CommScores: Predictive metrics of microbial communities via biochemical interactions

Andrew Freiburger[1], Dileep Kishore[2], Priya Ranjan[2], **Christopher Henry**[1]

[1]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL ; [2]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN ; [2]

Contact Us

INTRODUCTION

Microbial communities are complex consortia with profound potential for industrial bioproduction, therapeutics, agricultural tools, and basic biological research. The chemical interaction between these microbes increase combinatorially with community size and therefore remain experimentally intractable for all but a few experimental methods. Computational tools can better accommodate these exponential interaction networks, yet current methods focus on cross-feeding and co-growth simulations and fail to capture other biological dimensions in which microbial members interact, such as resource competition, or, in a few biological dimension analysis tools such as the SMETANA suite (Species METabolic interaction ANAlysis), miss other biological dimensions like as functional similarity and isolate growth capacity.

We, therefore, developed **CommScores** as a suite of both curated scores from literature and novel scores of new interaction dimensions in microbial communities. We conserved the original interaction dimensions of published scores while improving their computational methods with greater accuracy, efficiency, and transparency. Our novel scores introduce metrics for new dimensions of 1) functional similarity from genome ontologies (ecological similarity); 2) categorizing the biological interaction type by contrasting growth of organisms as community members versus as isolates; 3) contrasting the isolate growth rates of members (potential outperformance); and 4) evaluating costless excreta (possible cross-feeding). **CommScores** is offered as a Python package and as a KBase Application for seamless use in a web environment.

We exemplify **CommScores** by predicting member interactions in the *Arabdopsis* plant microbiome. Recapitulation of experimental data identified that metabolic support – the donation of cross-feeding metabolites in the *MIP*, the contributing of accessible nutrients to the environment via the *CIP*, or the capacity for supporting a member from cytoplasmic compounds via the *BSS* – principally governs interactions in the *Arabidopsis* microbiome. This elucidates basic biochemical dynamics that can be leveraged to rationally engineer the community for diverse applications in bioproduction, therapeutics, and agriculture.

Organisms
Metabolites
Environments

Predicted outcomes

E. coli   *P. fluorescens*

Acetate
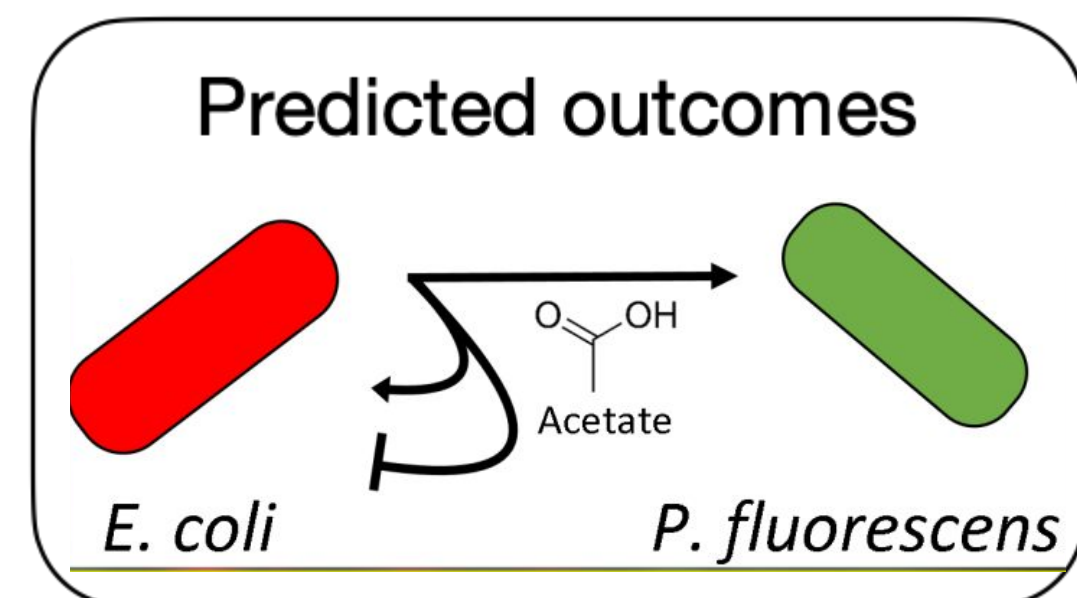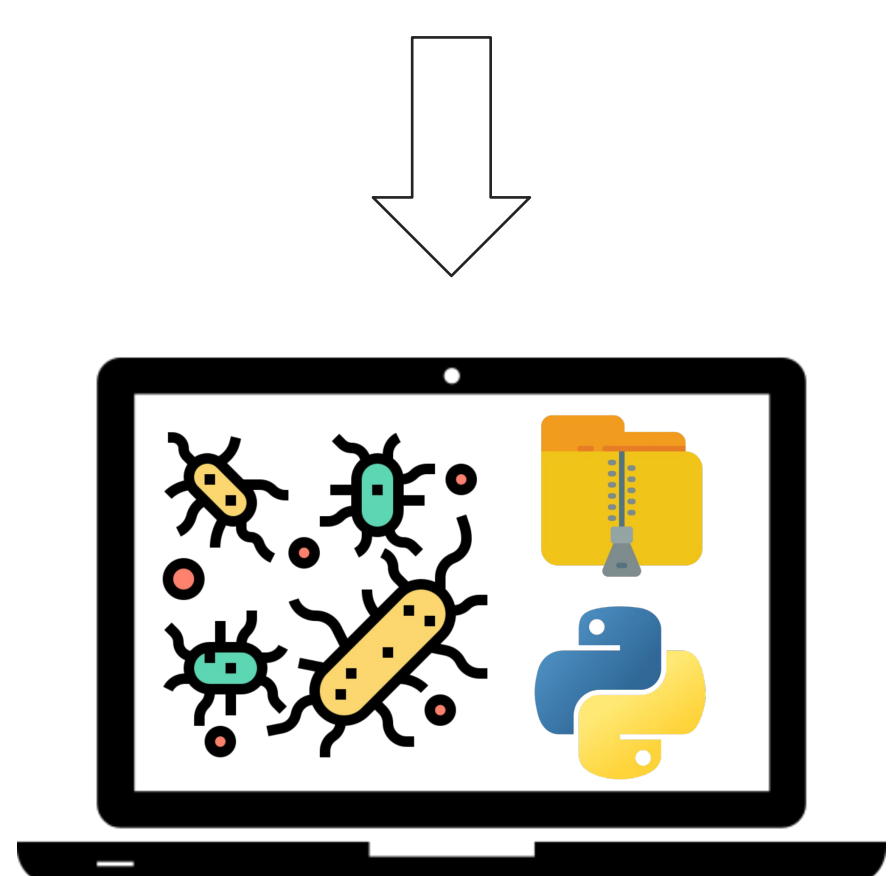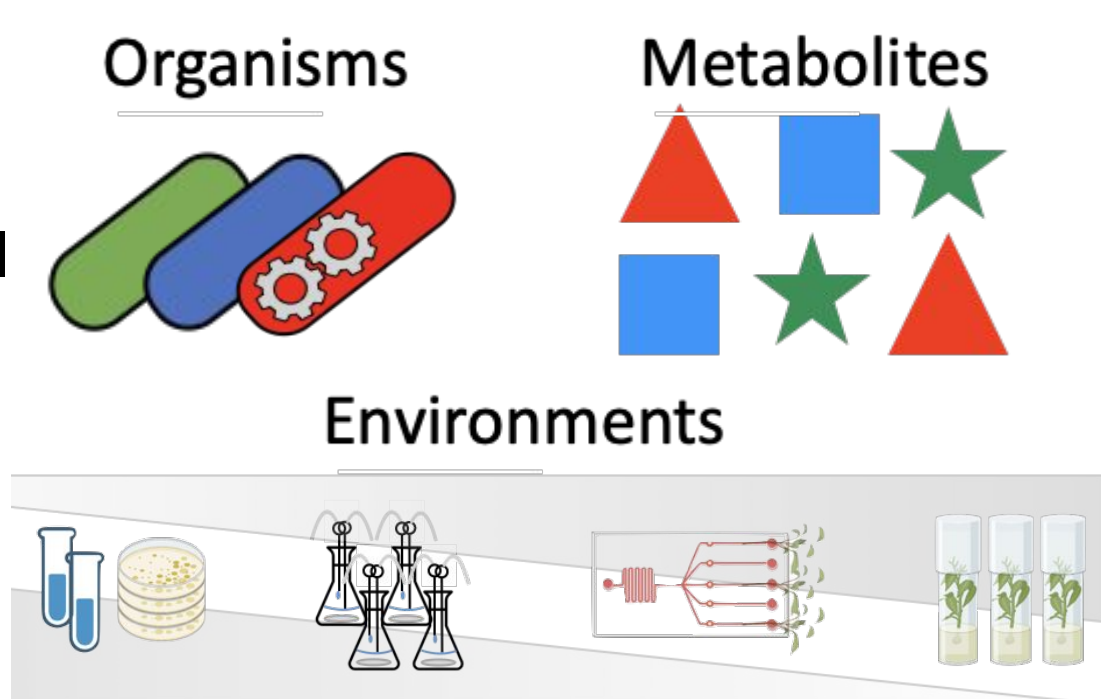
CommScores

Novel

The **FS** (*Function Similarity*) score evaluates the genetic similarity between two members, and thereby approximates ecological overlap and resulting competition between the members.

$$FS = \left(\frac{O_i \cap O_j}{O_i \cup O_j} \right)$$

The **GYD** (*Growth Yield Difference*) score evaluates the relative difference in growth yield between two members, and thereby approximates the propensity for one member to dominant the environment.

$$GYD = \frac{abs(G_i - G_j)}{min(G_i, G_j)}$$

The **BIT** (*Biological Interaction Type*) score categorizes the member interaction according to the isolate growth relative to the member growths in the community environment. This is the most comprehensive score, but does not necessarily capture all of the interaction dimensions.

1. *competition* $PC_i < th_{neg}$ & $PC_j < th_{neg}$
2. *parasitism* $PC_i > th_{pos}$ & $PC_j < th_{neg}$ or $PC_i < th_{neg}$ & $PC_j > th_{pos}$
3. *amensalism* $PC_i < th_{neg}$ & $PC_j \approx 0$ or $PC_i \approx 0$ & $PC_j < th_{neg}$
4. *neutral* $th_{neg} < PC_i < th_{pos}$ & $th_{neg} < PC_j < th_{pos}$
5. *commensalism* $PC_i > th_{pos}$ & $th_{neg} < PC_j < th_{pos}$ or $th_{neg} < PC_i < th_{pos}$ & $PC_j > th_{pos}$
6. *mutualism* $PC_i > th_{pos}$ & $PC_j > th_{pos}$

The **CIP** (*Costless Interaction Potential*) score determines the set of all excretions at optimal growth of the members, and thereby evaluates the breadth of media additions from the members that do not incur a fitness cost.

$$CIP = |CE_i \cup CE_j|$$

Curated

Two of the most prominent community scoring metrics in literature are the **MRO** (*Metabolic Resource Overlap*) and the **MIP** (*Metabolic Interaction Potential*) scores, which are notably apart of the *SMETANA* suite of scores. These scores capture two of the most putatively consequential interaction dimensions: nutrient competition and cross-feeding, respectively.

$$MIP = M_{non-interacting} - M_{interacting}$$

$$MCtI_i = \left(\frac{|M_i \cap M_j|}{|M_i|} \right)$$

The **BSS** (*Biosynthetic Support Score*) is a directional assessment of parasitic capacity of one member against the other member. The averaged value of both directions, by contrast, estimates the complementary support that each member can potentially provide.

$$BSS_i = \left(\frac{|M_i \cap R_j|}{|M_i|} \right)$$

The **PC** (*Potential Cooperation*) score determines the ratio of community growth to the sum of member growths, and thereby evaluates whether the whole is greater than the sum of the parts.

$$PC_i = \frac{G_{i,comm}}{G_{i,mono}}$$

$$PC_{comm} = \frac{G_{comm}}{G_{i,mono} + G_{j,mono}}$$

Validation with *Arabidopsis* microbiome

We exemplified CommScores by recapitulating an experimental study of pairwise interactions in an *Arabidopsis* microbiome (Schafer et al., *Science,* **2023**). The experimental data (sourced from Table S5) was labeled as positive or negative interactions as the figure at right.
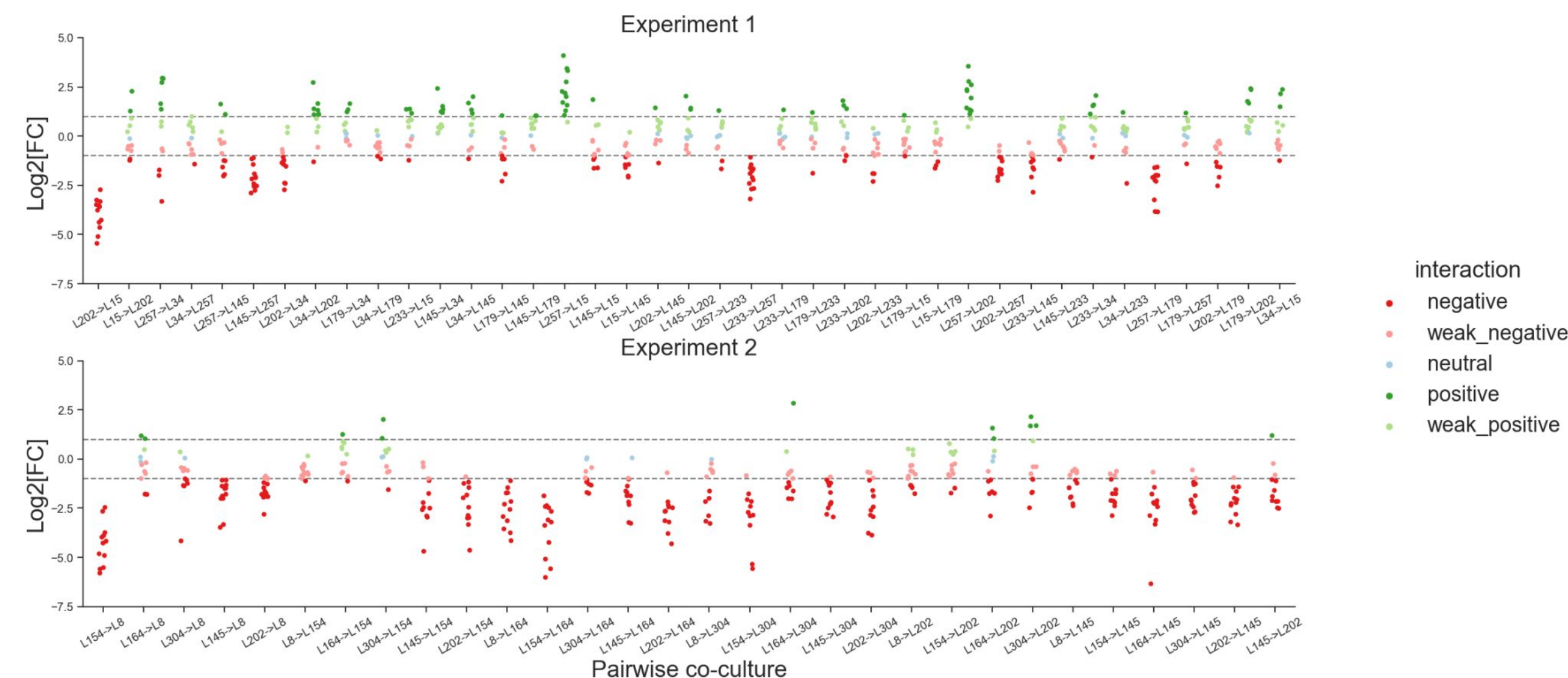
The ability of each CommScores metric to discern these interactions is evaluated in the following figure, where the donor-focused scores of costless MIP (mip_c1) and BSS (bss1) exhibited the greatest significance.
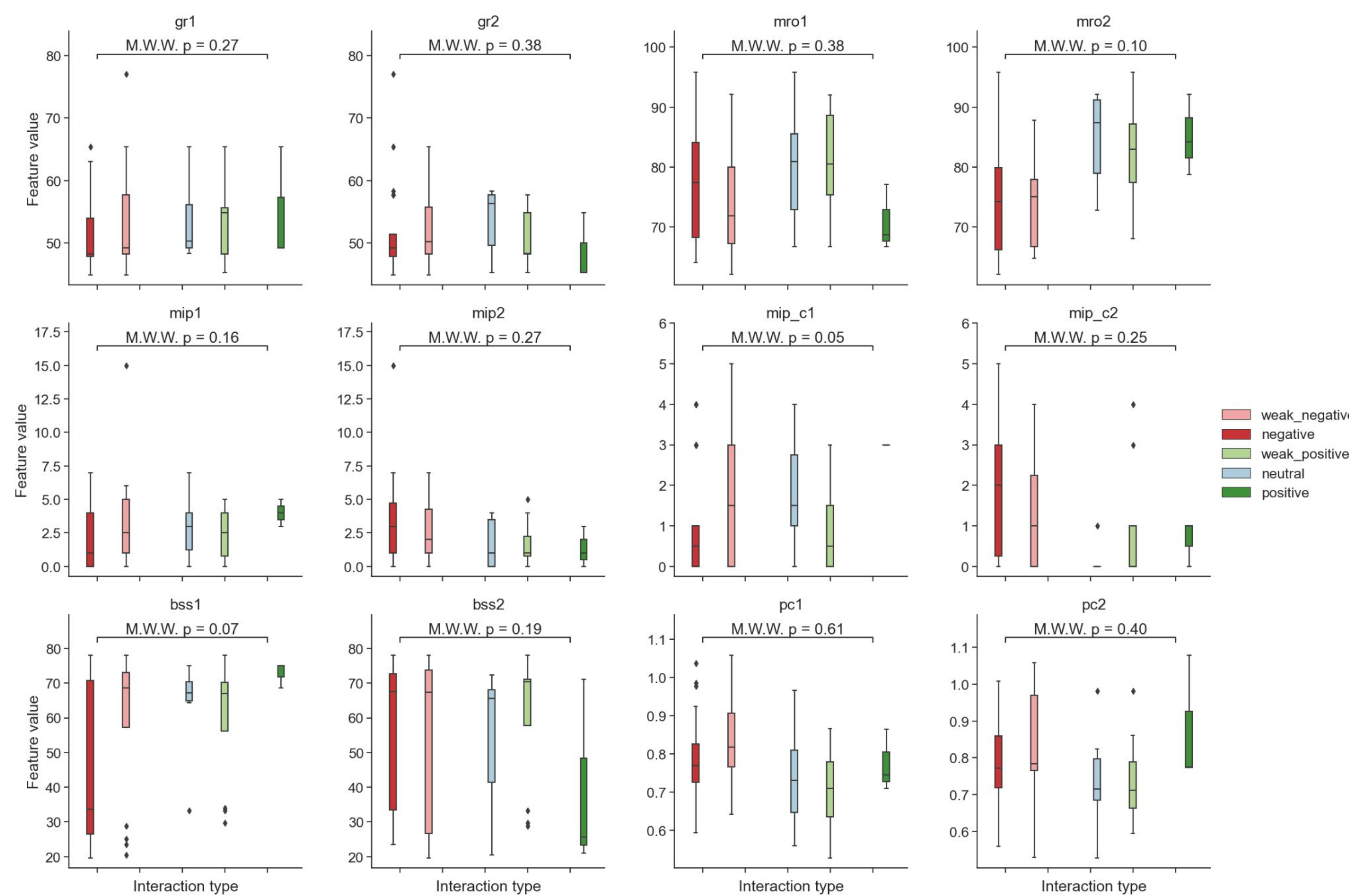
The correlation of each CommScores metric was also compared with predicting the experimental interactions, where the cross-feeding scores (MIP and BSS) elicited the greatest predictive power for interaction between the members.
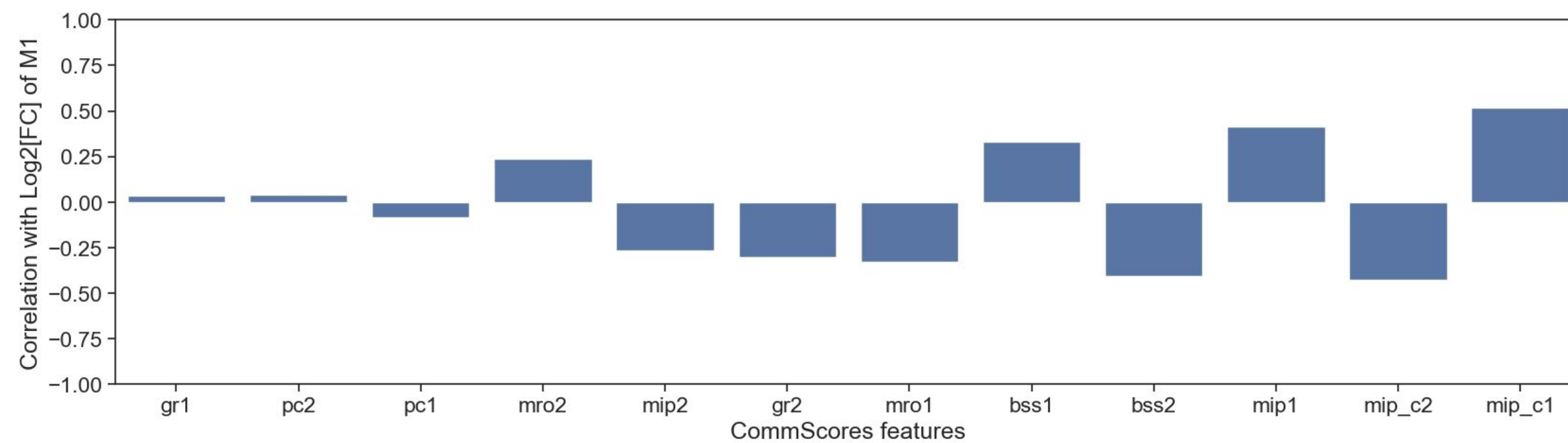
FUTURE WORK

The fundamental insights from **CommScores** will be applied to guide our experimental engineering efforts, which will invite new scientific possibilities and research applications.

We further intend to expand CommScores with additional scores: namely curation of the Metabolic Support Index (MSI) and Community Support Index (CSI) that topologically assess the ability of members to support each other from non-exchangeable metabolites, which is a subtle variation on the BSS score. We intend as well to develop a few novel scores that capture new biological dimensions, such as pH perturbation from each member's growth and the stability of relative abundance in the pairwise coculture to examine the significance of competitive dynamics. Finally, we intend to refine the **CommScores** KBase Application with more features that facilitate gleaning information from **CommScores** results for investigators..

OAK RIDGE National Laboratory

KBase PREDICTIVE BIOLOGY

Argonne NATIONAL LABORATORY