Using Gut Microbiome Profiles To Predict Growth Outcomes In Preterm Infants

A. Tandon[1], D. Genetti[1], J. Levesque[1], D. Gallagher[1], T. Warren[1], A. La Reau[2], E. Smith[2], F. Jia[2], E. Adams[2], T. Ward[2]
1. Astarte Medical, Yardley, PA, USA; 2. Diversigen, New Brighton, MN, USA

BACKGROUND

- Growth failure among preterm infants has been shown to have an impact on poor neurodevelopment outcomes, with the gut microbiota potentially playing an integral role (1).
- We hypothesized that we could identify infants at risk for growth failure in the first few weeks of life based on their microbiome profile, leading to better outcomes.
- To identify age and growth discriminatory taxa, we developed machine learning models for age prediction based on microbiome profiles as well as classification for normal growth or growth failure.

STUDY DESIGN

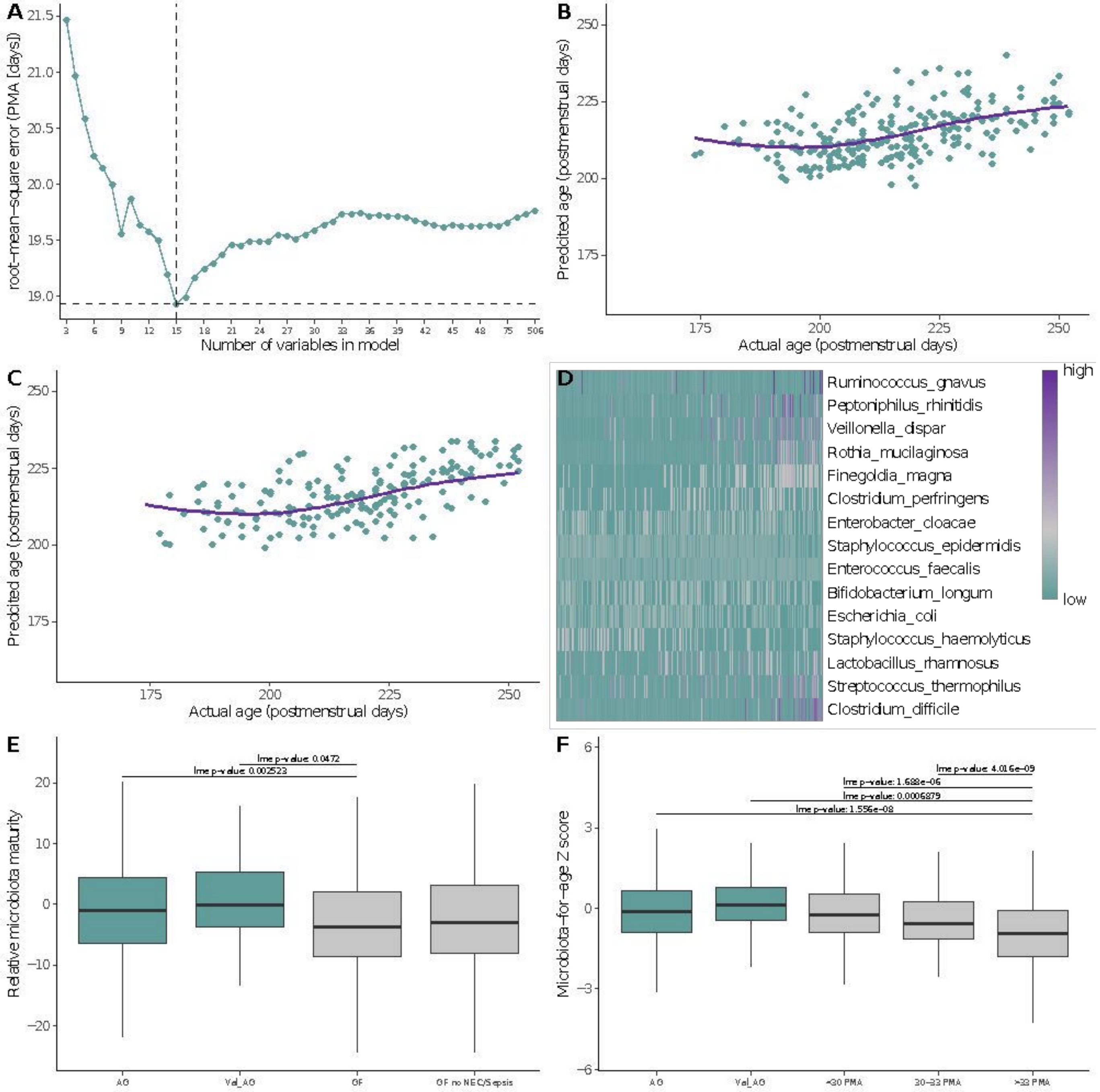
- **Study Population**: A total of 267 preterm infants from 3 different clinical sites were followed from birth to hospital discharge.
- **Sample collection**: Stool samples (n=2996) were collected longitudinally from 1 to 174 days of life, from infants with normal growth (n= 157), growth failure (n=102) and infants that died (n=8). Growth Failure is defined as birth-to-discharge weight z-score decline of ≥ 1.2.
- **Sample Processing**: Extracted DNA was sequenced via shotgun metagenomic sequencing at a mean depth of 28,390,685 sequences.
- **Data Generation:** Shotgun sequencing was annotated using MetaPhlAn2 and HUMAnN2 (3).
- **Covariates:** Growth Status, Clinical Sites, Probiotic (Yes/No), Sepsis, Necrotizing Enterocolitis, Mode of Birth, Gender, Gestational Age at Birth, Post-Menstrual Age.

ANALYSIS PLAN

- **Random Forest Regression:** Age predictions were generated for infants using their microbiome profiles followed by calculation of their microbiota for age Z-score (MAZ) and relative microbiota maturity (RMM). These were then compared between healthy infants and those with growth failure (2).
- **Random Forest Classification:** We used microbial profiles with additional metadata features to predict growth outcomes using a 10-fold cross-validation and feature step-down approach.
- **Alpha diversity:** SplinectomeR (4) was used to calculate differences in microbial richness between normal growth and growth failure infants over time.
- **Differential Taxa:** MetaLonDA (5) was to used assess the difference in absolute abundance of taxa between normal growth and growth failure over time.

CONCLUSIONS

- Growth failure infants showed lower microbial richness and altered microbial composition compared to normal growth infants.
- Our Random Forest Regression identified 15 taxa that were most important for age prediction, resulting in a final model explaining 79% of the variance.
- Random Forest Classification predicted overall growth status with 70% accuracy, with 10 taxa found to be most important in growth status prediction in the Samples ≤ 15 DOL model.



Random forest regression revealed differences in microbiota age between healthy infants and growth failure infants

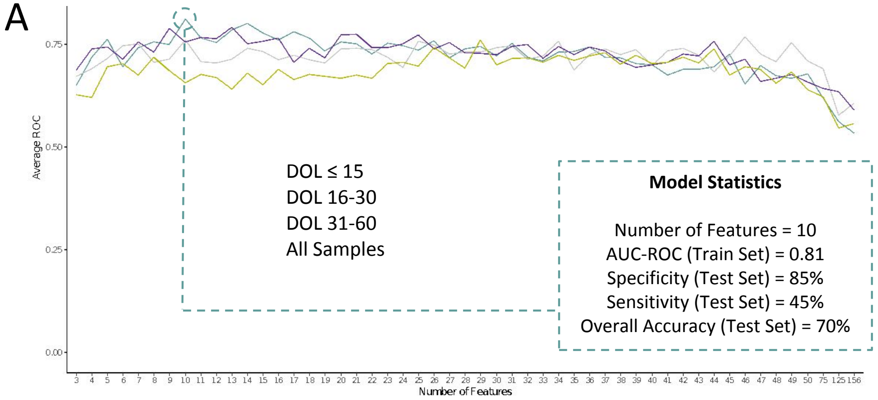
- Using a stepdown approach to systematically reduce model complexity and improve performance, it was determined that the minimal model with the best performance included 15 features and had an average error of ± 19 days postmenstrual age.
- Models consistently predict higher ages for young infants and lower ages for old infants.
- Several taxa included in the model show clear abundance patterns over infant's age. Some show higher relative abundance in older infants (e.g. *Finegoldia magna* and *Peptoniphilis rhinitidis*), while others show the opposite pattern (e.g. *Staphylococcus haemolyticus*).
- Growth failure infants have reduced relative microbiota maturity and microbiota-for-age Z scores relative to healthy infants.

Figure 1. Random forest regression using microbiome profiles of healthy infants for model generation. For all plots shown at left, healthy infants are shown in green and infants with growth failure are in gray. (A) Model performance (error rate) using a step-down approach for feature selection. (B-C) Actual age vs Predicted age for the model training set alone (25 infants, B), and the test set alone (17 infants, C). (D) Heatmap of the 15 taxa that were included in the final model. Only samples from healthy infants are shown and samples are ordered from left to right in increasing age of the infant from which the sample was taken. Colors indicate row-normalized abundance. (E) Relative microbiota maturity (RMM) for various subsets of infants (AG: appropriate growth, Val_AG: validation set of appropriate growth, GF: growth failure, GF no NEC/Sepsis: growth failure with no NEC or sepsis). (F) Microbiota for age Z-score (MAZ) for appropriate growth infants and several age subdivisions of growth failure infants.

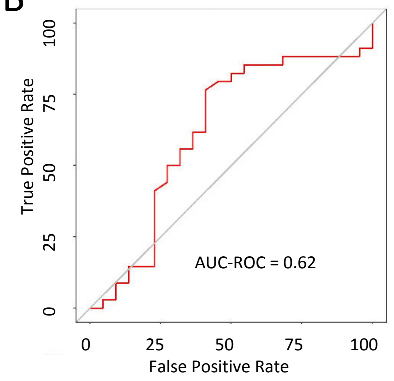
Random forest classification revealed taxa discriminatory for growth status and differences in classification accuracy over time

- Using a feature stepdown approach with samples subset to different DOL time windows, the best-performing model was found for the `Samples ≤ 15 DOL` time window for a reduced set of 10 features, with an average AUC-ROC value of 0.81.
- Overall, all models showed a marked improvement in ROC at the beginning of the stepdown, with gradual tapering off as fewer features remained.
- The best-performing model predicts growth status at 70% overall accuracy, with notably better performance when predicting Normal Growth compared to Growth Failure.

A



B



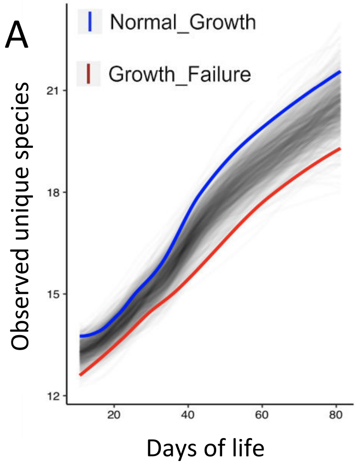
C

Taxon	Variable Importance
Veillonella parvula	8.695
Klebsiella oxytoca	8.671
Klebsiella unclassified	8.525
Dermabacter sp. HFH0086	8.162
Klebsiella pneumoniae	7.793
Streptococcus mutans	7.608
Bifidobacterium catenulatum	7.459
Clostridium innocuum	7.372
Staphylococcus caprae capitis	7.363
Streptococcus lutetiensis	7.229

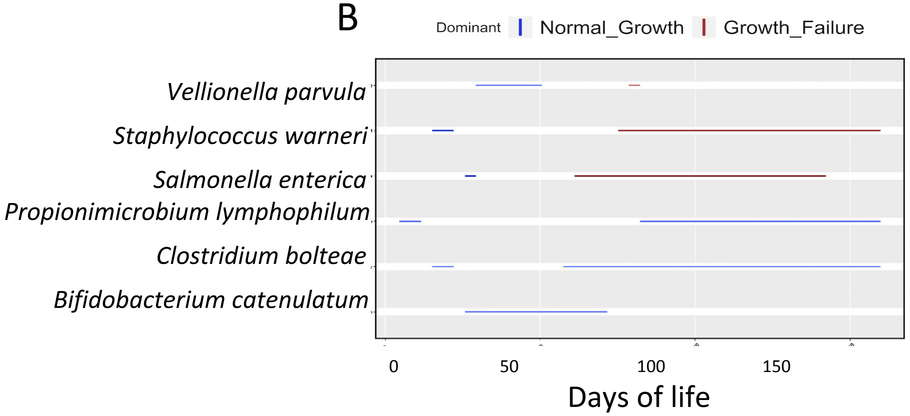
Figure 2. Random forest classification of growth status using microbiome profiles. (A) Feature stepdown line plot showing improvement in average area under the ROC curve over successive iterations for four separate time window models. Model statistics for the best-performing model are highlighted. (B) AUC-ROC plot for the best-performing step of the feature stepdown for the Samples DOL ≤ 15 DOL model against the Test Set samples. (C) Taxa by descending feature importance from the best-performing step of the feature stepdown for the Samples DOL ≤ 15 model.

Figure 3. Differences in alpha and beta diversity and taxa over time. (A) Microbial richness by growth status with respect to time. The shaded area represents the 95% confidence interval for no difference between two groups. (B) Microbial taxa differentially abundant by growth status with respect to time, by FDR corrected p value < 0.05.

A



B



Normal and growth failure infants have differing richness and composition

- Normal growth infants have significantly higher microbial richness ($p <$ 0.05 from day 30 to day 55) compared to those with growth failure status.
- *Vellionella parvula* & *Bifidobacterium catenulatum* were identified as differentially abundant taxa among normal growth infants.

References. (1) Younge NE, et al. *Scientific Reports.* **9**: 8167 (2019). (2) Subramanian, S, et al. *Nature.* **510,** 417–421 (2014). (3) Franzosa EA, et al. *Nat Methods*. **15**:962-968 (1015). (3) Shields-Cutler RR, et al. *Front. Microbiol.* **9**:785 (2018). (4) Wang X, et al. *PLOS ONE* **7**: e50267 (2012). (5) Metwally A, et al. *Microbiome.* **6,** 32 (2018).

Contact Information

- If you have any questions for (Tracy Warren, Ari Tandon), please text or video chat during the conference through the platform.

- You can also contact (Tracy Warren, Ari Tandon) via email at (tracy@astartemedical.com ; arti@astartemedical.com)

- Astarte Medical is a precision nutrition company using software and predictive analytics to improve outcomes during the first 1,000 days of life. With an initial focus on preterm infants, Astarte Medical supports feeding protocols, practice and decision-making in the neonatal ICU with a suite of digital tools and diagnostics designed to standardize feeding, optimize nutrition and quantify gut health. (https://astartemedical.com)