1. **VE303 colonization and species diversity analysis [Folder name – Figure\_1\_5A]**

Central R Script: VE303-Ph2\_TimeToEvent\_Engraftment\_Recurrence\_loops.Rmd

**Main text analysis:**

* Total VE303 strain detection and abundance/individual strain abundance over time across treatment groups, VE303 strain prevalence over the first 2 weeks in Dosed groups, KM curves for VE303-08 high vs low colonized subjects (Figure 1). Scripts called - "plot.detection.allmark.R", "plot.clean.abund.ve303\_det\_response.R", "KM exposure recurrence quartile loops.R".
* Diversity over time across treatments (Figure 5A). Scripts called: "plot.clean.abund.ve303\_det\_response.R".

**Extended data analysis:**

* Stacked bar charts of VE303 strain detection for individual subjects across treatment groups, mean strain abundance over the first 4 weeks per strain per treatment group. (Extended Data Figure 2). Scripts called - "plot.detection.allmark.R".
* Age violin plots for recurrent vs non-recurrent VE303 recipients. (Extended Data Figure 5B). Scripts called - "plot.detection.allmark.R".
* Brickplot showing colonization in all VE303 recipients. (Extended Data Figure 9). Scripts called - "plot.detection.allmark.R".

**Supplementary data analysis:**

* LME Models for VE303 detection and abundance across treatment groups up to Day 14, HD vs Placebo, LD vs Placebo, and HD vs LD; t-test and Wilcoxon test for HD vs LD at Day 14. (Table S1). Scripts called – VE303-Ph2-Dose-Response.RMD.
* VE303 strain detection summary at scheduled timepoints. (Table S2). Scripts called - "plot.detection.allmark.R".
* Cox models analysis for all VE303 strains. (Table S3). Scripts called - "KM exposure recurrence quartile loops.R".
* LME Models for change in shannon diversity and gini-simpson diversity over time and across treatment groups (Table S5). Scripts called – central script.

1. **Random forest analysis: multi-omic features predicting VE303 colonization. [Folder name – Figure\_2]**

**Main text analysis:**

* Random forest modeling to identify predictors of VE303 colonization using screening datasets - [scripts - "rf\_predict\_colonization.R”]; fit random forest models to infer feature importance.
* Generate the heatmap summary of important predictors of colonization using RF model results at screening. (Figure 2 heatmap). Script - "RF colonization heatmap.Rmd”.

**Extended data analysis:**

* Generate summaries of model performance for predicting colonization per-strain using screening timepoint datasets. (Extended data Figure 3). “RF script\_model\_colonization”: " rf\_importance\_colonization.R".

1. **Analysis of stool metabolites: SCFA, BA, and vancomycin over time. [Folder name – Figure\_3\_6A]**

Central R Script: VE303-Ph2-Metabolite-Analysis\_Vanco and Abx.Rmd

**Main text analysis:**

* Plotting VE303 strain exposure for subjects treated with vancomycin vs fidaxomicin over the first 2 weeks post-dosing (Figure 3A).
* Plotting Shannon diversity change from screening to Day 1 in VE303-dosed subjects treated with vancomycin vs fidaxomicin (Figure 3B).
* Plotting vancomycin concentration over time in dosed subjects (Figure 3C).
* Plotting correlations between VE303 strain colonization and residual vancomycin concentrations during the first 2 weeks after dosing (Figure 3D).
* Plotting mean concentrations of metabolites significantly associated with recurrence (Figure 6A).

**Extended data analysis:**

* Heatmap of mean metabolite concentrations of all primary/secondary bile acids and SCFAs from D1 to D14 in recurrent vs non-recurrent subjects(Extended data figure 7D).

**Supplementary data:**

* Plotting relationship between VE303 strain colonization and number of days of SoC antibiotics (Figure S1).
* LME modeling to determine the effect of antibiotics pre-treatment type, antibiotic treatment length, and residual vancomycin concentration on VE303 colonization (Table S4).

1. **Random forest analysis: multi-omic features predicting recurrence. [Folder name – Figure\_4]**

**Main text analysis:**

* Random forest modeling to identify predictors of recurrence using screening, Day 1, Day 7, and Day 14 datasets - [scripts -

“rf\_predict\_recurrence.R”: fit random forest models to infer feature importance.

* Generate barplots summary of important features predictive of clinical outcome using RF models at screening. (Figure 4 importance). Script - "rf\_importance\_recurrence.R”

**Extended data analysis:**

* Generate summaries of model performance for predicting recurrence using screening, Day 1, Day 7, and Day 14 datasets. (Extended data figure 3C). Script - : “rf\_importance\_recurrence.R”.
* Generate heatmap summaries of important features associated with clinical outcomes using screening, Day 1, Day 7, and Day 14 datasets. (Extended Data Figure 4A). Script - "make\_heatmap”
* Infer features predictive of recurrence, summarize model performance, and generate heatmap summaries of important features associated with early (by Day 14) recurrence. (Extended Data Figure 3D, Extended Data Figure 4B). Scripts in sub folder - "Figure\_S4\_early\_recurrence”.
* Plot cytokines found to associated with recurrence (Extended Data Figure S5A) and correlation plot for Age vs MIP-1b (Extended data Figure S5C). Script - : “rf\_importance\_recurrence.R”.

1. **LME analysis of endogenous stool microbiome and associations with recurrence and VE303 abundance [Folder name – Figure\_5BC]**

Central R Script: VE303-Ph2-Dose-Response Endogenous Microbes.Rmd

**Main text analysis:**

* Volcano plot showing species significantly associated with recurrence (Figure 5B)
* Ranked effect-size barplots showing top species associated with clinical outcomes (Figure 5C).

**Extended data analysis:**

* Ranked effect-size barplots showing top taxa associated with clinical outcomes and VE303 abundance at all taxonomic levels (Extended data figure 6).

1. **LME analysis of stool SCFA and bile acids and correlations with host endogenous bacteria [Folder name – Figure\_6B]**

**Main text analysis:**

* Fit LME models to identify metabolites significantly correlated with microbes. Generate heatmap summaries of significant metabolite-taxon pairs (Figure 6B). Script - " lme\_metab\_microbiome.R”.

**Extended data analysis:**

* Boxplots of total primary bile acids, secondary bile acids, and short-chain fatty acids over time across treatment groups (Extended data figure 7A,B,C). Scripts in sub folder - "Figure\_S7\_ ABC”; script - “plot\_SCFA\_BA\_comb.R”.
* Heatmaps of bile-acid genes (annotated using ShortBRED) found to be associated with treatment or recurrence outcomes, or with endogenous bacteria. (Extended data figure 8). Scripts in sub folder - "Figure\_S8”; scripts - “lme\_sba\_genes”, “lme\_sba\_microbes”.