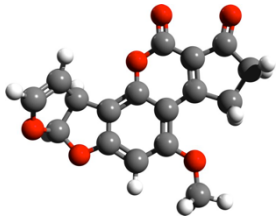
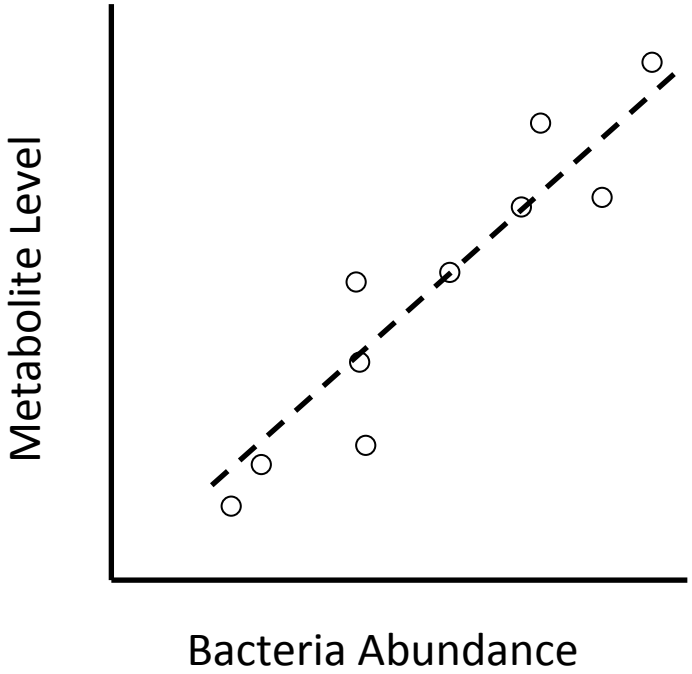
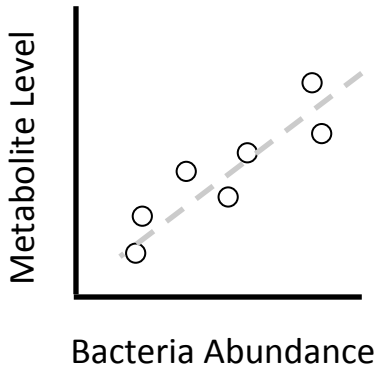
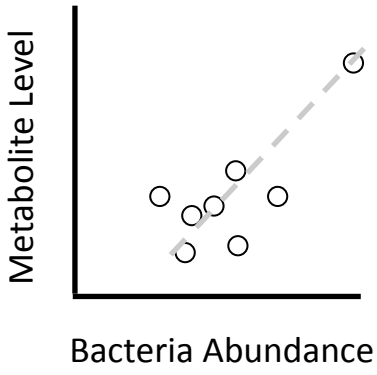
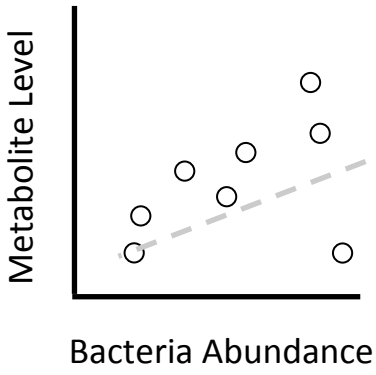
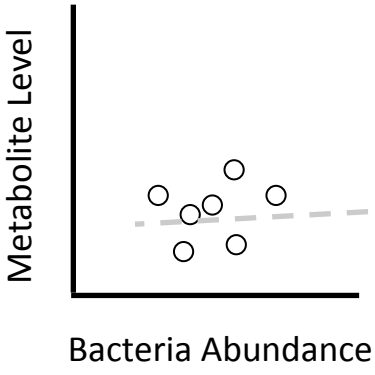


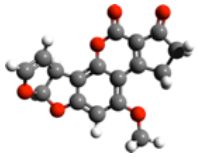


Correlations Under the Influence (CUTIE)

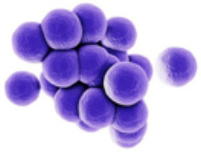
Code Review 4.8.19

Kevin Bu

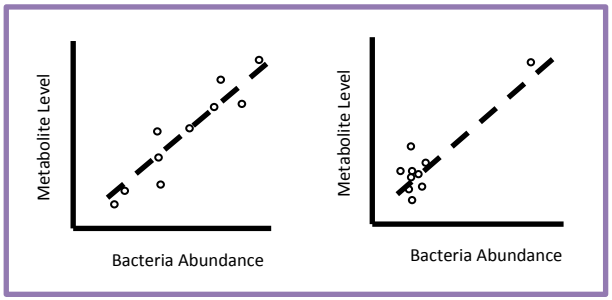
	True Correlation	False Correlation
Significant Correlation ($p < 0.05$)	✔ True Positive (Non-CUTIE) 	✖ False Positive (CUTIE) 
Insignificant Correlation ($p > 0.05$)	✖ False Negative 	✔ True Negative 

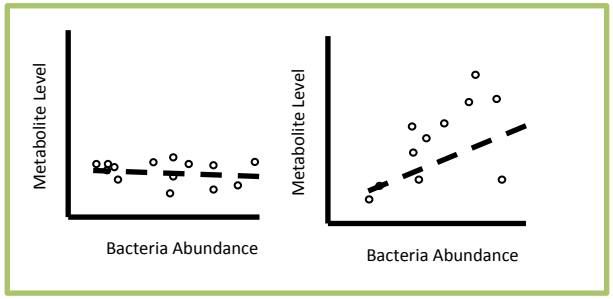


Metabolite	Bacteria
Glycine	Ecoli
Citrate	Shigella
...	...



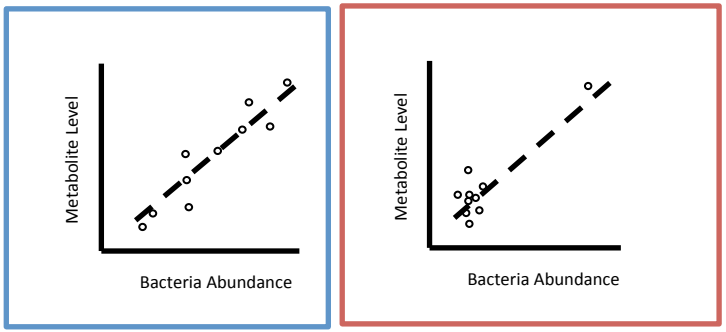
Compute Pairwise Correlations

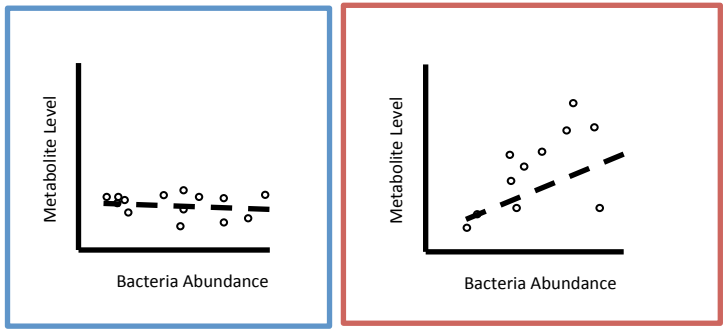




Examine all correlations with $p < 0.05$*

Examine all correlations with $p > 0.05$*





Perform CUTIE analysis

Perform rCUTIE analysis

Differentiate **non-CUtIe's (TP)** from **CUtIe's (FP)**

Differentiate **TN's** from **FN's**

Milestone 1: Code Review 4.8.19

(1) Tutorial (in the README)

(2) Unit tests for all core functions

CUTIE file structure

- **CUtIe/**
  - cutie/
    - __init__.py
    - output.py
    - parse.py
    - utils.py
    - statistics.py
  - README.md
  - setup.py
  - scripts/
    - calculate_cutie.py
    - configs/
  - tests/
    - config_defaults.ini
    - data/
      - Mapping.Pneumotype.Multiomics.RL.NYU.w_metabolites.w_inflamm.txt
      - otu_table_MultiO_merged___L6.txt
      - n=50,alpha=0.6.csv
    - lungpt_1pc_point_unit_test_kkc1fdr0.05/

Issues

- 1. Parse.py (Nan)
  - Check PEP8 compliance (md5Checksum, parse_input, process_df, read_taxa, parse_minep, parse_config)
- 2. Output.py (Zach)
  - 2a. Unit tests for generation of matrices (print_matrix, print_Rmatrix, print_true_false_corr, report_results, generate_pair_matrix)
  - 2b. Check for PEP8 compliance, potentially shorten graphing functions (graph_subsets, generate_dfs, plot_dfs, plot_figure, plot_pdist, plot_logp_and_logpfold, plot_corr, plot_corr_sets, diag_plots)
  - 2c. Unit tests for log handling (init_log, write_log)
- 3. Utils.py (David)
  - Unit tests for auxiliary functions (return_indicators, remove_nans, initialize_stat_dicts, get_param, calculate_intersection, indicator, init_var_indicators)
- 4. Statistics.py (4a: David, 4b-e: Kevin)
  - 4a. Unit tests for zero-handling (multi_zeros, multi_replacement, zero_replacement, log_transform)
  - 4b. Unit tests for initial stats (assign_statistics, initial_stats_SLR, initial_stats_MINE, set_threshold, get_initial_corr)
  - 4c. Unit tests for pointwise stat functions (resample1_cutie_pc, resample1_cutie_sc, cookd, dffits, dsr, return_influence, pointwise_comparison, calculate_FP_sets, compute_pc, compute_sc, compute_kc)
  - 4d. Unit tests for MINE-specific handling (compute_mine, str_to_pvalues, binarySearchBins)
  - 4e. Unit tests for CUTIE, bootstrap, and jackknife functions (updatek_cutie, cutiek_true_corr, evaluate_correlation_k, update_rev_extrema_rp, jackknifek_cutie, bootstrap_cutie, resamplek_cutie, get_pCI, test_CI)