**Pathway\_level\_flux\_analysis**

This folder contains the core scripts and representative outputs used for pathway-level flux analysis across pFBA and FVA-sampling approaches in Drosophila muscle-GEMs.

All analyses were performed using MATLAB, and visualization or overlap analyses were completed in both MATLAB and R.

**Contents**

• MATLAB scripts: Core analysis scripts for performing pFBA and FVA-sampling, computing pathway flux indices (PFI), comparing pathway-level fluxes across methods, conducting statistical tests (Z-test), and generating annotated Excel summaries and visualizations.

• R scripts: Visualization and overlap analysis scripts for comparing pFBA, and FVA-sampling results, performing threshold-based filtering, and generating dot plots highlighting significant pathway flux differences and overlaps.

**Notes**

Due to file size constraints, the sampling result files (HSD.mat, NSD.mat) are not included. Only representative .mat files, summary tables, and result figures are provided for illustration. For access to the full dataset or additional information, please contact: [sunjin\_moon@hms.harvard.edu](mailto:sunjin_moon@hms.harvard.edu)

**Summary of scripts**

1. **script\_01\_run\_pFBA.m**
2. Purpose:  
   Perform parsimonious flux balance analysis (pFBA) on NSD and HSD.
3. Input:
   1. model\_out\_cbra\_u.mat – contains preprocessed NSD and HSD muscle-GEMs
   2. COBRA Toolbox functions (e.g., optimizeCbModel, convertToIrreversible, changeRxnBounds)
4. Output:
   1. Folder: 1\_pFBA/
      1. pFBA\_results.mat – optimized fluxes, objective values, total and average fluxes per model
      2. model\_constrained\_out.mat – constrained models after applying reaction- and gene-specific limits
5. Note:
   1. This is the same as the previous pFBA analysis.
6. **script\_02\_PFI\_pFBA.m**
   1. Purpose:  
      Compute pathway flux and compare pathway-level flux statistics between NSD and HSD muscle-GEMs using pFBA results. The script summarizes mean, median, quartile, and total pathway fluxes, calculates log₂ fold-changes (HSD/NSD), identifies top up- and down-regulated pathways, and visualizes results with bar and grouped bar plots.
   2. Input:
      1. 01\_results/model\_constrained\_out.mat — constrained NSD and HSD muscle-GEMs with subsystem annotations
      2. 01\_results/pFBA\_results.mat — pFBA-derived flux distributions for each model
   3. Output:
      1. Folder: 2\_pFBA\_PFI/ containing
         1. pathway\_flux\_analysis.xlsx — comprehensive pathway-level summaries (mean, median, mode statistics, ratios, and reaction counts)
         2. pathway\_flux\_analysis.mat — MATLAB structures of pathway flux matrices and comparison tables
         3. Top\_Bottom\_Pathways\_Log2FC.png/.fig — bar plots of top 5 up- and down-regulated pathways
         4. reaction\_analysis\_top\_bottom\_pathways.xlsx — reaction-level comparisons within top and bottom pathways
         5. Grouped\_Bar\_Mean\_SD\_Pathways.png/.fig — grouped bar plots with error bars comparing HSD vs NSD fluxes for top/bottom pathways
7. **script\_03\_PFI\_sampling.m**
   1. Purpose:  
      Compute pathway flux and compare pathway-level flux statistics between NSD and HSD muscle-GEMs using sampling results.
   2. Input:
      1. NSD.mat, HSD.mat — flux sampling outputs containing samples.points and samples.subSystems
      2. Stored under: C\_sampling/
   3. Output:
      1. Folder: 3\_sampling\_PFI / containing
         1. pathway\_flux\_analysis.xlsx — sorted mean and median comparisons with log₂FC and SD
         2. Console summary of valid pathways and z-score–compatible pathway statistics
8. **script\_04\_compare\_all.m**
   1. Purpose:  
      Compile pathway-level flux comparison results from pFBA and FVA-sampling analyses, classify up- and down-regulated pathways based on log₂ fold-change (HSD/NSD), and generate binary presence matrices for UpSet or overlap visualization in R.
   2. Input:
      1. 2\_pFBA\_PFI /pathway\_flux\_analysis.xlsx — pFBA mean pathway flux results
      2. 3\_sampling\_PFI/pathway\_flux\_analysis.xlsx — FVA-sampling mean pathway flux results
   3. Output:
      1. Folder: 4\_compare\_all/ containing
         1. UpSet\_pathway\_data.xlsx — binary matrices for up/down-regulated pathways across pFBA and FVA-sampling (Up\_Regulated, Down\_Regulated sheets)
         2. Up\_Lists and Down\_Lists — pathway lists by method for direct inspection
         3. upset\_data.mat — MATLAB workspace with full up/down pathway sets and binary matrices for R UpSet plot generation
9. **script\_05\_PFI\_z\_test.m**
   1. Purpose:  
      Perform Z-test–based statistical comparison of sampling-derived pathway fluxes between NSD and HSD muscle-GEMs to identify pathways showing significant flux differences.
   2. Input:
      1. NSD.mat, HSD.mat – flux sampling results for NSD and HSD muscle-GEMs
      2. subsysAll.mat – list of metabolic subsystem definitions (pathway annotations)
   3. Output:
      1. 5\_PFI\_zTest\_sampling / folder containing
         1. Z\_test\_sampling\_results\_Bonferroni.xlsx – full pathway-wise Z-test results
         2. Z\_test\_sampling\_results\_Bonferroni\_sorted.xlsx – sorted by log₂ fold-change
         3. \*\_cleaned.xlsx versions without NaNs
         4. NonSignificant\_Pathways\_sampling\_Bonferroni\_0.05.txt – list of nonsignificant pathways
         5. Console summary of tested pathways, significance counts, and descriptive statistics
10. **script\_06\_R\_visualization\_threshold.R**
    1. Purpose:  
       Visualize pathway-level flux differences (HSD vs NSD) using Z-test results with a minimum reaction count threshold. Generates publication-ready dot plots highlighting significant pathways, significance magnitude, and overlap with pFBA/FVA predictions.
    2. Input:
       1. Z\_test\_results\_Bonferroni\_sorted\_cleaned.xlsx
       2. UpSet\_pathway\_data.xlsx — three-way overlap data from 6\_compare\_all/ (pFBA FVA-sampling)
    3. Output:
       1. Folder: 6\_sampling\_PFI\_vis
          1. Z\_test\_results\_threshold\_filtered.xlsx — filtered results (≥ reaction threshold)
          2. Z\_test\_results\_below\_threshold.xlsx — excluded pathways (below threshold)
          3. dotplot\_significant\_log2FC.png/.svg — significant pathways (colored by direction)
          4. dotplot\_all\_pathways.png/.svg/.pdf — all pathways colored by -log10(p\_adj) and sized by reaction count, |Z-score|, or |flux difference|
          5. highlighted\_pathways\_pFBA\_FVA\_overlap.xlsx — significant pathways overlapping in pFBA & FVA analyses
          6. Visualization\_Data.xlsx — combined dataset with summary statistics, legend, and color annotation for figure reproduction