Integrating Longitudinal, Clinical, and Microbiome Data to Predict Growth Failure in Preterm Infants

Presenter Name: David Genetti

Institution: Astarte Medical

Email: dave@astartemedical.com

PAS

2021 VIRTUAL

Astarte Medical

Systems Biology Group

Disclosure

David Genetti has disclosed the following financial relationships. Any real or apparent conflicts of interest related to the content of this presentation have been resolved.

Affiliation / Financial Interest	Organization
Chief Technology Officer	Astarte Medical

This investigation is the result of an ongoing, paid collaboration between Astarte Medical and Carnegie Mellon University.

Motivation

- Growth failure (GF) in preterm infants is associated with clinical disorders and can lead to poor neurodevelopmental outcomes.
- Gut microbiota for preterm infants is influenced by preterm birth, the environment, feeding, and clinical care.
- **Our goal:** Use Machine Learning methods to integrate clinical and microbiome data to Identify infants at risk of GF earlier in their NICU stay. Use the method to propose potential interventions (feeding, antibiotics) to improve outcomes.

Data And Clinical Features

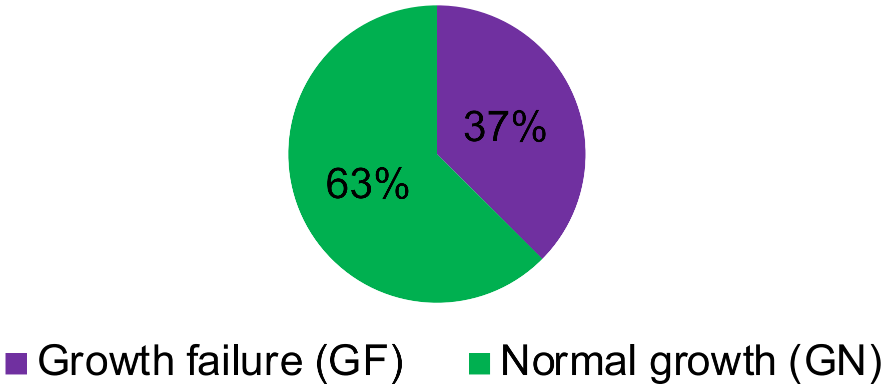
Total infants
259

United States
84 (2 sites)

United Kingdom
175 (1 site)

GF = birth-to-discharge weight z-score decline ≥1.2

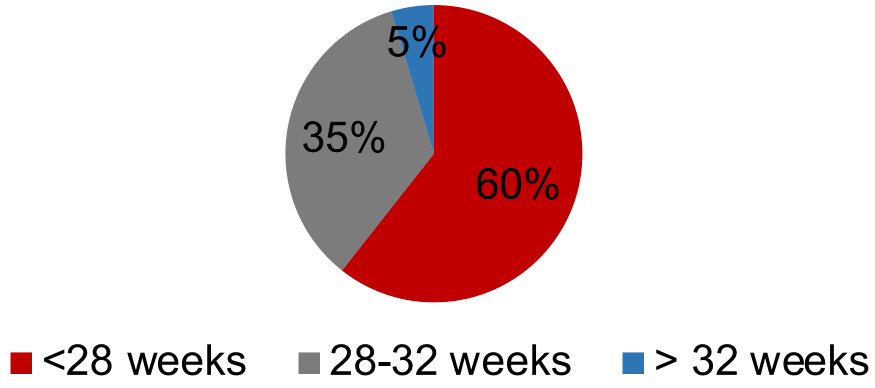
Growth status



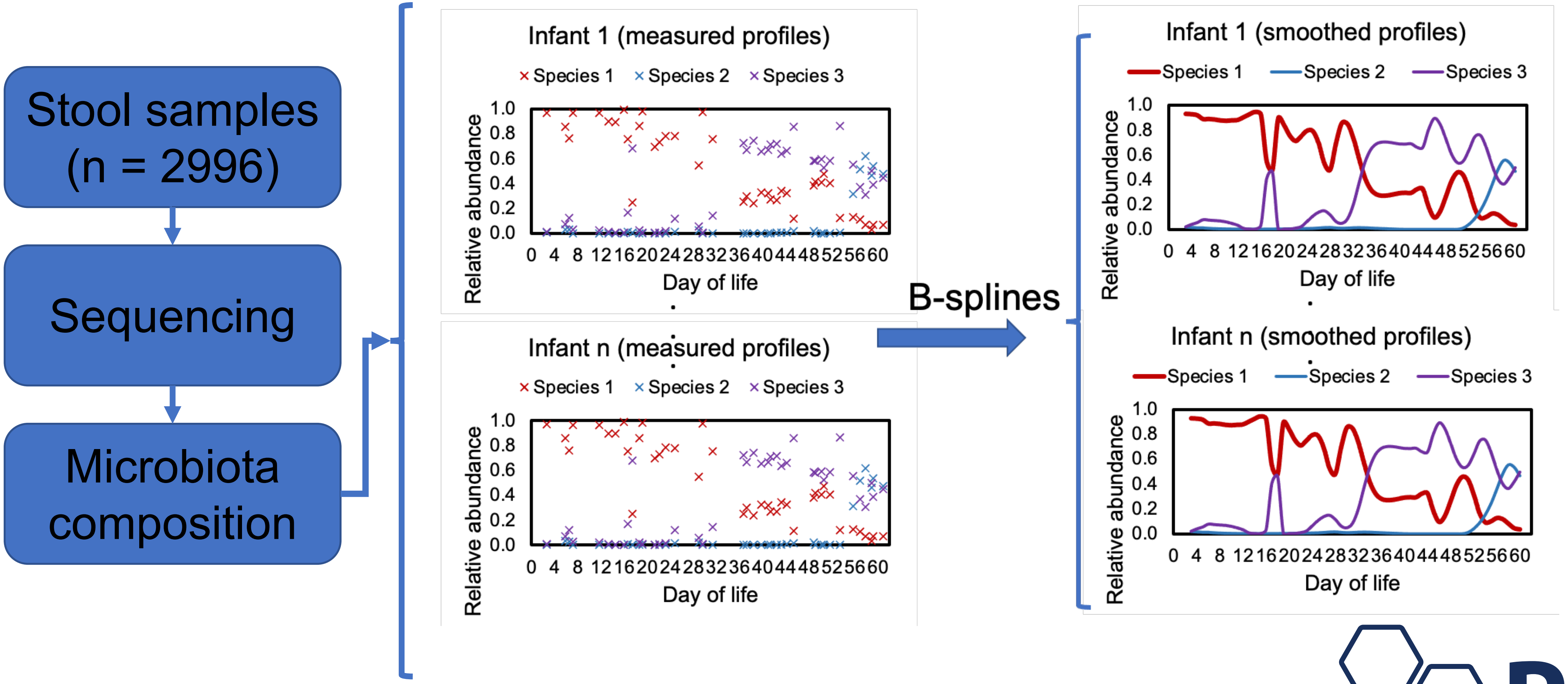
Clinical data includes

- Day of life
- Post-menstrual age (PMA)
- Longitudinal body weight
  - Imputed, if missing
- Feeding type and quantity
- Medication & probiotics

Gestational age at birth



Microbiome Data & Pre-processing



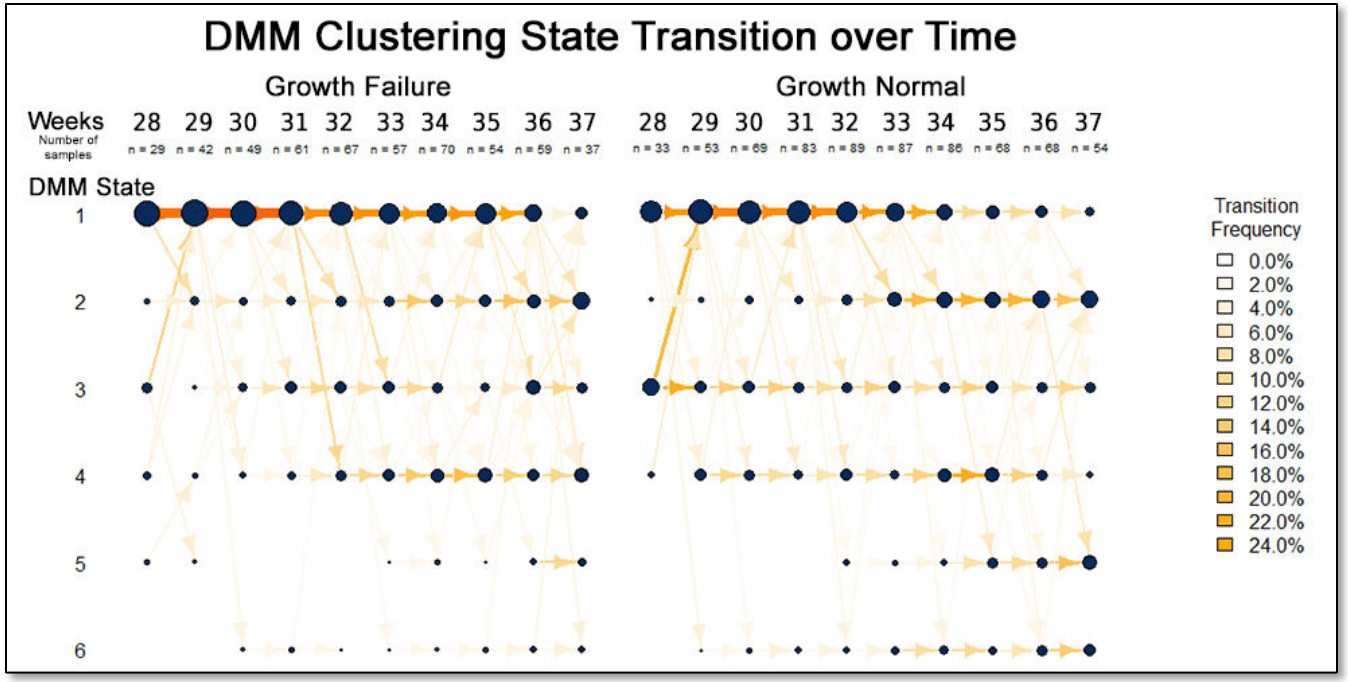
Stool samples (n = 2996)

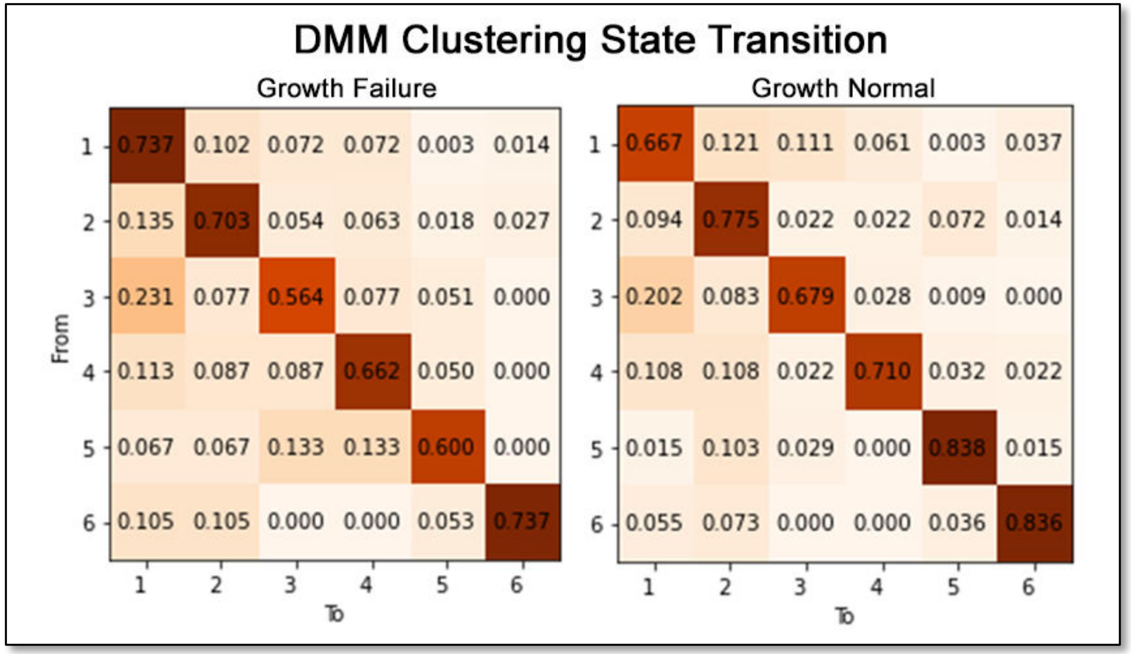
Sequencing

Microbiota composition

Lugo-Martinez et al *Microbiome,* 2019, Ruiz-Perez et al, *mSystems*, 2021

Initializing HMM With DMM Clustering




DMM Clustering State Transition

Growth Failure

From \ To	1	2	3	4	5	6
1	0.737	0.102	0.072	0.072	0.003	0.014
2	0.135	0.703	0.054	0.063	0.018	0.027
3	0.231	0.077	0.564	0.077	0.051	0.000
4	0.113	0.087	0.087	0.662	0.050	0.000
5	0.067	0.067	0.133	0.133	0.600	0.000
6	0.105	0.105	0.000	0.000	0.053	0.737

Growth Normal

From \ To	1	2	3	4	5	6
1	0.667	0.121	0.111	0.061	0.003	0.037
2	0.094	0.775	0.022	0.022	0.072	0.014
3	0.202	0.083	0.679	0.028	0.009	0.000
4	0.108	0.108	0.022	0.710	0.032	0.022
5	0.015	0.103	0.029	0.000	0.838	0.015
6	0.055	0.073	0.000	0.000	0.036	0.836

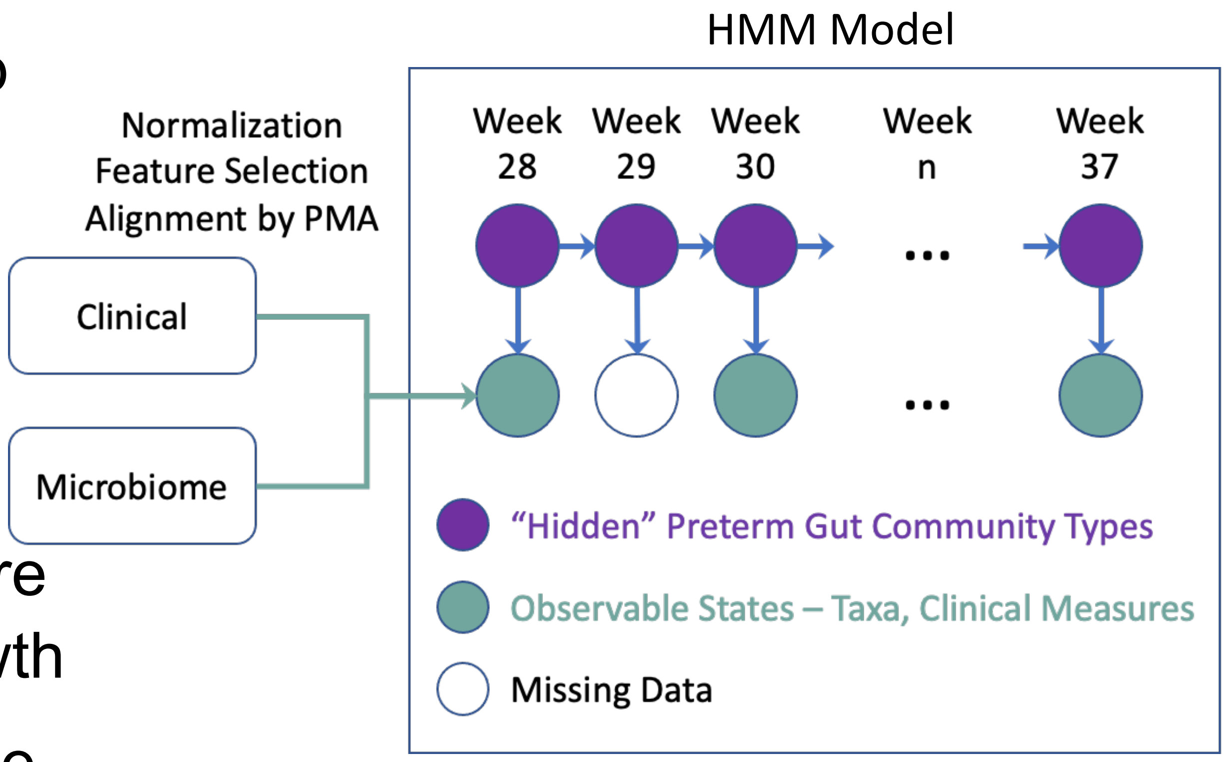
Poster # EP-187 for details on PGCT and DMM

Holmes et al *PLoS ONE,* 2012, Stewart et al, *Nature*, 2018

PAS
Pediatric Academic Societies

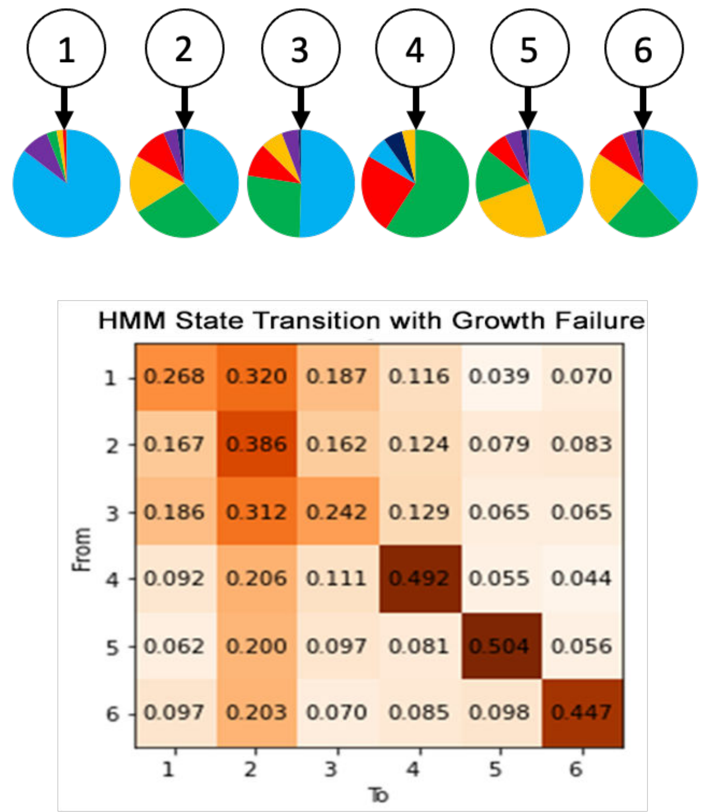
Hidden Markov Model (HMM)

- HMMs are used to model the relationship between observations and hidden states
    - *Observed data*: microbiome samples, clinical features
    - *Hidden states*: gut community types
- **Task:** Learn two different HMM models
    - GF model – infants with Growth Failure
    - GN model – infants with Normal Growth
- **Prediction:** For each new infant, compute likelihood ratio between the two models



Learned HMM And Results

GF HMM - Top 15 Taxa per State



HMM State Transition with Growth Failure

From \ To	1	2	3	4	5	6
1	0.268	0.320	0.187	0.116	0.039	0.070
2	0.167	0.386	0.162	0.124	0.079	0.083
3	0.186	0.312	0.242	0.129	0.065	0.065
4	0.092	0.206	0.111	0.492	0.055	0.044
5	0.062	0.200	0.097	0.081	0.504	0.056
6	0.097	0.203	0.070	0.085	0.098	0.447

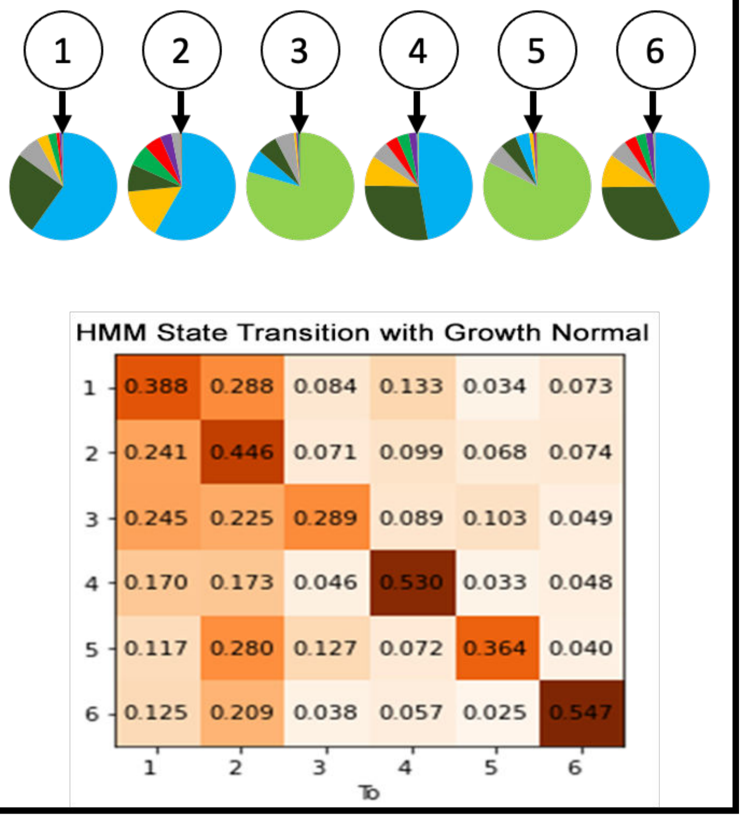
Representative List

E. faecalis
E. coli
B. Longum
E. unclassified

K. oxytoca
B. breve
S. epidermidis
E. cloacae

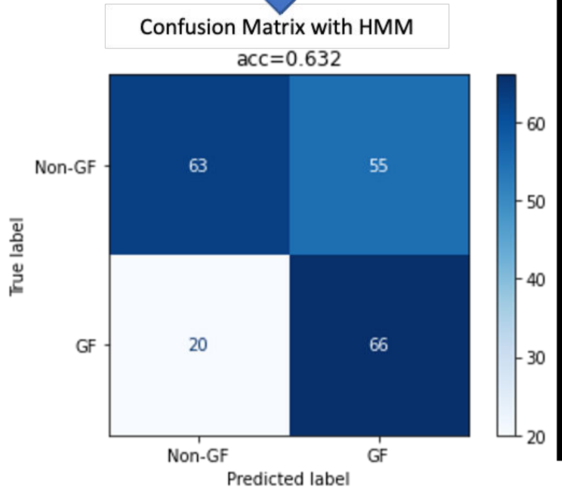
B. bifidum

GN HMM - Top 15 Taxa per State



HMM State Transition with Growth Normal

From \ To	1	2	3	4	5	6
1	0.388	0.288	0.084	0.133	0.034	0.073
2	0.241	0.446	0.071	0.099	0.068	0.074
3	0.245	0.225	0.289	0.089	0.103	0.049
4	0.170	0.173	0.046	0.530	0.033	0.048
5	0.117	0.280	0.127	0.072	0.364	0.040
6	0.125	0.209	0.038	0.057	0.025	0.547



Confusion Matrix with HMM

acc=0.632

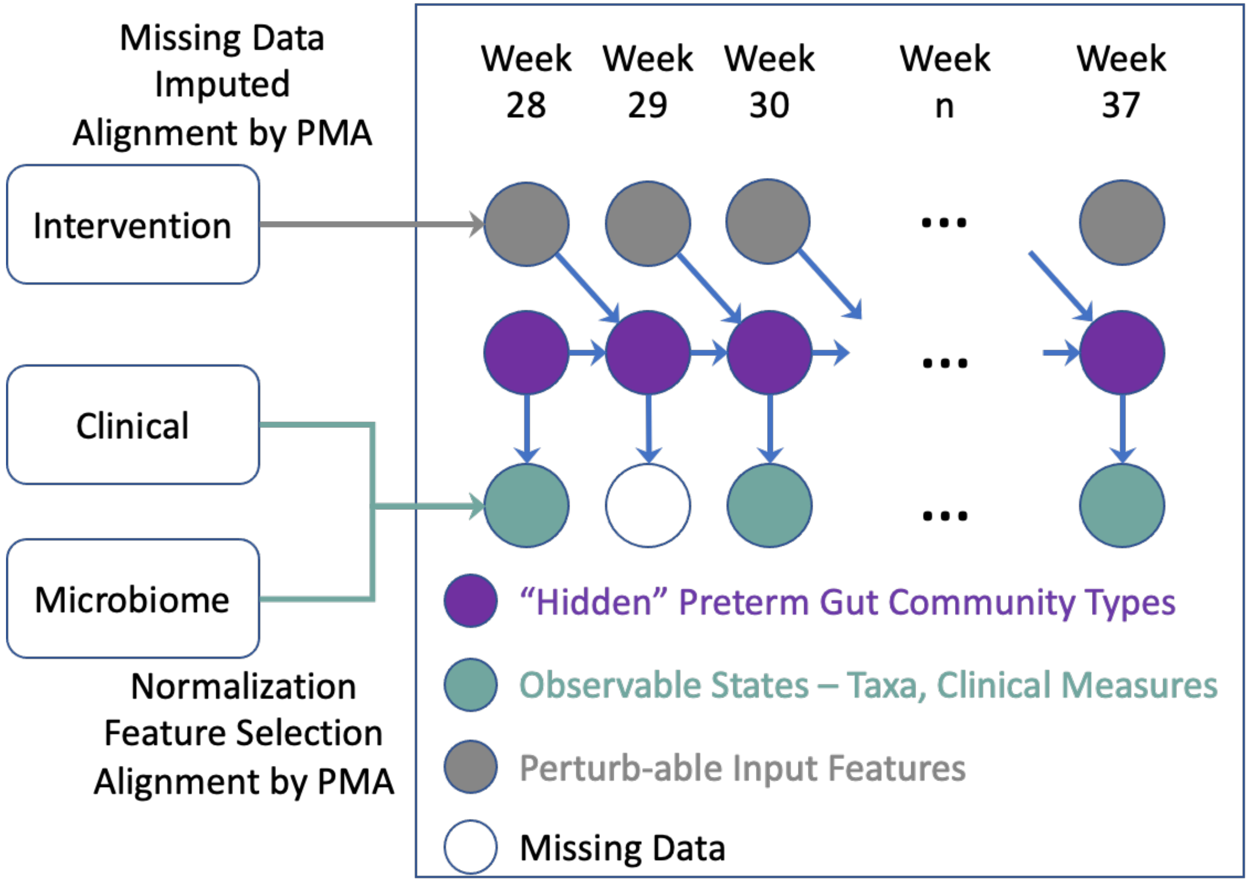
True label \ Predicted label	Non-GF	GF
Non-GF	63	55
GF	20	66

77% True Positive Rate

Pediatric Academic Societies

Input/Output HMM (IOHMM)

- Input/Output HMM adds an input layer to help determine state transitions
    - *Input layer*: intervention data such as feeding type or medication
- **Task:** Extend HMM to include input layers
- **Intervention analysis:** Systematically analyze effects of changing values for each intervention type



Missing Data
Imputed
Alignment by PMA

Intervention

Clinical

Microbiome

Normalization
Feature Selection
Alignment by PMA

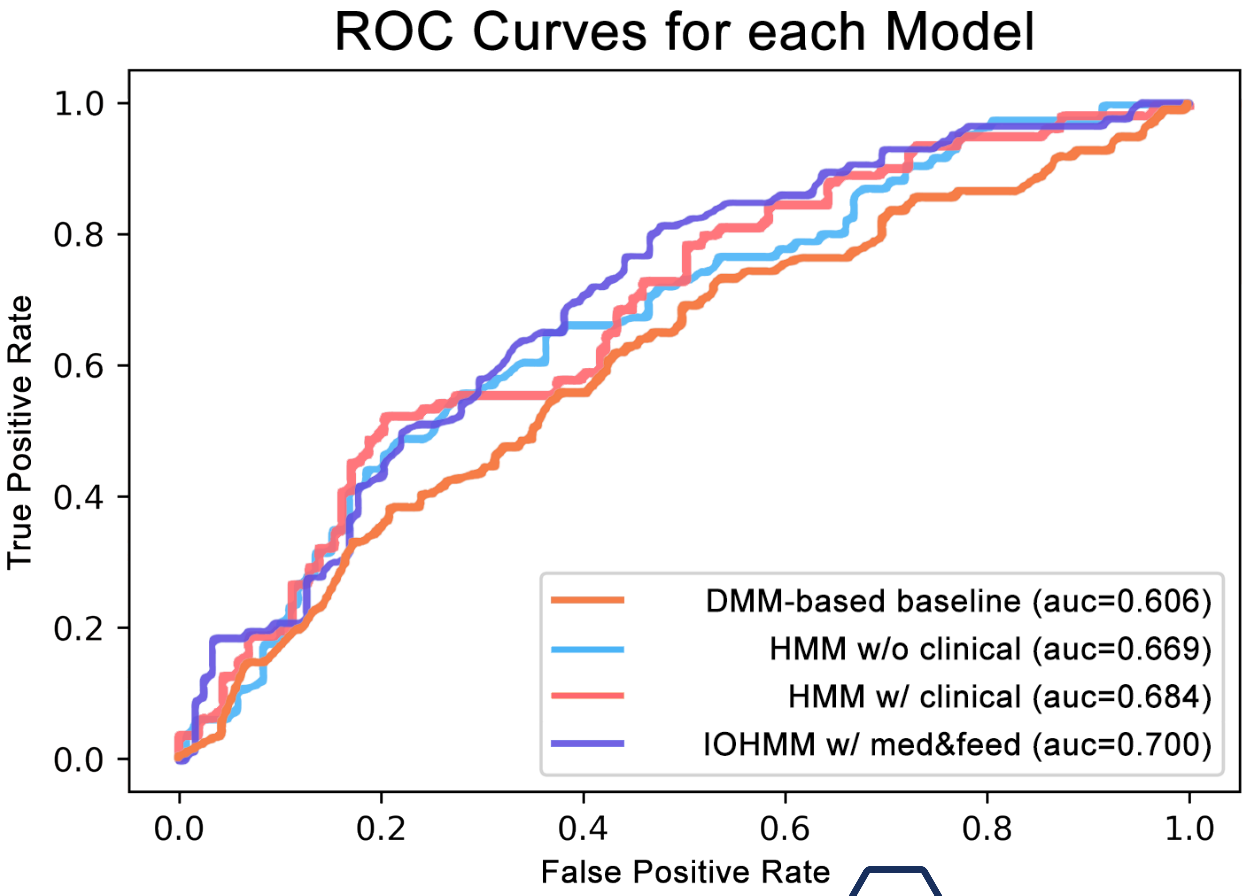
Week 28
Week 29
Week 30
Week n
Week 37

● "Hidden" Preterm Gut Community Types
● Observable States – Taxa, Clinical Measures
● Perturb-able Input Features
○ Missing Data

Prediction Results

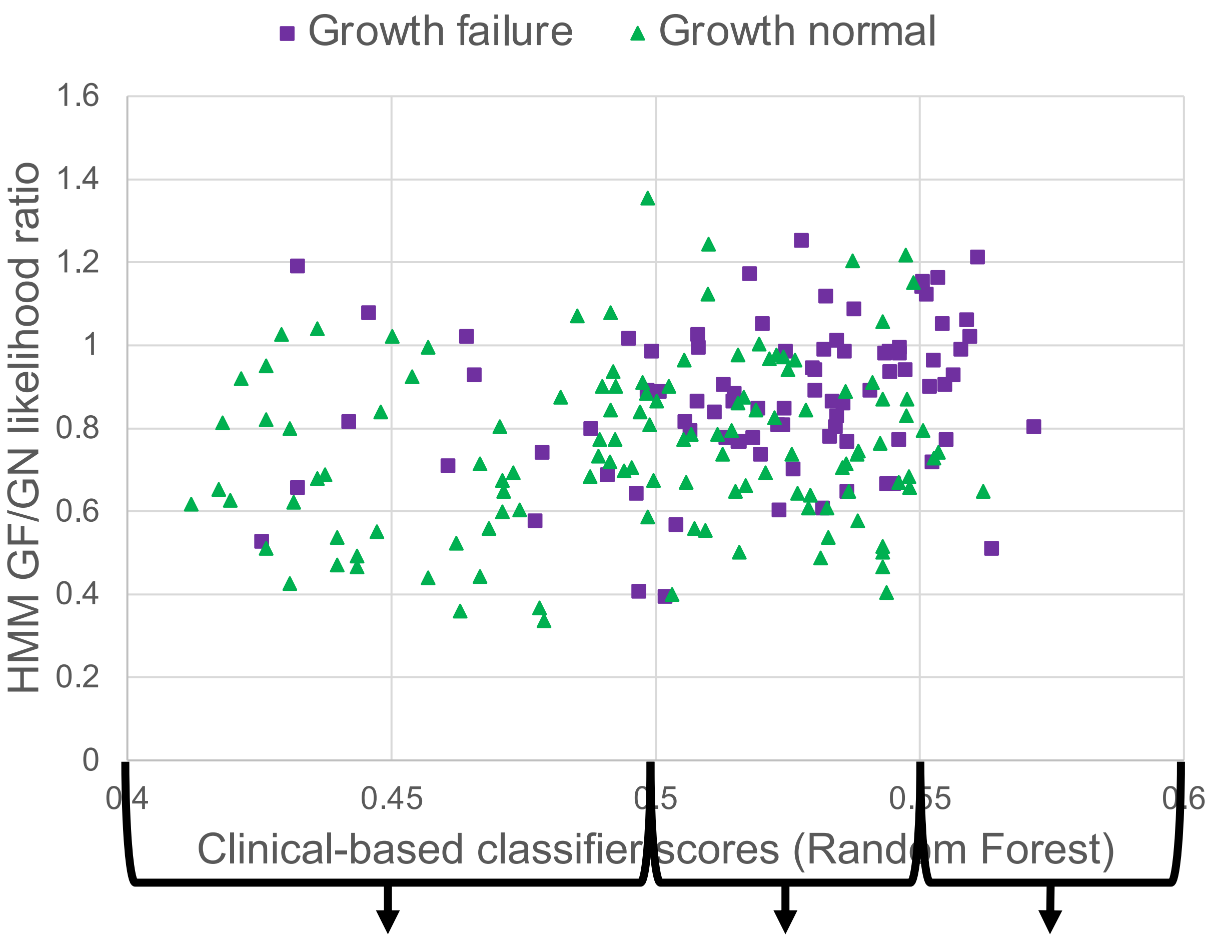
IOHMM with medication and feeding outperformed all HMM-based models

In silico perturbation analysis now possible to observe which factors influence growth failure and other outcomes – ongoing research



PAS
Pediatric Academic Societies

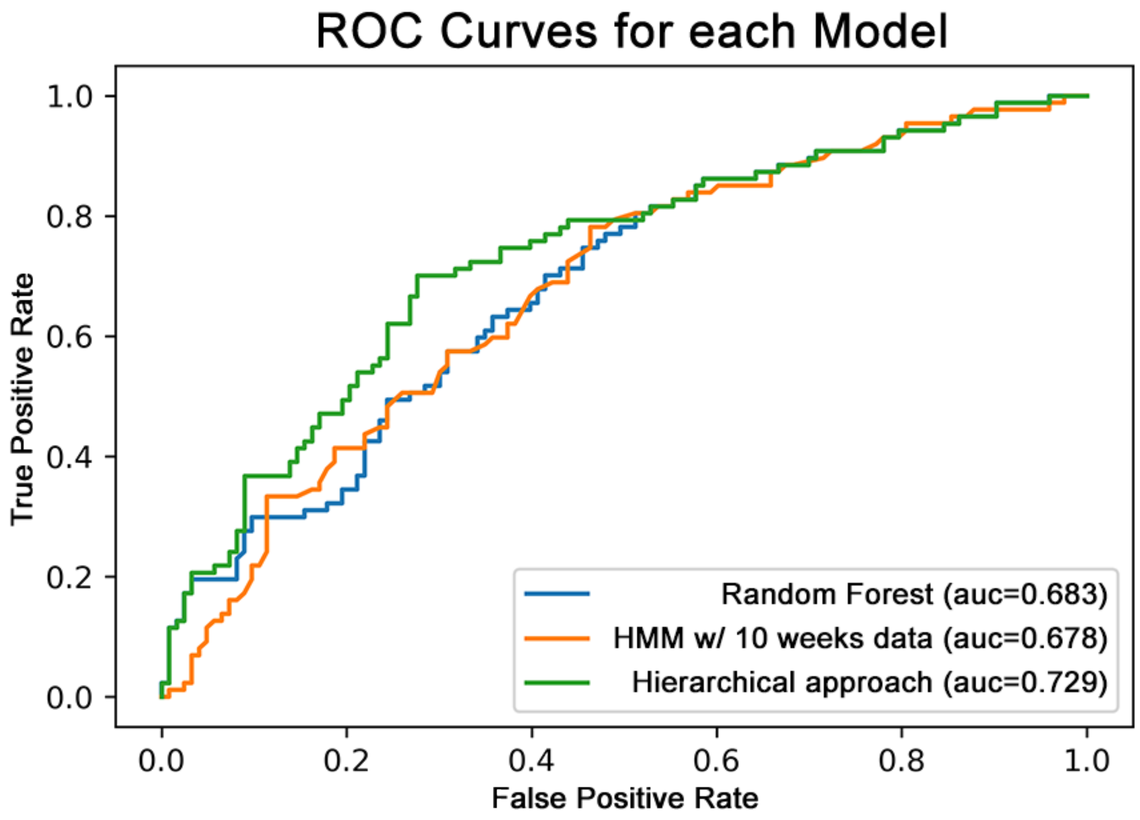
Two-stages Hierarchical Approach



RF
s ≤ 0.50

HMM
0.50 < s < 0.55

RF
s ≥ 0.55



Poster # EP-188 for details on clinical-based ML predictors

PAS
Pediatric Academic Societies

Summary

- Graphical models can successfully integrate longitudinal clinical and longitudinal microbiome data.
- This data and population remains nuanced and challenging, but we demonstrate increasing accuracy with more sophisticated models and approaches.
- Our best performing classifiers are hierarchical. Models based only on clinical data perform well apart from a confused middle population. Incorporating the microbiome via our HMM improves performance for the more difficult to predict cases.
- In silico perturbation of input factors is now possible with these models and research looking to identifying successful interventions is underway.

Acknowledgements

Carnegie Mellon University

- Ziv Bar-Joseph, PhD
- Jose Lugo-Martinez, PhD
- Siwei Xu

Astarte Medical

- Tracy Warren
- Arti Tandon, PhD
- Justine Levesque
- Daniel Gallagher

Newcastle University

- Chris Stewart, PhD
- Janet Berrington, MBBS, BMedSci, MD
- Nicholas Embleton, MD, MBBS, FRCPCH, BSc
- Gregory Young, PhD

Brigham & Women's

- Katherine Gregory, RN, PhD

Washington St. Louis

- Misty Good, MD

PAS
Pediatric Academic Societies

Integrating Longitudinal, Clinical, and Microbiome Data to Predict Growth Failure in Preterm Infants

Presenter Name: David Genetti
Institution: Astarte Medical
Email: dave@astartemedical.com

PAS
2021 VIRTUAL

Astarte Medical

Systems Biology Group