networks (Fig. 5A, Supplementary Fig. 4). Among the models of CST duration, of the potential 6318 possible associations between the 26 CSTs and 243 immunological parameters, only 10 Tphe and no ICS metaclusters achieved statistical significance after multiple test correction. CST-associated CD4+ metaclusters preceded, but CD8+ metaclusters followed, the average onset of their associated CST, suggesting a temporal directionality to these relationships. For the models of CST occurrence ever, of the 15 total ISTs, only 3 (Tphe1, Tphe3 and Tphe5) were significantly correlated with a single CST (nCST 8). The most striking finding in the network was that entry into Tphe5 by the time of hospital discharge (n=25 subject-samples) precluded a subject ever entering into nCST 8, which itself was only ever observed after hospital discharge (Fig. 5B).

In the time to first occurrence model, a greater number of gCSTs , CD8+ T cell populations and ICS metaclusters exhibited significant associations than in the duration and occurrence ever models (Supplementary Fig. 4B). Infants who were delayed in their entry into the *Streptococcus*-dominant nCST 4 had higher TNFα or IFNγ+ naïve CD8+ T cell frequencies at discharge and one year. Higher frequencies of effector CD8+ populations were also found in subjects with delayed gCST 9 entry (*Bifidobacterium* and *Bacterioides* low). Alternatively, the occurrence of gCST 3 was accelerated in subjects exhibiting Tphe2 at discharge (Supplementary Fig. 5). gCST 3 was the most diverse and mature gCST prior to discharge, and the earliest gCST in which *Clostridia* are prevalent. Notably, our previous study shows that the early presence of

Clostridia in newborns predicts better growth velocity(6). Overall, the sparsity of these associations underscores the predominant role that host age plays in driving the abundance of T cell populations and microbiota composition. However, significant relationships between T cells and microbiota do occasionally occur even after accounting for the influence of host age, and the sequence of occurrence of associated immune markers and microbial CSTs relative to one another imply bidirectional imprinting between these two systems.

An Early Maladaptive Immune State Type Precludes Allociococcus Colonization, Which Increases Disease Risk

Closer examination of the infrequent but significant T cell-microbiota associations revealed the involvement of an organism with previously described clinically relevant functions. nCST 8, which was common overall but never observed in infants who manifested the Tphe5 immune state at either the birth or discharge sampling timepoints, was dominated by *Allociococcus*. *Allociococcus* was virtually absent in samples collected prior to hospital discharge, but appeared soon after, maintaining stable mean abundance through one year. Previous reports indicate that in children, *Alloiococcus* in the respiratory tract is negatively associated with respiratory infections, while its presence in the ear is positively associated with antibiotic-resistant otitis media (31, 32). Additionally, Tphe5 is notable for its relationship to prenatal inflammatory exposures, being dramatically more prevalent in infants who experienced chorioamnionitis or exposure to antenatal antibiotics, as described above. Because the predominance of *Alloiococcus* is the distinguishing feature of nCST 8, and the occurrence of nCST 8 was precluded by the occurrence of Tphe5 in early life, we sought to assess the relationship between *Alloiococcus* abundance in the nose, acute respiratory illness, and early

immunophenotype, controlling for multiple confounders. To identify episodes of respiratory illness post-NICU discharge, infants were scored by parents using a self-reported modified COAST score when respiratory symptoms arose (33). If a threshold score of 3 was met, an in-person study visit was initiated, during which symptom scores were reviewed, nasal and rectal swabs were obtained, and a physical exam was performed.

As expected, the Tphe5 immunophenotype at birth or discharge was associated with diminished *Alloiococcus* abundance in the nose across all post-discharge timepoints, yielding a 7-fold reduction (3-14 fold, 95% CI, p-value < 0.001; Fig. 5C), while controlling for DOL, GA, mode of delivery, and repeated sampling of subjects. Additionally, we found a 1.4-fold reduction in the odds of a sample being taken during acute illness for every 10% increase in *Alloiococcus* relative abundance (1.1-1.8 fold, 95% CI, p-value < 0.003), controlling for confounders as above. Considering the joint effects of acute illness and Tphe5 occurrence at birth or discharge as predictors in the same model, we found that both were associated with reduced *Alloiococcus* abundance, (log ratios -0.9 ± 0.4 and -1.9 ± 0.8, respectively, 95% CI; p-values < 0.001; Fig. 5D). However, despite negative associations between Tphe5 and *Alloiococcus* abundance, and *Alloiococcus* abundance with illness, Tphe5 was not significant as a predictor of illness, either by itself (log odds = 0.5 ± 0.7, 95% CI, p-value = 0.13) or in conjunction with *Alloiococcus* relative abundance (log odds = -0.4 ± 0.7, 95% CI, p-value = 0.33), controlling for confounders in both cases. Taken together with the immune-CST associations described above, these results show that bidirectional T cell-microbiota relationships occur infrequently, but when present, can be strongly linked with critical health outcomes. Furthermore, the temporal progression from prenatal inflammatory exposure to T cell phenotype to microbiota to clinical outcome suggest that the cascade of events leading to disease states in

infancy is initiated early and involves a complex but observable interplay between exposure, host response and mucosal niche development.

Mistimed Immune and Microbial Development Predict Respiratory Morbidity

Observing that rare T cell-microbiota interactions occurring independently of PMA impacted respiratory morbidity led us to hypothesize that mistimings in development of T cells or microbiota increased the risk of PRD. To test this hypothesis, we developed a quantitative model of the "normal" relationship between PMA and T cell and microbiota composition. We trained two sparse regression models that used the T cell populations and OTU abundance vectors to predict log2-transformed PMA at sample collection. Holding out a subject's longitudinal record, the cross-validated models strongly predicted PMA using either T cell populations (R^2=0.77) or bacterial taxa (R^2=0.65) (Fig. 6A). For each subject, the fitted intercepts of these models, which here represent the predicted PMA at 37 weeks actual PMA, indicate the subject's microbiota and T cell maturity relative to "normal" at 37 weeks PMA (see "Immunological and microbial developmental indices" for details). The fitted slopes of the models indicate a subject's *rate* of microbiota and T cell maturation over the first year, again relative to normal. These four fitted parameters define a *developmental index* (DI) for each subject, which was used to assess mistiming with respect to age, or asynchrony between age, T cell and microbiota development.

We used random forest classification models to compare the predictive power of the DI alone to that of a set of known clinical risk factors for PRD. The clinical features were race, maternal education, sex, GA, birthweight, season at birth and oxygen supplementation integrated over the first 14 days of life (34). The four DI features were the z-scores of the microbiota and T

cell slopes and intercepts. In cross-validation, the clinical features predicted PRD with area under the curve (AUC) of 0.69 (0.59-0.79 95% CI). The features contributing most to the outcome were increased oxygen exposure, lower birthweight and younger GA (Supplementary Fig. 6). Notably, when compared to clinical predictors, the developmental index had statistically equivalent skill in predicting PRD (Fig. 6B, AUC 0.64, 0.54-0.74 95% CI). Combining the clinical features and the developmental variables did not improve the AUC of the predictive model, further evidencing that T cell and microbial development may have durable effects on health outcomes that are equal in their impact to traditional clinical characteristics. Of the four components generating the DI, the microbiome intercept and immune slope had the largest variable importance scores. In exploring the functional relationship between PRD and these factors, we observed that an immature microbiota at term equivalent PMA increased the risk of PRD by over 2-fold, and this effect was magnified in subjects with accelerated T cell maturation (Fig. 6C). These results indicate that the timing of T cell and microbiome maturation *relative to an infant's age* that plays an integral role in promoting or interfering with respiratory health.

Discussion

Birth marks the commencement of a dynamic interplay between innate developmental programming, colonization and assembly of the microbiome, and differentiation and maturation of the adaptive immune system which influences health from infancy through adulthood. In healthy infants, this process balances the accommodation of commensal microbiota, appropriate immune response to pathogens, and functional maturation of the organs at the mucosal interface between human host and environment. Previous studies have generally applied a cross-sectional approach to summarizing microbiota and immune systems in infant, which neglects the rapidly changing infant mucosa as a factor in pathophysiology or homeostasis. By developing new,

longitudinal models of microbiota composition and T cell populations, we were able to establish conceptually and analytically tractable representations of these systems, and to interrogate their maturation and co-development, revealing several key findings. First, T cells and microbiota exhibit structured patterns of progression synchronized by postmenstrual age, with pronounced differences between pre- and full-term infants in very early life and a tendency towards convergence over the first year of life. Furthermore, within the framework of development driven by PMA, interactions occur between T cell population profiles and microbiota community structure. Finally, early atypical or asynchronous immune and microbiota development is a precipitating event in the cascade of disease in infants. To our knowledge, this is the first study to successfully model the influence of this triad of T cell, microbiota, and host development on clinical outcomes in a cohort of both preterm and full-term human infants.

In recent years the concept of a "neonatal window of opportunity" of exposure-mediated immune priming has emerged as a potential causal factor underlying chronic immune-mediated diseases (12). This window of opportunity represents a promising target for clinical intervention and disease prevention. Frequent or severe respiratory infections can result in chronic respiratory insufficiency, and are the leading cause of outpatient visits and hospitalizations in children (35-37). Premature infants, who begin life with diminished respiratory function, have up to a 50% risk for recurrent cough and rehospitalization in the first year, most frequently associated with viral infections. Moreover, chronic respiratory morbidity is associated with dysregulated or poorly targeted immune system activation – particularly T cell activation – brought on by viral or bacterial exposures during infancy (38-40), and understanding the earliest immune- and microbial-related events is essential to interrupting a pathologic program. We therefore focused on respiratory morbidity as a useful and highly relevant outcome to study the

relationship between microbiota, T cell and host development in our birth cohort. By defining developmental indices based on microbiota and T cell populations, we establish that maturity at term and rate of maturation over the first year are indicators of respiratory disease outcome at one year (PRD). Specifically, precocious immune development in conjunction with an immature microbiome at term corresponds to substantially elevated risk of PRD, while either one of these factors by themselves has an attenuated effect. This indicates that mistimed or discordant maturation between the microbiome and immune system is a correlate of respiratory morbidity. Previous reports have used age, microbiota, or immune variables as independent factors in predicting respiratory outcome (2, 5, 25, 37). These studies do not address the possibility that a newborn's immune system is not simply deficient, but rather under normal developmental conditions, is uniquely balanced to provide protection against novel pathogens while minimizing immunopathology. Exposures that accelerate or delay the normal maturation of T cells and microbiota during infancy, such as in utero infection promoting the early occurrence of Tphe5, may disrupt this age-specific balance that has served human evolution so well.

The ability to predict a subject's PMA based on their T cell phenotype is strong evidence that developmental state is a key driving factor in immune maturation, which is further reinforced by convergence of PT and FT phenotypes over time. Characterizing a T cell trajectory during infancy revealed greater heterogeneity than has been appreciated, especially between PT and FT subjects and even within their predominantly naïve T cell pool. As an example, early PT and FT ISTs were both enriched for naïve T cell subpopulations characterized by high CCR7 and CD28 and low CD45RO expression. Unique subpopulations within the naïve T cell pool, however, were distributed across the GA and PMA spectrum. For example, PT and FT naïve T cells showed differential abundance of CD31, IL-7rα and CXCR5, implicating age-dependent

cell survival and provision of B-cell help. Our results were also aligned with previous studies demonstrating that PT infants have higher proportions of CD45RO+ T cells in their cord blood (41, 42). CD45RO+ CD4+ and CD8+ T cells associated with a younger PMA fell within an effector, and potentially short-lived, phenotype, which emerging evidence from animal studies also supports. Many of the PT-enriched CD4+ CD45RO+ subpopulations also were of a Treg phenotype. This result is consistent with prior studies showing that activated fetal naive T cells have a propensity towards Treg differentiation, though co-enrichment for a Treg memory phenotype in early gestation has not previously been shown (43, 44).

Trajectories of individual T cell populations can be informative, but measured together in state types enables more nuanced biological interpretations. Appearance of effector CD4+ and CD8+ T cells at early PMA and PT in Tphe1 and Tphe3, for example, can be seen in inflammatory states, but Tphe1 and Tphe3 effector expansion is also accompanied by expansion of Tregs, which can counter T cell-mediated inflammation. Whether or not Tregs derived in earlier gestation harbor full suppressor function was not addressed in our study, but prior studies have proposed that immune suppression by Tregs contributes to a PT susceptibility to infection (45). The co-expansion of effectors and Tregs in the same IST suggests a state of dysregulation rather than immune suppression, and interestingly, the CCR7-FOXP3+ CD4+ subpopulation associated with Tphe1 arise in inflammatory states and contribute to immune dysregulation in CCR7 null mice (46). In contrast, the higher abundance of virtual memory CD8+ T cells but lower CD45RO+ T cells in FT-associated IST's suggests T cells may acquire a memory phenotype through homeostatic expansion rather than inflammatory stimuli (47-49).

By applying longitudinal models, we gain further insight into the transient nature of many activated T cell populations. For example, we found a direct correlation between GA and IL-8+

T cells at birth, but an indirect correlation with PMA postnatally. This finding, which is consistent with our previous research, is notable in that the recent study by Olin et al. shows enhanced plasma IL-8 in PT when compared to FT (4, 50). Our unique focus on T cells specifically, rather than secreted mediators in plasma, sheds light on a non-linear T cell-specific functional trajectory during pre- and postnatal development that may be distinct from the innate compartment of the immune system . Indeed, an apparent deviation from a typical longitudinal trajectory, as was seen in Tphe5 and Tphe6, was partially attributed to the presence of select exposures (antenatal antibiotics and CMV, respectively) that more durably shape an individual's immune trajectory independent of development.

Together, the microbiome of the airway and gastrointestinal tract, and their interaction with one another and with the host, constitute the gut-lung axis, a system increasingly implicated in immune development and health outcomes such as respiratory morbidity (51, 52). Notably, while the relationship between immune development and the gut microbiome has featured prominently in the literature, we identified more frequent and stronger associations between T cell populations and nasal microbiota when controlling for age. This may be in part due to the central role that intestinal mucosa plays in coordinating the gut microbiome and immune cells (53); while the niche matures, PMA is likely to be an overwhelming factor shaping microbiota and immunity. The respiratory niche is arguably the more relevant site for studying respiratory outcomes. In support of this assertion is a recent study showing that nasal colonization by *Veillonella* and *Prevotella* in infancy alters the nasal immune secretome, which predicts asthma outcome at 6 years (54). Our data did not reveal a correlation between *Veillonella* or *Prevotella*-dominant CSTs and PRD, but *Veillonella* is lowly abundant in the protective nCST8.

One of the outstanding questions from this and other studies linking microbiota to outcome, is whether or not there are modifiable events that precede either harmful or protective microbiota colonization. Using a longitudinal, systems-based approach, our results can offer some insight. *Alloiococcus* is a common post-discharge colonizer, but in our cohort, antenatal exposure to inflammation or antibiotics was associated with early Tphe5 expression (at birth or discharge), which then precluded subsequent *Alloiococcus*-dominant nCST8 in the nose. The additional observation that *Alloiococcus* is substantially diminished during acute respiratory illness reveals a previously undescribed sequence initiated by perinatal exposures, impacting T cell development at birth, then post-discharge airway colonization, and ultimately susceptibility to respiratory infection throughout the first year of life. While these antenatal exposures are not readily modified, there may be some benefit to targeting probiotic treatments in chorioamnionitis-exposed mothers and neonates. In fact, probiotics have been successful in several studies in the prevention of illnesses including reductions in pediatric upper respiratory tract infections (55-57). On the other hand, a more conservative, informed approach to immune- or microbiota-based therapy in infants is also called for by a recent report showing that probiotic treatment in healthy newborns had only a transient effect on stool microbiota, and was associated with an increased risk of enteric and respiratory infections. Evidence from our study also indicates that an infant's health is influenced by timely, synchronous development of microbiota, T cells and the infant, which underscores the need to tread cautiously when considering interventions that may disrupt this normal balance.

Our results show that substantial changes in the immune system occur between NICU discharge and 12 months, but without the benefit of intensive interim sampling, it is difficult to comprehensively account for all factors occurring between timepoints that may shape an

individual's immune trajectory. However, the detailed characterization of relationships between T cell populations and microbiota, and the demonstrated associations between the development of these systems and infant health, represent novel insights into the clinical relevance of microbiome-immune co-development and will inform causal models and mechanistic hypotheses that can be used to develop novel interventions and guide treatment decisions by furthering understanding of the gut-lung axis and the neonatal window of opportunity.

Acknowledgments:

Authors would like to acknowledge the University of Rochester Pediatrics Translational Biospecimen Laboratory, the University of Rochester Genomics Research Core, the Flow Cytometry Core, and the Human Immunology Center. We also thank Richard Simon and Kimberly Baldo (The Harley School), for their assistance with figure preparation and graphic design. This study was dependent on the nurses and staff at the Golisano Children's Hospital NICU and URMC Strong Beginnings Maternity Services, and most of all to our families who generously consent to research studies. Funding provided by NIH NIAID HHSN272201200005C (Respiratory Pathogens Research Center), NIH NIAID 1K08AI108870-01A1 (CD8 T Cell Dysregulation in Premature Infants), NIH NHLBI U01 HL101813-01 (Prematurity and Respiratory Outcomes Program), NIH NCATS UL1 TR000042 (Clinical and Translational Science Institute).

Author Contributions:

Conception, execution, interpretation and manuscript preparation: KMS, AM, AG, NL, GSP, MC, AD

Project PI's, University of Rochester: GSP, MC, SG, AF, DJT

Acquisition and analysis of experimental data: NL, AG, AG, HK, JC, AG, KMS

Clinical and sample data collection, study coordination and recruitment: HH, GSP, MC, KMS

Biostatistical analysis: AM, AG

Clinical and laboratory data integration and management: JHW, SB

Data and Materials Availability:

Annotated datasets for 16S sequencing and flow cytometry results can be found in dbGaP, accession number phs001347. Code supporting this paper is available at https://github.com/amcdavid/CoordinatedTCellsMicrobiota.

Declaration of Interests:

The authors declare no competing interests.

Materials and Methods

Study Design

All study procedures were approved by the University of Rochester Medical Center (URMC) Internal Review Board (IRB) (Protocol # RPRC00045470 & 37933) and all subjects' caregivers provided informed consent. The infants included in the study were enrolled within 7 days of life for the University of Rochester Respiratory Pathogens Research Center PRISM and were cared for in the URMC Golisano Children's Hospital. Clinical data including nutrition, respiratory support, respiratory symptoms, medications, comorbidities, were entered into REDCap (58, 59), then integrated with laboratory results using the URMC Bio Lab Informatics Server, a web-based data management system using the open source LabKey Server (60). Blood was collected at birth, time of NICU discharge or 36-42 weeks PMA (whichever occurred first), and at 12 months of life. We collected 2729 gut (842 from NICU and 1887 post-discharge), and 2210 nasal (619 from NICU and 1591 post-discharge) usable microbiota samples longitudinally from 139 pre-term and 98 full-term infants and worked with the most extensive subset of these possible depending on the analysis in question (Supplementary Table 2). From the PRISM study cohort, fecal (rectal) and nasal material was collected from pre-term infants (23 to 37 weeks gestational age at birth (GA)) weekly from the first week of life until hospital discharge, and then monthly through one year of gestationally corrected age. Rectal and nasal samples were collected from full-term infants at enrollment and monthly through one year. Additionally, rectal and nasal samples were collected from all infants whenever they exhibited symptoms of acute respiratory illness after discharge from the hospital. Symptoms of acute respiratory illness prompting sample collection were summarized by the primary caregiver using a symptom

COAST (Childhood Origins of Asthma) score sheet (35). Parents were instructed to notify the study team if the infant had symptom score of three or greater. Among subjects who completed study procedures through 12 months, 52 PT subjects (43%) and 17 FT subjects (21%) met the criteria for PRD. All blood samples generating usable data were included in all analyses. Sufficient blood to perform T cell phenotyping by flow cytometry at three pre-defined timepoints was collected from 55% of subjects at birth, 61% of subjects at NICU discharge, and 38% at 12 months. For training the PMA predictions models (described below), all microbiota samples were used. For all other analyses, microbiota samples from subjects that did not have any usable data from blood were excluded. Two staining panels, covering i) intracellular cytokine production (ICS) and ii) T cell surface phenotyping (Tphe) were designed (Supplementary Table 3). Complete immunophenotyping for all three timepoints was performed on 25% of subjects, and 63% of subjects had complete immunophenotyping for at least one timepoint.

Flow Cytometry Methods

Sample collection, isolation, storage, thawing, stimulation and staining for flow cytometry was performed as detailed previously (61). In short, cord blood and peripheral blood mononuclear cells were isolated via Ficoll centrifugation, cryopreserved and stored in liquid nitrogen, and rapidly thawed and washed with pre-warmed RPMI-1640 (10% FBS and 1x L-glutamine); thawing was done in ‘subject-balanced’ batches (equal mix of pre and full-term subjects, each with three time points) and an aliquot of each freshly thawed sample was plated and stained with a T-cell phenotyping (‘Tphe’) panel with the remainder of the sample rested overnight in an incubator, plated and stimulated with *Staphylococcus aureus*, Enterotoxin Type B (SEB), and stained with a T-cell functional panel (‘ICS’). Panel compositions are as shown in Supplementary Table 3.

Samples were acquired on a BD LSRII (core facility instrument QC-ed daily with BD CS&T beads); PMT voltages normalized per run to pre-determined/optimized 'Peak-6' (Spherotech) median fluorescence values. R-based packages and scripts were used for all post-acquisition processing and analysis. Reading of raw .fcs files, compensation, transformation, and subsetting/writing of .fcs files was performed using flowCore (62). To minimize inter-run variation associated with the Tphe panel, the flowStats (63) warpSet function was used to normalize arcsinh transformed channel data using a healthy donor adult PBMC control as reference. For analysis with the clustering algorithm FlowSOM, an iterative approach was used for both panels to first cluster on live, intact, lymphoid-sized CD4+ and CD8+ T-cell subsets (in the case of the ICS panel, including activated, CD69+ subsets); those subsets were then re-clustered to capture rare populations and optimally resolve phenotypic heterogeneity and associated function. Over-clustering followed by expert-guided merging was favored when defining the final number of T cell populations. FlowSOM clustering results used in downstream analysis were represented as proportion of the respective T-cell subset, per sample. All scripts, including Tphe arcsinh cofactors, warpSet and FlowSOM parameters, and final clustering counts are available in Supplementary R-Code.

Microbiota Identification

Microbiota sample collection and storage techniques, genomic DNA extraction and background control methods were as previously published (64). Raw data from the Illumina MiSeq was first converted into FASTQ format 2□×□312 paired-end sequence files using the bcl2fastq program (v1.8.4) provided by Illumina. Format conversion was performed without de-multiplexing, and the EAMMS algorithm was disabled. All other settings were default. Samples were multiplexed using a configuration described previously (65). The *extract_barcodes*.py

script from QIIME (v1.9.1) (66) was used to split read and barcode sequences into separate files suitable for import into QIIME 2 (v2018.11) (67) which was used to perform all subsequent read processing and characterization of sample composition. Reads were demultiplexed requiring exact barcode matches, and 16S primers were removed allowing 20% mismatches and requiring a matching window of at least 18 bases. Cleaning, joining, and denoising were performed using DADA2 (68): reads were truncated (forward reads to 260 bps and reverse reads to 240 bps for rectal V3-V4 samples and forward reads to 275 bps and reverse reads to 260 bps for nasal V1-V3 samples), error profiles were learned with a sample of one million reads per sequencing run, and a maximum expected error of two was allowed. Taxonomic classification was performed with naïve Bayesian classifiers trained on target-region specific subsets of the August, 2013 release of GreenGenes (69). Sequence variants that failed to classify to the phylum level or deeper were discarded. Sequencing variants observed fewer than ten times total, or in only one sample, were discarded. Rectal samples with fewer than 2250 reads and nasal samples with fewer than 1200 reads were discarded. Phylogenetic trees were constructed for each body site using MAFFT (70) for sequence alignment and FastTree (71) for tree construction. For the purposes of β-diversity analysis, rectal and nasal samples were rarefied to depths of 2250 and 1200 reads, respectively, and the Unweighted Unifrac (72) metric was applied.

Statistical analyses

Multivariate ANOVA. We used a sequence of multivariate ANOVA (MANOVA) models to estimate the amount of variance one set of variables could explain in another. We modeled T cell population relative abundances, gut, and nasal species-level relative abundances pairwise each as predictor and response matrices. DOL and PMA served only as predictors. Each pair of variables types was joined, with missing samples deleted casewise. T cell populations and

microbiome taxa with a variance of less than .0001 were removed in each comparison. The remaining variables were renormalized to sum to one, and transformed using the isometric log ratio, then modeled using a multivariate linear model with a matrix response. R^2_{adj} was calculated as $1 - \text{MSE}_{\text{full}}/\text{MSE}_{\text{reduced}}$ where the mean squared error (MSE) was the total sum of squared residuals in the response matrix, divided by the residual degrees of freedom, thus approximately unbiased for the residual variance. The PMA-adjusted model used PMA, and the set variables of interest as a predictor in the full model, retaining only PMA in the reduced model. Wilks' lambda was used to test for association between response and predictor variables.

T cell PMA- and GA- associated trajectories. For each T cell subpopulation, a linear regression was fit of GA and PMA on the arcsin-sqrt transformed relative abundance of that population using a continuous and piecewise linear model with a single knot at 37 weeks PMA and interaction with GA (Fig. 2D-2E and Supplementary Fig. 1). In symbols, we fit the model

$$\text{asin}\left(\sqrt{y}\right) \sim \left[p1_{p<0}\right] + \left[p1_{p\geq 0}\right] + \text{GA}' + \text{GA}' \times \left[p1_{p<0}\right] + \text{GA}' \times \left[p1_{p\geq 0}\right] + (1|\text{Subject}),$$

where y is the relative abundance of T cell population (relative to other populations that share same CD4 vs CD8 status and Tphe vs ICS), $p = \text{PMA} - 37$, $\text{GA}' = \text{GA} - 37$ are the PMA and GA of a sample, shifted so that term-equivalent samples and gestational ages are zero, and (1|Subject) is a random intercept for each subject. The intercept of this model represents the abundance in subjects of 37 weeks GA at birth, and the remaining terms are identified by interpolation and extrapolation of the time points actually sampled in an individual. Figure 2D plots the GA' term and its 95% CI for T cell populations with significant (Bonferroni-adjusted p<.05, 80 tests) GA' effects. Non-monotone populations were determined by testing three contrasts i) the $\left[p1_{p<0}\right]$ terms, ii) the $\left[p1_{p\geq 0}\right]$ terms, and iii) the difference between them for joint statistical significance (Bonferroni-adjusted p<.05, 80 tests).

CST and IST Assembly. Microbial community state types (CSTs) were defined for each body site by fitting Dirichlet multinomial mixture (DMM) models (73) using the R package DirichletMultinomial (v1.22.0) (74, 75), R version 3.5.0. Sample composition was represented using normalized counts of the most specific operational taxonomic units (OTUs) present in at least 5% of the samples from a given body site. Normalization was performed on a per sample basis by taking the relative abundance of each OTU and multiplying by 2250 for rectal samples and 1200 for nasal samples. Resulting non-integer counts were rounded down. For each body site, the DMM model was fit with one through twenty Dirichlet components and the optimal number of components was selected by minimizing the Laplace approximation of the negative-log model evidence. In this model, CSTs are synonymous with Dirichlet components, and each sample was assigned to the CST from which it had the highest posterior probability of be derived. This procedure was repeated with the immunological data in order to define immune state types (ISTs), using relative abundances of FlowSOM defined T cell populations in the place of OTUs. Relative abundances were computed within assays (TPHE and ICS) and major populations (CD4 and CD8) separately, and converted to counts by multiplying by 50,000 and rounding down. CD4 and CD8 counts were combined to fit the DMM for each assay.

Microbiota-T cell Associations. Associations between microbiome development and the immune system were modeled using microbiome CST occurrence patterns as outcome variables and iterating through the relative abundances of each FlowSOM T cell population or observed IST at each time point as predictors. In symbols, we used the model

CST ~ immune_parameter + MOD + GA + (sampling_intensity).

For each CST, each of these immunological parameters (T cell population relative abundances and IST, hereafter referred to as the immunological variables of interest [VOIs]) at

each of the three time points when the immune system was sampled (birth, discharge, and one year) was assessed independently.

CST occurrence patterns were related to immunological VOIs by testing three types of associations between every CST-VOI combination at the level of individual subjects, while controlling for mode of delivery (MOD), gestational age at birth (GA) and, in model (i) only (see below), the number of microbiome samples (sampling_intensity) that were collected from an individual. These models differed in the aspect of CST occurrence that was modeled as the outcome. Model (i) tests associations between the VOI and whether or not a CST occurs at all in an individual; (ii) tests associations between the VOI and how persistent a CST is in an individual; and (iii) tests associations between the VOI and the days to first occurrence of a CST in an individual. Model (i) was tested using logistic regression with VOI, MOD, GA and the number of microbiome observations from a given individual as the sampling intensity. The outcome indicated whether or not a given CST was ever observed in the individual. We tested the VOI association by dropping that term and calculating a likelihood ratio test. Model (ii) was tested using a quasi-Poisson regression model with MOD, GA, and the VOI as covariates, and total number of days the subject was assigned to *any* CST as an offset. The number of days a subject was assigned to a given CST was the outcome and was calculated by summing the interval lengths between CST change points. Intervals were calculated from midpoint to midpoint on the sampled days of life. At birth, subjects were placed in the first observed CST if the first sample occurred within 14 days of life, otherwise the first interval was excluded. Subjects were assumed to remain in their final observed CST for an interval equal to half the interval length between the penultimate and ultimate sample. Significance of the VOI was assessed as in model (i). Model (iii) was tested using interval censored, accelerated log logistic

failure time models (R package icenReg v2.0.9) (76) with MOD, GA, and the VOI as covariates and the interval preceding the first observation of a given CST as the outcome. For gCST 1 and nCST 1, which on average were the earliest CSTs, we modeled the interval preceding the first observation of a CST other than NAS 1 or REC 1. For each CST, only subjects that were ever observed in that CST at some point were included. Significance was assessed based on Wald test p-values of the terms in the fitted full models.

For models (i)-(iii), subjects with fewer than one sample taken per 30 NICU-days or fewer than six samples post discharge were excluded. We filtered immune VOI with fewer than ten observations, and CSTs present in fewer than 10% of the remaining observations. Numerical covariates were converted into z-scores, except GA which we modeled as $(\mathrm{GA} - 37)/37$. Within each model (i)-(iii), multiple testing across all CSTs and VOIs was corrected for using the Benjamini-Hochberg method at 10% FDR.

Tphe5, Alloiococcus abundance, and acute illness associations. Using only post-discharge nasal samples, the abundance of *Alloiococcus* represented as read counts was modeled as a function of day of life, GA, MOD, and the occurrence of Tphe5 at birth or discharge using a generalized estimating equation fit with the geeglm function in R (77). Subject was used as the clustering variable, an exchangeable working correlation structure was specified, total reads per sample was used as an offset, and the family was Poisson with a log link function. This model was repeated with the addition of acute illness as a covariate. The probability of a sample coming from an illness or healthy surveillance visit was modeled using mixed effects logistic regression fit with the glmer function (78), using *Alloiococcus* relative abundance, DOL, GA, and MOD as covariates, with Subject as a random effect. This model was repeated with the addition of Tphe5 at birth or discharge as a covariate.

Prediction of PMA. Two separate elastic net regression models (79) were trained to predict (80) the log2-transformed PMA with a) T cell immunological features and b) microbial abundance. In (a) the four feature sets were CD4 ICS, CD8 ICS, CD4 Tphe and CD8 Tphe populations, while in (b) the two feature sets consisted of nasal and rectal species-level relative abundances from samples collected prior to DOL 450, filtered to remove taxa present in fewer than 3% of samples. A total of 433 samples from 185 subjects and 80 features were included in (a). Model (b) was trained on 3032 samples from 237 subjects and 218 features. Some samples had incomplete feature sets, e.g., if only the ICS panel was run then both the CD4 and CD8 Tphe sets were missing, or if only the nasal microbiome was sampled and the rectal abundances were missing. We treated this as a missing data problem, and imputed the values with their mean values among non-missing cases. Imputation was chained onto the elasticnet model (occurred only using the training data, in each fold) for the purposes of tuning and validation. Within each feature set, we used the relative proportions, transformed into z-scores.

Cross validation for tuning and prediction. We tuned the model and estimated its performance using cross-validation by holding out a subject's entire longitudinal record. We tuned the elastic net alpha in [0, 1] and lambda in [.001, .5] parameters by randomly selecting 50 combinations of (alpha, lambda) and evaluating the test mean-square error (MSE) via 5-fold cross-validation. After finding a minimizing pair of (alpha, lambda), the model was refit with 10-fold cross-validation. For each subject i, this provides two sequences of fitted values, representing the log2-transformed PMA prediction. For instance, for the microbiome, we have

$$\widehat{Y_{ij}} = f^{-i}(x_{ij}), j = 1, \dots, n_i,$$

where x_{ij} represent microbial feature vectors, n_i indexes the number of longitudinal samples for subject *i*, and f^{-i} represent the elastic net model trained excluding subject i. For the

T cell immunome, the analogous model is fit. The back-transformed values $2^{\widehat{Y_{ij}}}$ were used to calculate each model's R^2.

Immunological and microbial developmental indices. The longitudinal sequence of cross-validated fitted values $\widehat{Y_{ij}}$ were compared to the true PMA for each subject using a linear mixed model. We fit the model

$$\hat{Y} - \log_2(37) \sim \log_2(\mathrm{PMA}/37) + (1 + \log_2(\mathrm{PMA}/37)|\mathrm{Subject})$$

thus $\widehat{Y_{ij}} = \alpha_i + \beta_i \times \mathrm{PMA}_{ij} + \epsilon_{ij}$ and calculated the best linear unbiased predictor of each subject's 37-week intercept α_i, slope β_i and their conditional standard errors se(α_i), se(β_i). These are transformed into a quantity similar to a z-score by subtracting the median of α_i, β_i over subjects i, and dividing by its conditional standard error se(α_i) or se(β_i).

Prediction of PRD. We used random forest classification models to predict PRD using two feature sets: clinical and developmental index. The clinical features were race, maternal education, the baby's sex, gestational age, weight and season at birth, and oxygen supplementation integrated over the first 14 days of life. The developmental index features were the z-scores of the microbiome and T-immune slopes and intercepts. The random forest hyperparameters mtry, ntree and nodesize were tuned separately for each feature set with random search using 5-fold cross-validation. After the optimal parameters were found for each feature set, a second round of 20-fold cross validation was used to evaluate the area under the ROC curve (AUC). The fitted values from the random forest regression were calculated using the function generatePartialDependenceData.

References:

1. Dimmitt RA, Staley EM, Chuang G, Tanner SM, Soltau TD, Lorenz RG. Role of postnatal acquisition of the intestinal microbiome in the early development of immune function. J Pediatr Gastroenterol Nutr. 2010;51(3):262-73. Epub 2010/07/20. doi: 10.1097/MPG.0b013e3181e1a114. PubMed PMID: 20639773; PMCID: 2932839.
2. Rechavi E, Lev A, Lee YN, Simon AJ, Yinon Y, Lipitz S, Amariglio N, Weisz B, Notarangelo LD, Somech R. Timely and spatially regulated maturation of B and T cell repertoire during human fetal development. Science translational medicine. 2015;7(276):276ra25. doi: 10.1126/scitranslmed.aaa0072. PubMed PMID: 25717098.
3. Lee AH, Shannon CP, Amenyogbe N, Bennike TB, Diray-Arce J, Idoko OT, Gill EE, Ben-Othman R, Pomat WS, van Haren SD, Cao KL, Cox M, Darboe A, Falsafi R, Ferrari D, Harbeson DJ, He D, Bing C, Hinshaw SJ, Ndure J, Njie-Jobe J, Pettengill MA, Richmond PC, Ford R, Saleu G, Masiria G, Matlam JP, Kirarock W, Roberts E, Malek M, Sanchez-Schmitz G, Singh A, Angelidou A, Smolen KK, Consortium E, Brinkman RR, Ozonoff A, Hancock REW, van den Biggelaar AHJ, Steen H, Tebbutt SJ, Kampmann B, Levy O, Kollmann TR. Dynamic molecular changes during the first week of human life follow a robust developmental trajectory. Nat Commun. 2019;10(1):1092. Epub 2019/03/14. doi: 10.1038/s41467-019-08794-x. PubMed PMID: 30862783; PMCID: PMC6414553.
4. Olin A, Henckel E, Chen Y, Lakshmikanth T, Pou C, Mikes J, Gustafsson A, Bernhardsson AK, Zhang C, Bohlin K, Brodin P. Stereotypic Immune System Development in Newborn Children. Cell. 2018;174(5):1277-92 e14. Epub 2018/08/25. doi: 10.1016/j.cell.2018.06.045. PubMed PMID: 30142345; PMCID: PMC6108833.
5. Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, Ross MC, Lloyd RE, Doddapaneni H, Metcalf GA, Muzny D, Gibbs RA, Vatanen T, Huttenhower C, Xavier RJ, Rewers M, Hagopian W, Toppari J, Ziegler AG, She JX, Akolkar B, Lernmark A, Hyoty H, Vehik K, Krischer JP, Petrosino JF. Temporal development of the gut microbiome in early childhood from the TEDDY study. Nature. 2018;562(7728):583-8. Epub 2018/10/26. doi: 10.1038/s41586-018-0617-x. PubMed PMID: 30356187; PMCID: PMC6415775.
6. Grier A, Qiu X, Bandyopadhyay S, Holden-Wiltse J, Kessler HA, Gill AL, Hamilton B, Huyck H, Misra S, Mariani TJ, Ryan RM, Scholer L, Scheible KM, Lee YH, Caserta MT, Pryhuber GS, Gill SR. Impact of prematurity and nutrition on the developing gut microbiome and preterm infant growth. Microbiome. 2017;5(1):158. Epub 2017/12/13. doi: 10.1186/s40168-017-0377-0. PubMed PMID: 29228972; PMCID: PMC5725645.
7. Bisgaard H, Li N, Bonnelykke K, Chawes BL, Skov T, Paludan-Muller G, Stokholm J, Smith B, Krogfelt KA. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. J Allergy Clin Immunol. 2011;128(3):646-52 e1-5. Epub 2011/07/26. doi: 10.1016/j.jaci.2011.04.060. PubMed PMID: 21782228.
8. Smith MI, Yatsunenko T, Manary MJ, Trehan I, Mkakosya R, Cheng J, Kau AL, Rich SS, Concannon P, Mychaleckyj JC, Liu J, Houpt E, Li JV, Holmes E, Nicholson J, Knights D, Ursell LK, Knight R, Gordon JI. Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. Science. 2013;339(6119):548-54. Epub 2013/02/01. doi: 10.1126/science.1229000. PubMed PMID: 23363771; PMCID: PMC3667500.
9. Bosch A, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Biesbroek G, Kool J, Pernet P, de Groot PCM, Eijkemans MJC, Keijser BJF, Sanders EAM, Bogaert D. Maturation of the Infant Respiratory Microbiota, Environmental Drivers, and Health Consequences. A Prospective Cohort Study. Am J Respir Crit Care Med. 2017;196(12):1582-90. Epub 2017/07/01. doi: 10.1164/rccm.201703-0554OC. PubMed PMID: 28665684.
10. Korten I, Mika M, Klenja S, Kieninger E, Mack I, Barbani MT, Gorgievski M, Frey U, Hilty M, Latzin P. Interactions of Respiratory Viruses and the Nasal Microbiota during the First Year of Life in Healthy Infants. mSphere. 2016;1(6). Epub 2016/12/03. doi: 10.1128/mSphere.00312-16. PubMed PMID: 27904883; PMCID: PMC5120172.
11. Man WH, van Houten MA, Merelle ME, Vlieger AM, Chu M, Jansen NJG, Sanders EAM, Bogaert D. Bacterial and viral respiratory tract microbiota and host characteristics in children with lower respiratory tract infections: a matched case-control study. Lancet Respir Med. 2019. Epub 2019/03/20. doi: 10.1016/S2213-2600(18)30449-1. PubMed PMID: 30885620.

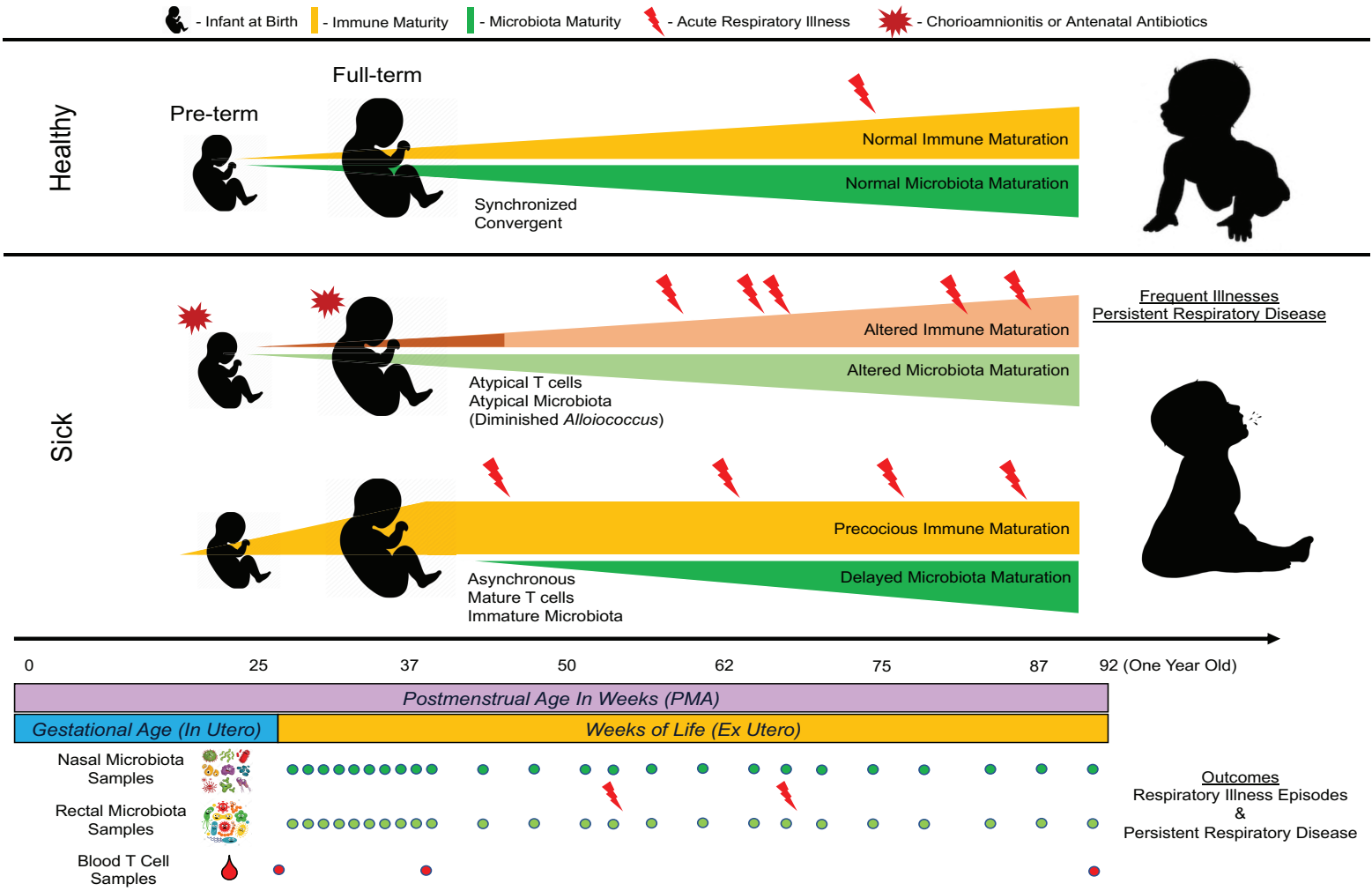
12. Torow N, Hornef MW. The Neonatal Window of Opportunity: Setting the Stage for Life-Long Host-Microbial Interaction and Immune Homeostasis. J Immunol. 2017;198(2):557-63. Epub 2017/01/11. doi: 10.4049/jimmunol.1601253. PubMed PMID: 28069750.
13. Reardon S. Microbiome therapy gains market traction. Nature. 2014;509(7500):269-70. Epub 2014/05/16. doi: 10.1038/509269a. PubMed PMID: 24828169.
14. McQuade JL, Daniel CR, Helmink BA, Wargo JA. Modulating the microbiome to improve therapeutic response in cancer. Lancet Oncol. 2019;20(2):e77-e91. Epub 2019/02/05. doi: 10.1016/S1470-2045(18)30952-5. PubMed PMID: 30712808.
15. Uchiyama K, Naito Y, Takagi T. Intestinal microbiome as a novel therapeutic target for local and systemic inflammation. Pharmacol Ther. 2019;199:164-72. Epub 2019/03/17. doi: 10.1016/j.pharmthera.2019.03.006. PubMed PMID: 30877020.
16. Watkins C, Stanton C, Ryan CA, Ross RP. Microbial Therapeutics Designed for Infant Health. Front Nutr. 2017;4:48. Epub 2017/11/11. doi: 10.3389/fnut.2017.00048. PubMed PMID: 29124056; PMCID: PMC5662644.
17. Pryhuber GS, Maitre NL, Ballard RA, Cifelli D, Davis SD, Ellenberg JH, Greenberg JM, Kemp J, Mariani TJ, Panitch H, Ren C, Shaw P, Taussig LM, Hamvas A, Prematurity, Respiratory Outcomes Program I. Prematurity and respiratory outcomes program (PROP): study protocol of a prospective multicenter study of respiratory outcomes of preterm infants in the United States. BMC Pediatr. 2015;15:37. Epub 2015/04/18. doi: 10.1186/s12887-015-0346-3. PubMed PMID: 25886363; PMCID: PMC4407843.
18. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, Saeys Y. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytometry A. 2015;87(7):636-45. Epub 2015/01/13. doi: 10.1002/cyto.a.22625. PubMed PMID: 25573116.
19. Marchant A, Goldman M. T cell-mediated immune responses in human newborns: ready to learn? Clin Exp Immunol. 2005;141(1):10-8. Epub 2005/06/17. doi: 10.1111/j.1365-2249.2005.02799.x. PubMed PMID: 15958064; PMCID: PMC1809424.
20. Schmiedeberg K, Krause H, Rohl FW, Hartig R, Jorch G, Brunner-Weinzierl MC. T Cells of Infants Are Mature, but Hyporeactive Due to Limited Ca2+ Influx. PLoS One. 2016;11(11):e0166633. Epub 2016/11/29. doi: 10.1371/journal.pone.0166633. PubMed PMID: 27893767; PMCID: PMC5125607.
21. Debock I, Flamand V. Unbalanced Neonatal CD4(+) T-Cell Immunity. Front Immunol. 2014;5:393. Epub 2014/09/16. doi: 10.3389/fimmu.2014.00393. PubMed PMID: 25221551; PMCID: PMC4145351.
22. Saso A, Kampmann B. Vaccine responses in newborns. Semin Immunopathol. 2017;39(6):627-42. Epub 2017/11/11. doi: 10.1007/s00281-017-0654-9. PubMed PMID: 29124321; PMCID: PMC5711983.
23. Bains I, Antia R, Callard R, Yates AJ. Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. Blood. 2009;113(22):5480-7. Epub 2009/01/31. doi: 10.1182/blood-2008-10-184184. PubMed PMID: 19179300; PMCID: PMC2689049.
24. De Rosa SC, Herzenberg LA, Herzenberg LA, Roederer M. 11-color, 13-parameter flow cytometry: identification of human naive T cells by phenotype, function, and T-cell receptor diversity. Nat Med. 2001;7(2):245-8. Epub 2001/02/15. doi: 10.1038/84701. PubMed PMID: 11175858.
25. Gerlach C, Rohr JC, Perie L, van Rooij N, van Heijst JW, Velds A, Urbanus J, Naik SH, Jacobs H, Beltman JB, de Boer RJ, Schumacher TN. Heterogeneous differentiation patterns of individual CD8+ T cells. Science. 2013;340(6132):635-9. Epub 2013/03/16. doi: 10.1126/science.1235487. PubMed PMID: 23493421.
26. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ, Teichmann SA, Marioni JC, Stegle O. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol. 2015;33(2):155-60. Epub 2015/01/20. doi: 10.1038/nbt.3102. PubMed PMID: 25599176.
27. Farber DL, Yudanin NA, Restifo NP. Human memory T cells: generation, compartmentalization and homeostasis. Nat Rev Immunol. 2014;14(1):24-35. Epub 2013/12/18. doi: 10.1038/nri3567. PubMed PMID: 24336101; PMCID: PMC4032067.
28. Mahnke YD, Brodie TM, Sallusto F, Roederer M, Lugli E. The who's who of T-cell differentiation: human memory T-cell subsets. Eur J Immunol. 2013;43(11):2797-809. Epub 2013/11/22. doi: 10.1002/eji.201343751. PubMed PMID: 24258910.
29. Kern F, Khatamzas E, Surel I, Frommel C, Reinke P, Waldrop SL, Picker LJ, Volk HD. Distribution of human CMV-specific memory T cells among the CD8pos. subsets defined by CD57, CD27, and CD45 isoforms. Eur J Immunol. 1999;29(9):2908-15. Epub 1999/10/03. doi: 10.1002/(SICI)1521-4141(199909)29:09<2908::AID-IMMU2908>3.0.CO;2-8. PubMed PMID: 10508265.
30. Brenchley JM, Karandikar NJ, Betts MR, Ambrozak DR, Hill BJ, Crotty LE, Casazza JP, Kuruppu J, Migueles SA, Connors M, Roederer M, Douek DC, Koup RA. Expression of CD57 defines replicative senescence

and antigen-induced apoptotic death of CD8+ T cells. Blood. 2003;101(7):2711-20. Epub 2002/11/16. doi: 10.1182/blood-2002-07-2103. PubMed PMID: 12433688.
31. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, Holt BJ, Hales BJ, Walker ML, Hollams E, Bochkov YA, Grindle K, Johnston SL, Gern JE, Sly PD, Holt PG, Holt KE, Inouye M. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell Host Microbe. 2015;17(5):704-15. Epub 2015/04/14. doi: 10.1016/j.chom.2015.03.008. PubMed PMID: 25865368; PMCID: PMC4433433.
32. Harimaya A, Takada R, Hendolin PH, Fujii N, Ylikoski J, Himi T. High incidence of Alloiococcus otitidis in children with otitis media, despite treatment with antibiotics. J Clin Microbiol. 2006;44(3):946-9. Epub 2006/03/07. doi: 10.1128/JCM.44.3.946-949.2006. PubMed PMID: 16517881; PMCID: PMC1393137.
33. Lemanske RF, Jr. The childhood origins of asthma (COAST) study. Pediatr Allergy Immunol. 2002;13(s15):38-43. Epub 2003/04/12. PubMed PMID: 12688623.
34. Keller RL, Feng R, DeMauro SB, Ferkol T, Hardie W, Rogers EE, Stevens TP, Voynow JA, Bellamy SL, Shaw PA, Moore PE, Prematurity, Respiratory Outcomes P. Bronchopulmonary Dysplasia and Perinatal Characteristics Predict 1-Year Respiratory Outcomes in Newborns Born at Extremely Low Gestational Age: A Prospective Cohort Study. J Pediatr. 2017;187:89-97 e3. Epub 2017/05/22. doi: 10.1016/j.jpeds.2017.04.026. PubMed PMID: 28528221; PMCID: PMC5533632.
35. Busse WW, Lemanske RF, Jr., Gern JE. Role of viral respiratory infections in asthma and asthma exacerbations. Lancet. 2010;376(9743):826-34. Epub 2010/09/08. doi: 10.1016/S0140-6736(10)61380-3. PubMed PMID: 20816549; PMCID: PMC2972660.
36. Tregoning JS, Schwarze J. Respiratory viral infections in infants: causes, clinical symptoms, virology, and immunology. Clin Microbiol Rev. 2010;23(1):74-98. Epub 2010/01/13. doi: 10.1128/CMR.00032-09. PubMed PMID: 20065326; PMCID: PMC2806659.
37. Chen W, Subbarao P, McGihon RE, Feldman LY, Zhu J, Lou W, Gershon AS, Abdullah K, Moraes TJ, Dubeau A, Sears MR, Lefebvre DL, Turvey SE, Mandhane PJ, Azad MB, To T. Patterns of health care use related to respiratory conditions in early life: A birth cohort study with linked administrative data. Pediatr Pulmonol. 2019;54(8):1267-76. Epub 2019/06/07. doi: 10.1002/ppul.24381. PubMed PMID: 31172683.
38. Restori KH, Srinivasa BT, Ward BJ, Fixman ED. Neonatal Immunity, Respiratory Virus Infections, and the Development of Asthma. Front Immunol. 2018;9:1249. Epub 2018/06/20. doi: 10.3389/fimmu.2018.01249. PubMed PMID: 29915592; PMCID: PMC5994399.
39. Turi KN, Shankar J, Anderson LJ, Rajan D, Gaston K, Gebretsadik T, Das SR, Stone C, Larkin EK, Rosas-Salazar C, Brunwasser SM, Moore ML, Peebles RS, Jr., Hartert TV. Infant Viral Respiratory Infection Nasal Immune-Response Patterns and Their Association with Subsequent Childhood Recurrent Wheeze. Am J Respir Crit Care Med. 2018;198(8):1064-73. Epub 2018/05/08. doi: 10.1164/rccm.201711-2348OC. PubMed PMID: 29733679; PMCID: PMC6221572.
40. Henrickson SE, Manne S, Dolfi DV, Mansfield KD, Parkhouse K, Mistry RD, Alpern ER, Hensley SE, Sullivan KE, Coffin SE, Wherry EJ. Genomic Circuitry Underlying Immunological Response to Pediatric Acute Respiratory Infection. Cell Rep. 2018;22(2):411-26. Epub 2018/01/11. doi: 10.1016/j.celrep.2017.12.043. PubMed PMID: 29320737; PMCID: PMC5796675.
41. Muench MO, Pott Bartsch EM, Chen JC, Lopoo JB, Barcena A. Ontogenic changes in CD95 expression on human leukocytes: prevalence of T-cells expressing activation markers and identification of CD95-CD45RO+ T-cells in the fetus. Dev Comp Immunol. 2003;27(10):899-914. Epub 2003/07/26. doi: S0145305X03000818 [pii]. PubMed PMID: 12880639.
42. Tezuka T, Sugita K, Mizobe N, Goi K, Miyamoto N, Nakamura M, Kagami K, Yokoyama T, Nakazawa S. Transient increase in CD45RO expression on T lymphocytes in infected newborns. Pediatr Res. 1998;43(2):283-90. Epub 1998/02/25. PubMed PMID: 9475298.
43. Mold JE, Venkatasubrahmanyam S, Burt TD, Michaelsson J, Rivera JM, Galkina SA, Weinberg K, Stoddart CA, McCune JM. Fetal and adult hematopoietic stem cells give rise to distinct T cell lineages in humans. Science. 2010;330(6011):1695-9. doi: 10.1126/science.1196509. PubMed PMID: 21164017; PMCID: 3276679.
44. Michaelsson J, Mold JE, McCune JM, Nixon DF. Regulation of T cell responses in the developing human fetus. J Immunol. 2006;176(10):5741-8. PubMed PMID: 16670279.
45. Pagel J, Hartz A, Figge J, Gille C, Eschweiler S, Petersen K, Schreiter L, Hammer J, Karsten CM, Friedrich D, Herting E, Gopel W, Rupp J, Hartel C. Regulatory T cell frequencies are increased in preterm infants with clinical early-onset sepsis. Clin Exp Immunol. 2016;185(2):219-27. Epub 2016/05/11. doi: 10.1111/cei.12810. PubMed PMID: 27163159; PMCID: PMC4955004.

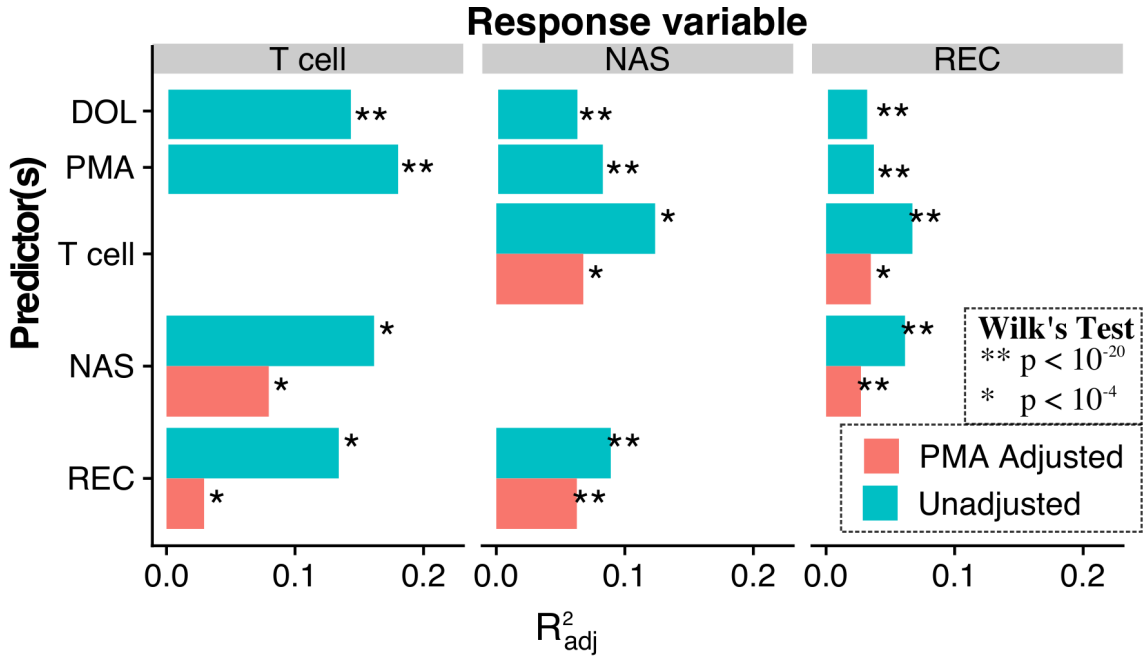
46. Smigiel KS, Richards E, Srivastava S, Thomas KR, Dudda JC, Klonowski KD, Campbell DJ. CCR7 provides localized access to IL-2 and defines homeostatically distinct regulatory T cell subsets. J Exp Med. 2014;211(1):121-36. Epub 2014/01/01. doi: 10.1084/jem.20131142. PubMed PMID: 24378538; PMCID: PMC3892972.
47. Zhivaki D, Lo-Man R. In utero development of memory T cells. Semin Immunopathol. 2017;39(6):585-92. Epub 2017/09/14. doi: 10.1007/s00281-017-0650-0. PubMed PMID: 28900758.
48. Sprent J, Surh CD. Normal T cell homeostasis: the conversion of naive cells into memory-phenotype cells. Nat Immunol. 2011;12(6):478-84. Epub 2011/07/09. PubMed PMID: 21739670; PMCID: PMC3434123.
49. Byrne JA, Stankovic AK, Cooper MD. A novel subpopulation of primed T cells in the human fetus. J Immunol. 1994;152(6):3098-106. Epub 1994/03/15. PubMed PMID: 8144905.
50. Scheible KM, Emo J, Laniewski N, Baran AM, Peterson DR, Holden-Wiltse J, Bandyopadhyay S, Straw AG, Huyck H, Ashton JM, Tripi KS, Arul K, Werner E, Scalise T, Maffett D, Caserta M, Ryan RM, Reynolds AM, Ren CL, Topham DJ, Mariani TJ, Pryhuber GS. T cell developmental arrest in former premature infants increases risk of respiratory morbidity later in infancy. JCI Insight. 2018;3(4). Epub 2018/02/23. doi: 10.1172/jci.insight.96724. PubMed PMID: 29467329; PMCID: PMC5916253.
51. Madan JC, Koestler DC, Stanton BA, Davidson L, Moulton LA, Housman ML, Moore JH, Guill MF, Morrison HG, Sogin ML, Hampton TH, Karagas MR, Palumbo PE, Foster JA, Hibberd PL, O'Toole GA. Serial analysis of the gut and respiratory microbiome in cystic fibrosis in infancy: interaction between intestinal and respiratory tracts and impact of nutritional exposures. MBio. 2012;3(4). Epub 2012/08/23. doi: 10.1128/mBio.00251-12. PubMed PMID: 22911969; PMCID: PMC3428694.
52. Marsland BJ, Trompette A, Gollwitzer ES. The Gut-Lung Axis in Respiratory Disease. Ann Am Thorac Soc. 2015;12 Suppl 2:S150-6. Epub 2015/11/26. doi: 10.1513/AnnalsATS.201503-133AW. PubMed PMID: 26595731.
53. Allaire JM, Crowley SM, Law HT, Chang SY, Ko HJ, Vallance BA. The Intestinal Epithelium: Central Coordinator of Mucosal Immunity. Trends Immunol. 2018;39(9):677-96. Epub 2018/05/03. doi: 10.1016/j.it.2018.04.002. PubMed PMID: 29716793.
54. Thorsen J, Rasmussen MA, Waage J, Mortensen M, Brejnrod A, Bonnelykke K, Chawes BL, Brix S, Sorensen SJ, Stokholm J, Bisgaard H. Infant airway microbiota and topical immune perturbations in the origins of childhood asthma. Nat Commun. 2019;10(1):5001. Epub 2019/11/05. doi: 10.1038/s41467-019-12989-7. PubMed PMID: 31676759; PMCID: PMC6825176.
55. Hempel S, Newberry SJ, Maher AR, Wang Z, Miles JN, Shanman R, Johnsen B, Shekelle PG. Probiotics for the prevention and treatment of antibiotic-associated diarrhea: a systematic review and meta-analysis. JAMA. 2012;307(18):1959-69. Epub 2012/05/10. doi: 10.1001/jama.2012.3507. PubMed PMID: 22570464.
56. Wang Y, Li X, Ge T, Xiao Y, Liao Y, Cui Y, Zhang Y, Ho W, Yu G, Zhang T. Probiotics for prevention and treatment of respiratory tract infections in children: A systematic review and meta-analysis of randomized controlled trials. Medicine (Baltimore). 2016;95(31):e4509. Epub 2016/08/09. doi: 10.1097/MD.0000000000004509. PubMed PMID: 27495104; PMCID: PMC4979858.
57. Hao Q, Dong BR, Wu T. Probiotics for preventing acute upper respiratory tract infections. Cochrane Database Syst Rev. 2015(2):CD006895. Epub 2015/05/01. doi: 10.1002/14651858.CD006895.pub3. PubMed PMID: 25927096.
58. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, McLeod L, Delacqua G, Delacqua F, Kirby J, Duda SN, Consortium RE. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform. 2019;95:103208. Epub 2019/05/13. doi: 10.1016/j.jbi.2019.103208. PubMed PMID: 31078660.
59. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-81. Epub 2008/10/22. doi: 10.1016/j.jbi.2008.08.010. PubMed PMID: 18929686; PMCID: PMC2700030.
60. Nelson EK, Piehler B, Eckels J, Rauch A, Bellew M, Hussey P, Ramsay S, Nathe C, Lum K, Krouse K, Stearns D, Connolly B, Skillman T, Igra M. LabKey Server: an open source platform for scientific data integration, analysis and collaboration. BMC bioinformatics. 2011;12:71. doi: 10.1186/1471-2105-12-71. PubMed PMID: 21385461; PMCID: 3062597.
61. Scheible K, Secor-Socha S, Wightman T, Wang H, Mariani TJ, Topham DJ, Pryhuber G, Quataert S. Stability of T cell phenotype and functional assays following heparinized umbilical cord blood collection. Cytometry A. 2012;81(11):937-49. Epub 2012/10/03. doi: 10.1002/cyto.a.22203. PubMed PMID: 23027690; PMCID: PMC4029066.

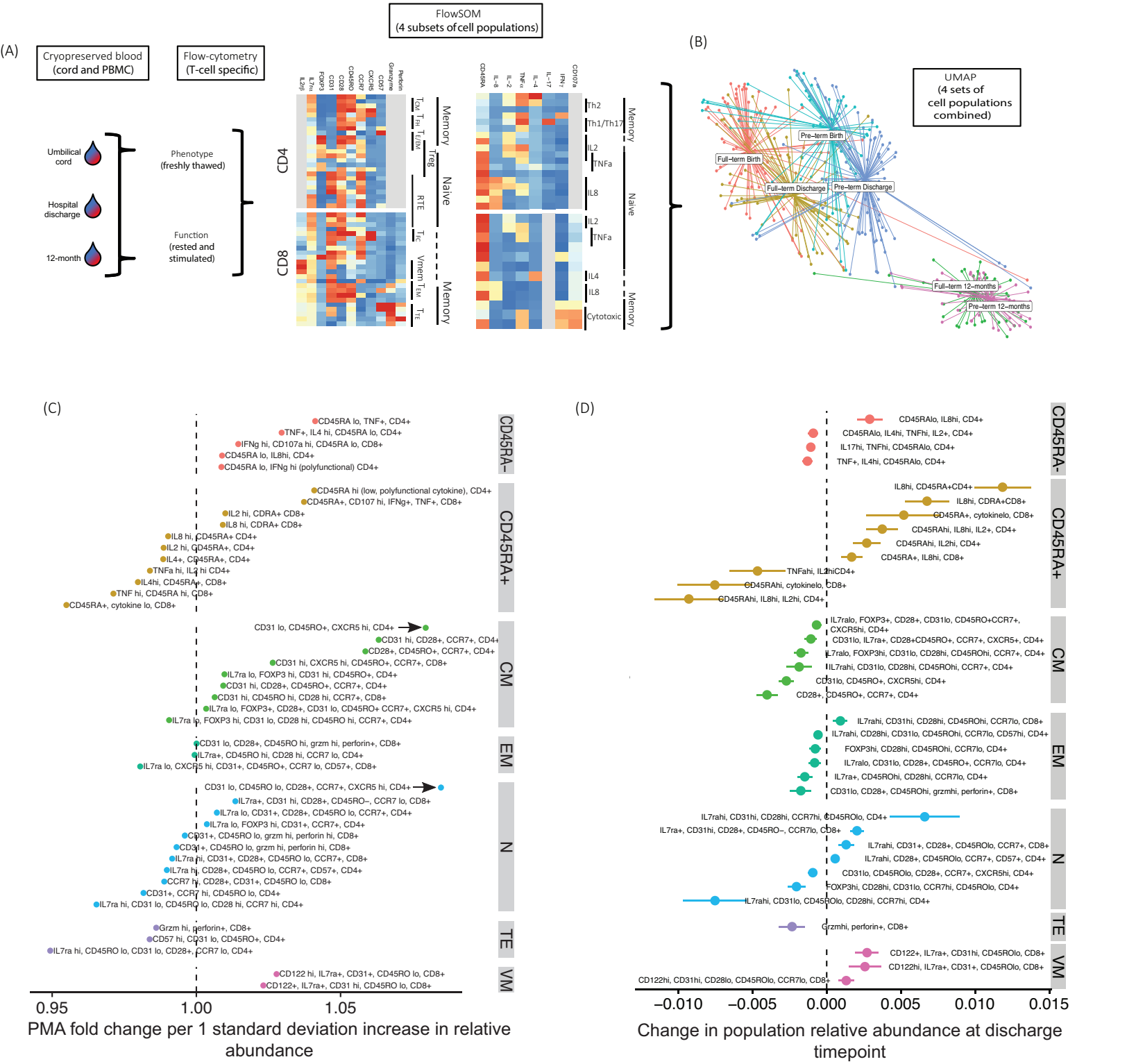
62. Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, Spidlen J, Strain E, Gentleman R. flowCore: a Bioconductor package for high throughput flow cytometry. BMC bioinformatics. 2009;10:106. Epub 2009/04/11. doi: 10.1186/1471-2105-10-106. PubMed PMID: 19358741; PMCID: PMC2684747.
63. Hahne FG, N.; Khodabakhshi, AH.; Wong, C.; Lee, K. . flowStats: Statistical methods for the analysis of flow cytometry data. 3.40.1 ed2019.
64. Grier A, McDavid A, Wang B, Qiu X, Java J, Bandyopadhyay S, Yang H, Holden-Wiltse J, Kessler HA, Gill AL, Huyck H, Falsey AR, Topham DJ, Scheible KM, Caserta MT, Pryhuber GS, Gill SR. Neonatal gut and respiratory microbiota: coordinated development through time and space. Microbiome. 2018;6(1):193. Epub 2018/10/28. doi: 10.1186/s40168-018-0566-5. PubMed PMID: 30367675; PMCID: PMC6204011.
65. Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome. 2014;2(1):6. Epub 2014/02/25. doi: 10.1186/2049-2618-2-6. PubMed PMID: 24558975; PMCID: PMC3940169.
66. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6. Epub 2010/04/13. doi: 10.1038/nmeth.f.303. PubMed PMID: 20383131; PMCID: PMC3156573.
67. Bolyen E. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints. 2018;e27285v1.
68. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3. Epub 2016/05/24. doi: 10.1038/nmeth.3869. PubMed PMID: 27214047; PMCID: PMC4927377.
69. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72(7):5069-72. Epub 2006/07/06. doi: 10.1128/AEM.03006-05. PubMed PMID: 16820507; PMCID: PMC1489311.
70. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. Epub 2013/01/19. doi: 10.1093/molbev/mst010. PubMed PMID: 23329690; PMCID: PMC3603318.
71. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5(3):e9490. Epub 2010/03/13. doi: 10.1371/journal.pone.0009490. PubMed PMID: 20224823; PMCID: PMC2835736.
72. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011;5(2):169-72. Epub 2010/09/10. doi: 10.1038/ismej.2010.133. PubMed PMID: 20827291; PMCID: PMC3105689.
73. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One. 2012;7(2):e30126. Epub 2012/02/10. doi: 10.1371/journal.pone.0030126. PubMed PMID: 22319561; PMCID: PMC3272020.
74. Morgan M. DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome Data. R package version. 2019;1.22.0.
75. Team RC. R: A language and environment for statistical computing. . R Foundation for Statistical Computing, Vienna, Australia. 2013.
76. Anderson-Bergman C. icenReg: Regression models for interval censored data in R. Journal of Statistical Software. 2017;81(12):1-23.
77. Halekoh UH, S; Yan, J. The R package geepack for generalized estimating equations. Journal of Statistical Software 2006;15(2):1-11.
78. Bates DM, M; Bolker, B; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2015;67(1):1-48.
79. Friedman JH, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Statistical Software. 2010;33(1).
80. Bischl BL, M.; Kotthoff, L.; Schiffner, J.; Richter, J.; Studerus, E.; Casalicchio, G.; Jones, Z. mlr: Machine Learning in R. J Machine Learning Research 1. 2000:1-48.

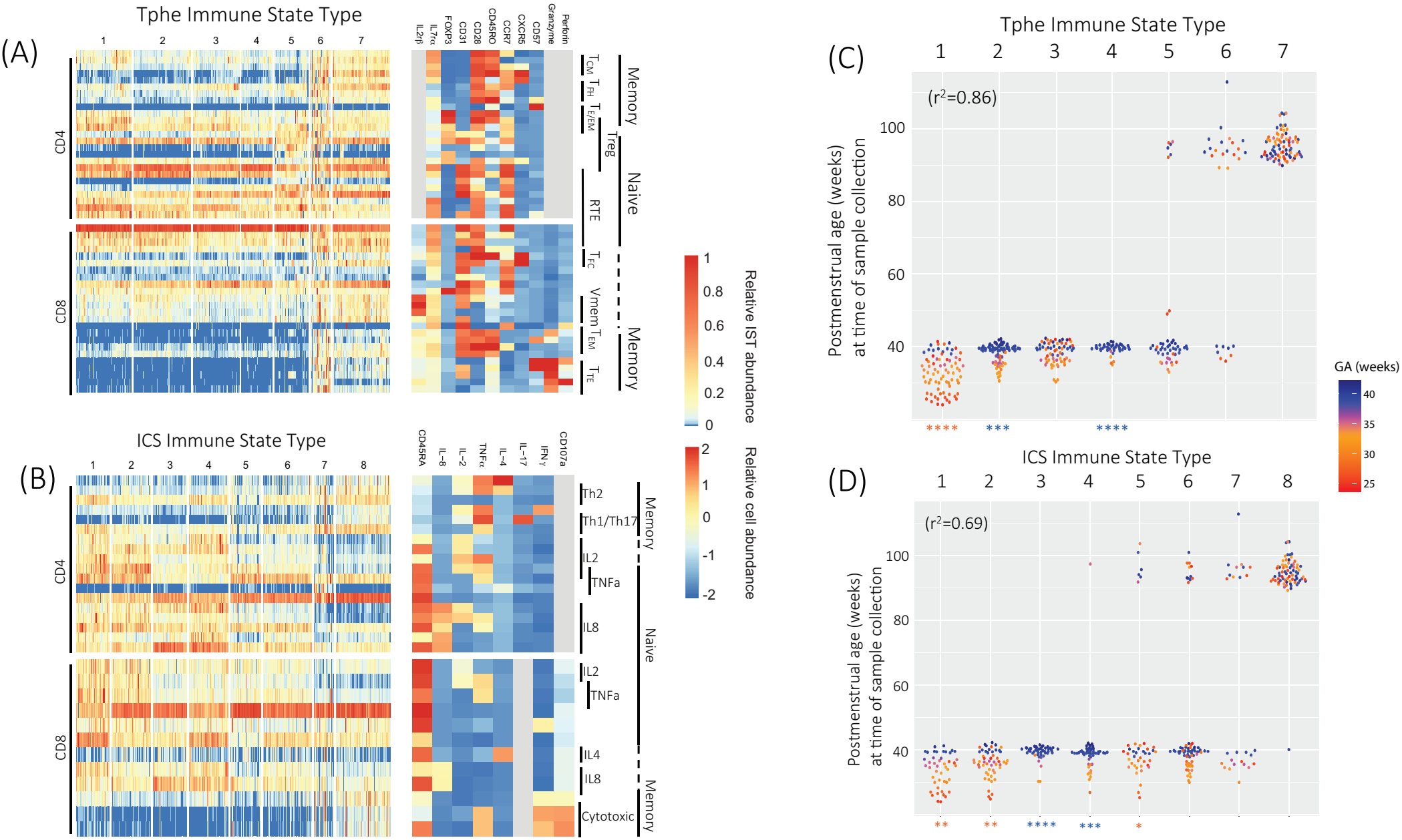
(A)

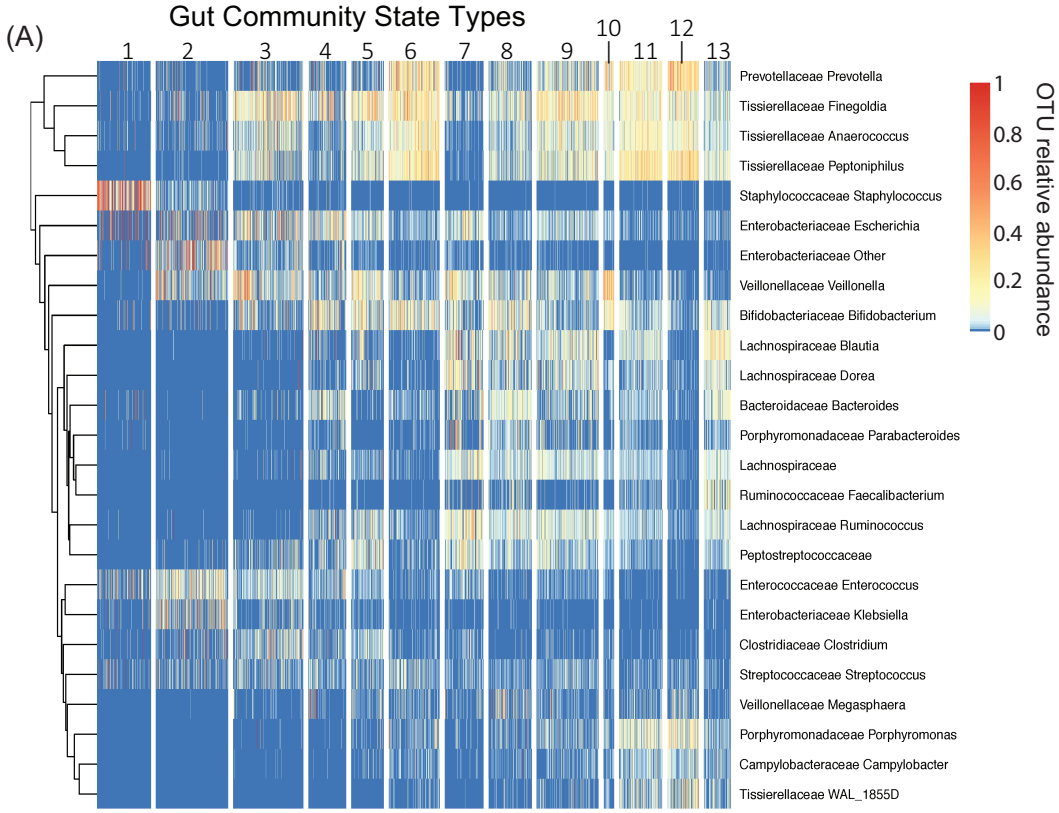

- Infant at Birth
- Immune Maturity
- Microbiota Maturity
- Acute Respiratory Illness
- Chorioamnionitis or Antenatal Antibiotics
Healthy
Pre-term
Full-term
Normal Immune Maturation
Normal Microbiota Maturation
Synchronized
Convergent
Sick
Altered Immune Maturation
Altered Microbiota Maturation
Atypical T cells
Atypical Microbiota
(Diminished Alloiococcus)
Frequent Illnesses
Persistent Respiratory Disease
Precocious Immune Maturation
Delayed Microbiota Maturation
Asynchronous
Mature T cells
Immature Microbiota
0
25
37
50
62
75
87
92 (One Year Old)
Postmenstrual Age In Weeks (PMA)
Gestational Age (In Utero)
Weeks of Life (Ex Utero)
Nasal Microbiota Samples
Rectal Microbiota Samples
Blood T Cell Samples
Outcomes
Respiratory Illness Episodes
&
Persistent Respiratory Disease

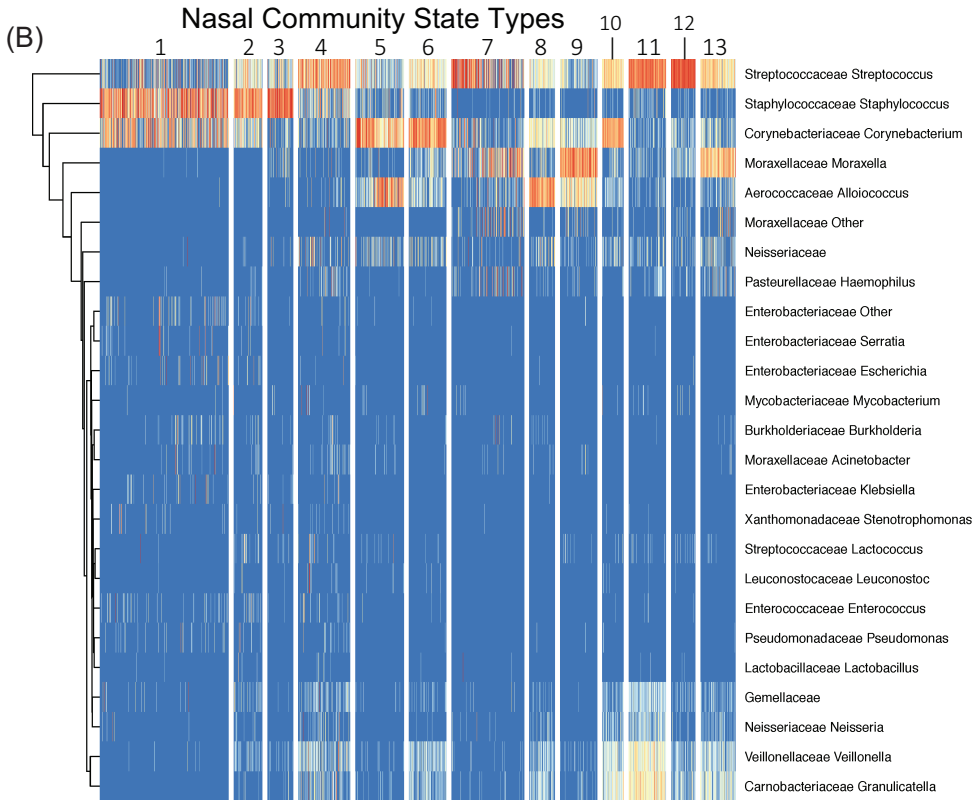
(B)

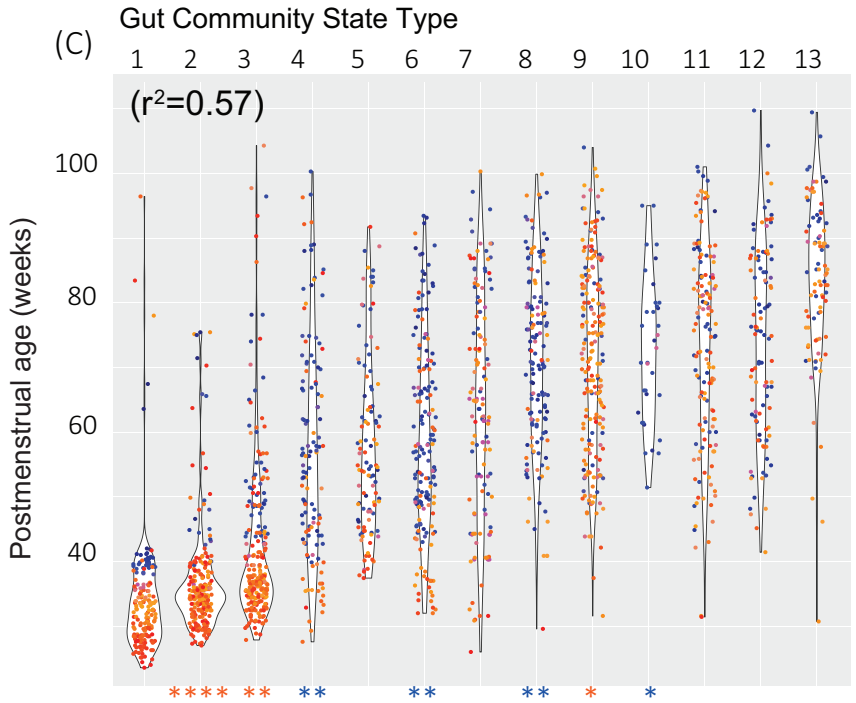

Response variable
T cell
NAS
REC
Predictor(s)
DOL
PMA
T cell
NAS
REC
Wilk's Test
** p < 10^-20
* p < 10^-4
PMA Adjusted
Unadjusted
0.0
0.1
0.2
R²adj

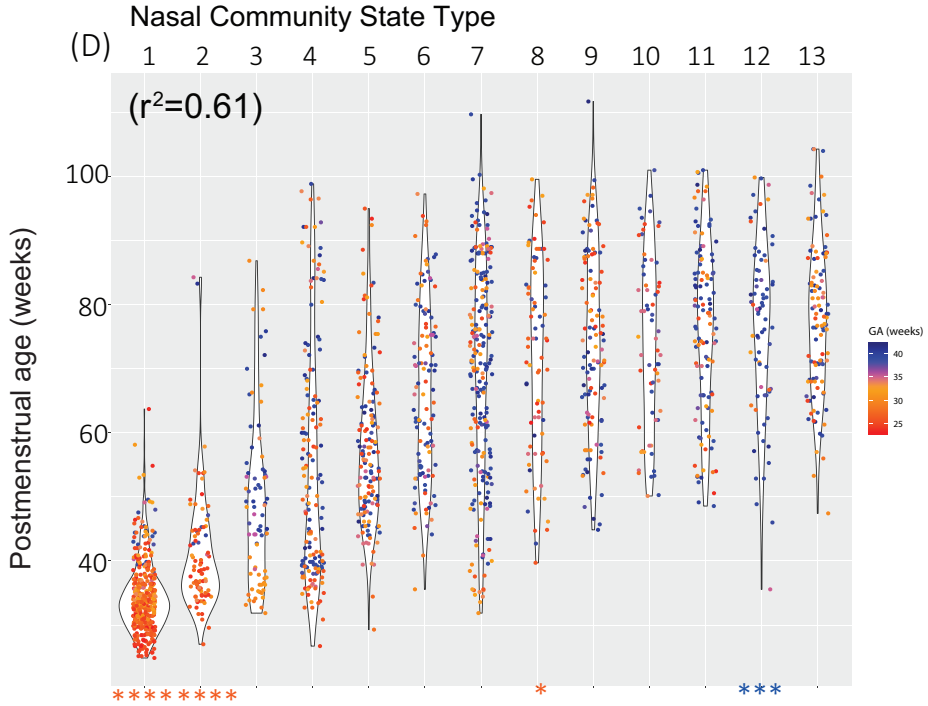

(A)
Cryopreserved blood (cord and PBMC)
Flow-cytometry (T-cell specific)
FlowSOM (4 subsets of cell populations)
Umbilical cord
Hospital discharge
12-month
Phenotype (freshly thawed)
Function (rested and stimulated)
CD4
CD8
Memory
Naive
Memory
Th2
Th1/Th17
IL2
TNFa
IL8
IL2
TNFa
IL4
IL8
Cytotoxic
(B)
UMAP (4 sets of cell populations combined)
Pre–term Birth
Full–term Birth
Full–term Discharge
Pre–term Discharge
Full–term 12–months
Pre–term 12–months
(C)
CD45RA–
CD45RA+
CM
EM
N
TE
VM
PMA fold change per 1 standard deviation increase in relative abundance
(D)
Change in population relative abundance at discharge timepoint
(E)
Gestational age at birth
Full term
Late preterm (32 weeks)
Early preterm (25 weeks)
Arcsin (square root (p))
Postmenstrual Age (weeks)

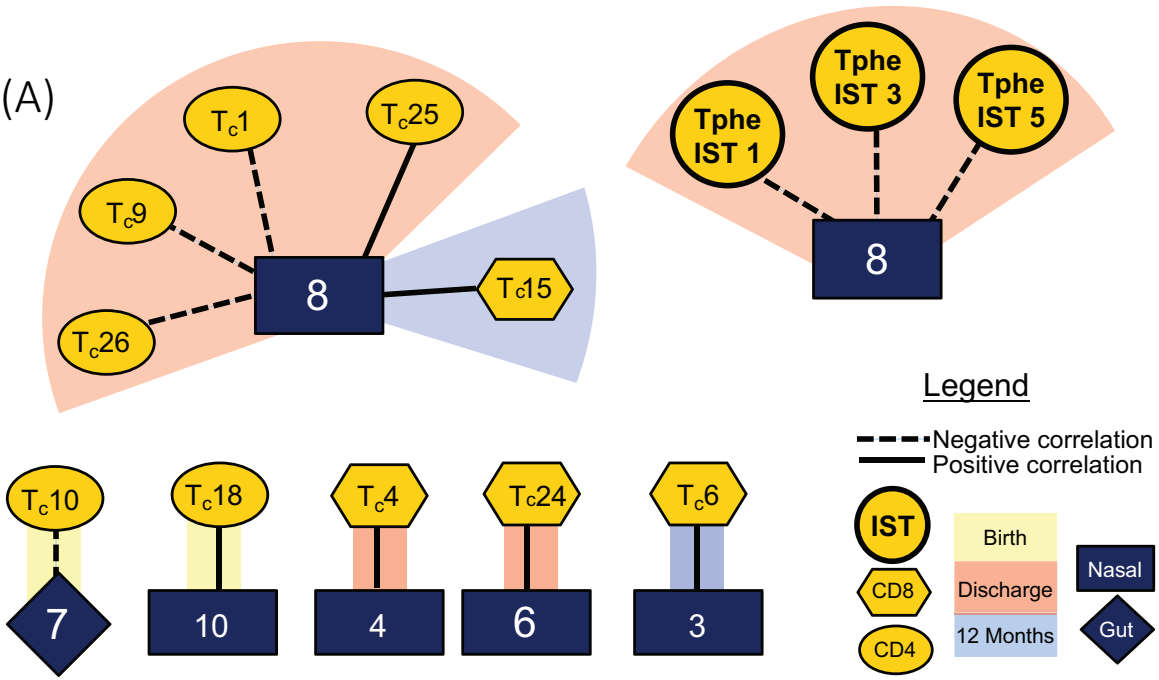

(A)
Tphe Immune State Type
1 2 3 4 5 6 7
CD4
CD8
IL2rβ
IL7rα
FOXP3
CD31
CD28
CD45RO
CCR7
CXCR5
CD57
Granzyme
Perforin
TCM
TFH
TE/EM
Treg
RTE
TFC
Vmem
TEM
TTE
Memory
Naive
Memory
Relative IST abundance
1
0.8
0.6
0.4
0.2
0
(B)
ICS Immune State Type
1 2 3 4 5 6 7 8
CD4
CD8
CD45RA
IL-8
IL-2
TNFα
IL-4
IL-17
IFNγ
CD107a
Th2
Th1/Th17
IL2
TNFa
IL8
IL2
TNFa
IL4
IL8
Cytotoxic
Memory
Naive
Memory
Relative cell abundance
2
1
0
-1
-2
(C)
Tphe Immune State Type
1 2 3 4 5 6 7
(r2=0.86)
Postmenstrual age (weeks)
at time of sample collection
100
80
60
40
****
***
****
GA (weeks)
40
35
30
25
(D)
ICS Immune State Type
1 2 3 4 5 6 7 8
(r2=0.69)
Postmenstrual age (weeks)
at time of sample collection
100
80
60
40
**
**
****
***
*


(A)
Gut Community State Types
1 2 3 4 5 6 7 8 9 10 11 12 13
Prevotellaceae Prevotella
Tissierellaceae Finegoldia
Tissierellaceae Anaerococcus
Tissierellaceae Peptoniphilus
Staphylococcaceae Staphylococcus
Enterobacteriaceae Escherichia
Enterobacteriaceae Other
Veillonellaceae Veillonella
Bifidobacteriaceae Bifidobacterium
Lachnospiraceae Blautia
Lachnospiraceae Dorea
Bacteroidaceae Bacteroides
Porphyromonadaceae Parabacteroides
Lachnospiraceae
Ruminococcaceae Faecalibacterium
Lachnospiraceae Ruminococcus
Peptostreptococcaceae
Enterococcaceae Enterococcus
Enterobacteriaceae Klebsiella
Clostridiaceae Clostridium
Streptococcaceae Streptococcus
Veillonellaceae Megasphaera
Porphyromonadaceae Porphyromonas
Campylobacteraceae Campylobacter
Tissierellaceae WAL_1855D
1
0.8
0.6
0.4
0.2
0
OTU relative abundance

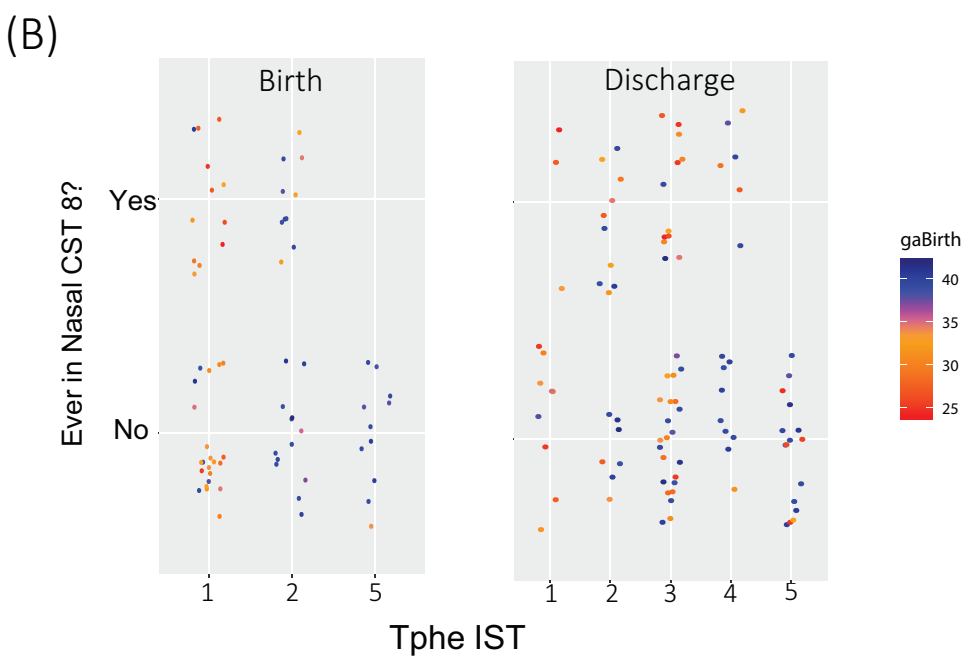

(B)
Nasal Community State Types
1 2 3 4 5 6 7 8 9 10 11 12 13
Streptococcaceae Streptococcus
Staphylococcaceae Staphylococcus
Corynebacteriaceae Corynebacterium
Moraxellaceae Moraxella
Aerococcaceae Alloiococcus
Moraxellaceae Other
Neisseriaceae
Pasteurellaceae Haemophilus
Enterobacteriaceae Other
Enterobacteriaceae Serratia
Enterobacteriaceae Escherichia
Mycobacteriaceae Mycobacterium
Burkholderiaceae Burkholderia
Moraxellaceae Acinetobacter
Enterobacteriaceae Klebsiella
Xanthomonadaceae Stenotrophomonas
Streptococcaceae Lactococcus
Leuconostocaceae Leuconostoc
Enterococcaceae Enterococcus
Pseudomonadaceae Pseudomonas
Lactobacillaceae Lactobacillus
Gemellaceae
Neisseriaceae Neisseria
Veillonellaceae Veillonella
Carnobacteriaceae Granulicatella

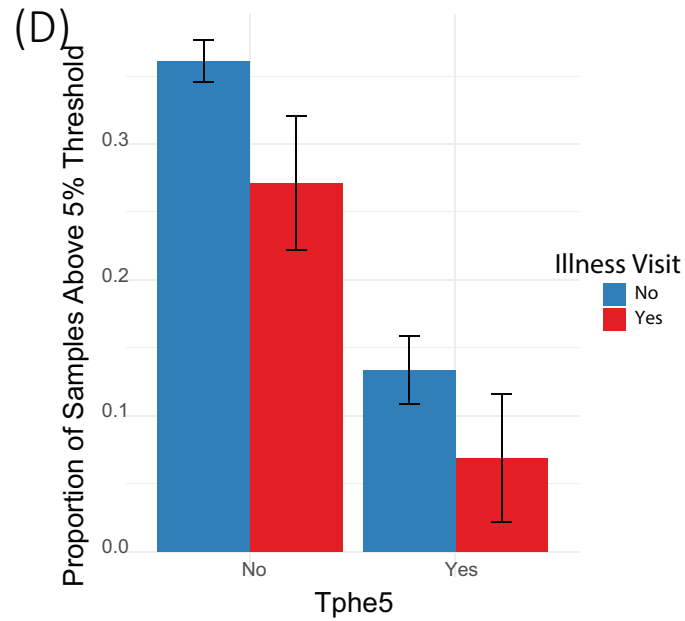

(C)
Gut Community State Type
1 2 3 4 5 6 7 8 9 10 11 12 13
(r²=0.57)
Postmenstrual age (weeks)
100
80
60
40

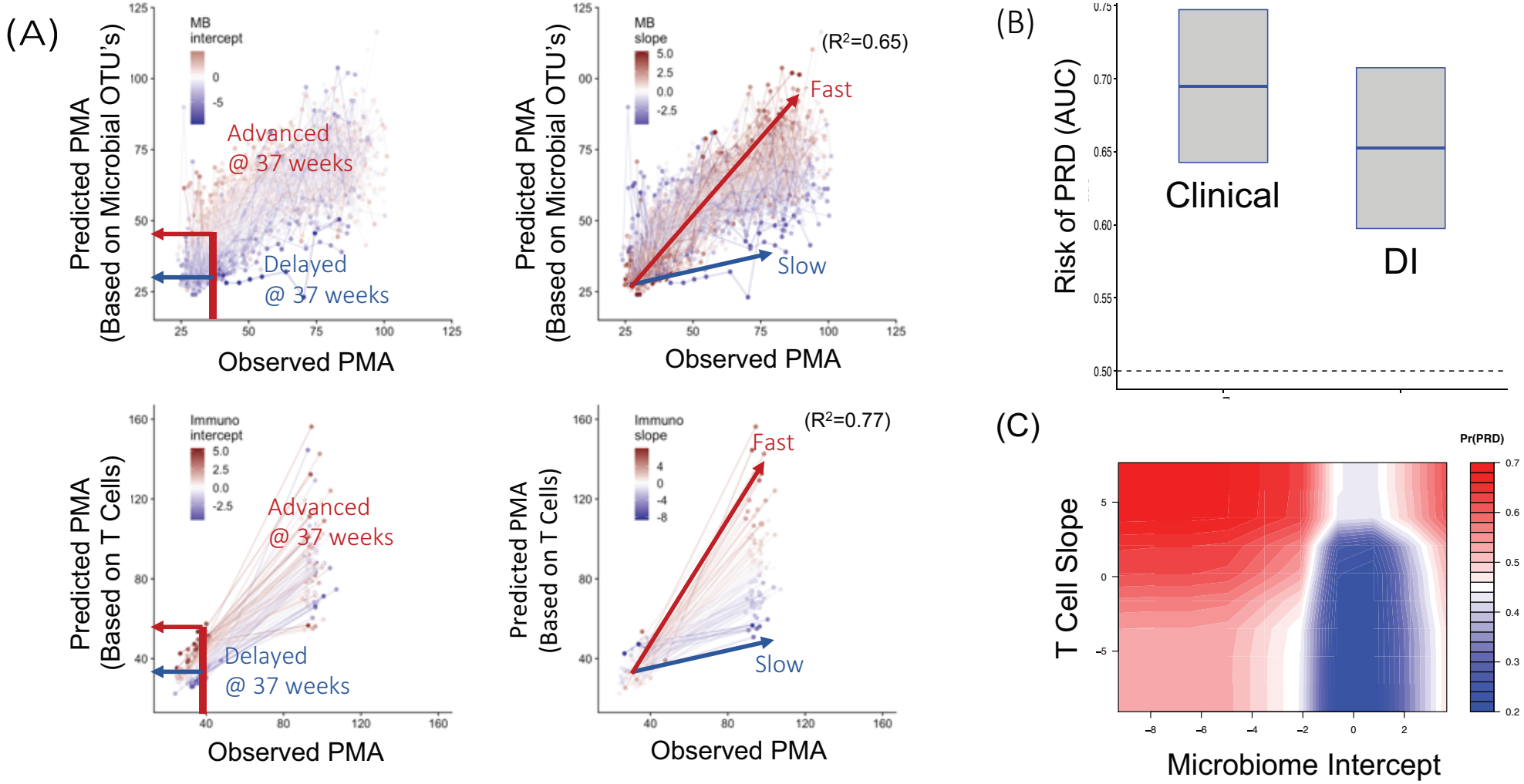

(D)
Nasal Community State Type
1 2 3 4 5 6 7 8 9 10 11 12 13
(r²=0.61)
Postmenstrual age (weeks)
100
80
60
40
GA (weeks)
40
35
30
25


(A)
Tc1
Tc25
Tc9
8
Tc15
Tc26
Tphe IST 1
Tphe IST 3
Tphe IST 5
8
Tc10
Tc18
Tc4
Tc24
Tc6
7
10
4
6
3
Legend
Negative correlation
Positive correlation
IST
CD8
CD4
Birth
Discharge
12 Months
Nasal
Gut

(C)
No Tphe5
Yes Tphe5
Relative Alloiococcus Abundance (Nasal)
1.00
0.75
0.50
0.25
0.00
40
60
80
100
Postmenstrual Age (weeks)


(B)
Birth
Discharge
Ever in Nasal CST 8?
Yes
No
1
2
5
1
2
3
4
5
Tphe IST
gaBirth
40
35
30
25


(D)
Proportion of Samples Above 5% Threshold
0.3
0.2
0.1
0.0
No
Yes
Tphe5
Illness Visit
No
Yes


(A)
Predicted PMA (Based on Microbial OTU's)
Observed PMA
MB intercept
Advanced @ 37 weeks
Delayed @ 37 weeks
MB slope
(R²=0.65)
Fast
Slow
Predicted PMA (Based on T Cells)
Immuno intercept
Immuno slope
(R²=0.77)
(B)
Risk of PRD (AUC)
Clinical
DI
(C)
T Cell Slope
Microbiome Intercept
Pr(PRD)