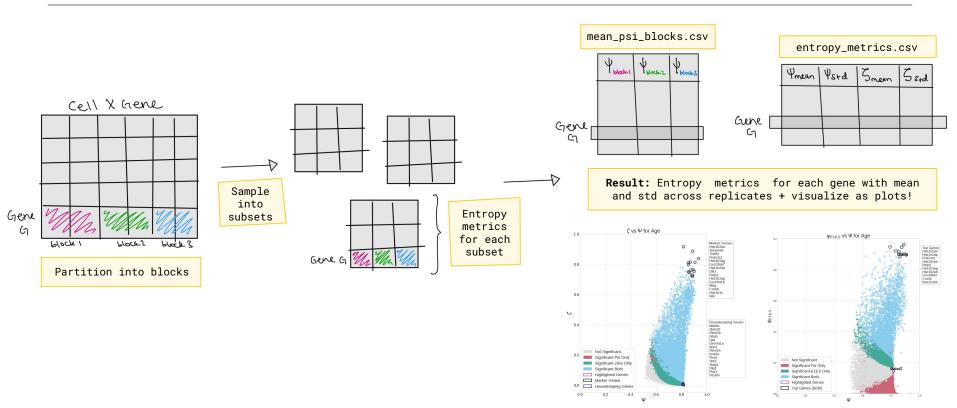


Nikhila P. Swarna September 18 2025

Pachter Lab
Division of Biology and Bioengineering
California Institute of Technology

.



light_ember

- one stop shop for generating entropy metrics and p-values

INPUT

- h5ad_dir (path to adata)
- partition_label (col in adata.obs)
- save_dir (path to save results)
- sampling= True
 - sample_id_col=None
 - category_col=None
 - o condition col=None
 - o num draws=100
 - save_draws = False
 - o seed = 42
- partition pvals=True
- block_pvals=False
 - block label=None
 - n_pval_iterations=1000
- n_cpus=1 (for parallel processing of sampling)



OUTPUT

- csv file in save_dir with all entropy metrics
- csv file in Psi_block_df folder with psi block
 - Separate file for pvals
 - Separate files for each partition
 - Alternate file names depending on sampling on or off.

generate_pvals

- manual access to generate pvals after initial investigation using light_ember

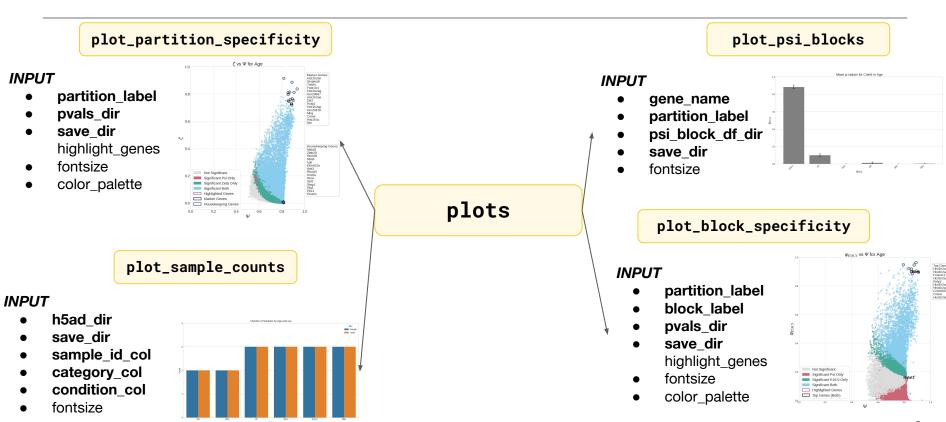
INPUT

- **h5ad_dir** (path to adata)
- partition_label (col in adata.obs)
- entropy_metrics_dir (path to light_ember output files)
- save_dir (path to save results)
- sample_id_col
- category_col
- condition col
- block_label=None
- seed = 42
- n_iterations=1000
- n_cpus=1



OUTPUT

- csv file in save_dir with entropy metrics and corresponding p-values and FDR q-values
 - Separate files for each partition



Defining entropy metrics for biological exploration

For a given gene in a count matrix that can be partitioned into r blocks (based on sex, strain, cell type, tissue, etc), we introduce **3 measures of specificity**:

| block r | block 3 | block 2 | block 1 |
|---------|---------|---------|---------|
| | | | |
| n | | | |

- Psi (Ψ)
- Psi_{block} (Ψ_{block})
- Zeta (ζ)

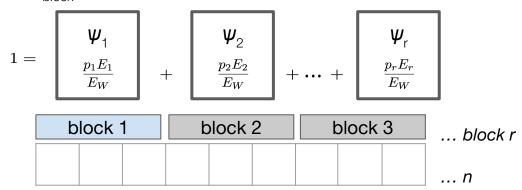
| Ψ | $oldsymbol{\psi}_{block}$ | ζ |
|---|---------------------------|----------------------------|
| Fraction of information explained by partitioning | Specificity to a block | Specificity to a partition |

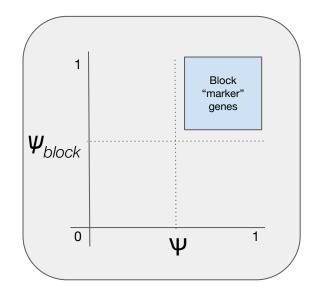
Ψ- Information Fraction by Partition

The fraction of information explained by using a particular partition on gene *g*'s counts is given by:

$$\Psi = \frac{E_W}{E_T} = 1 - \frac{E_B}{E_T}$$

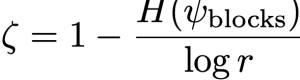
The Specificity of Information to Block, denoted by ψ_{block} , is the contribution of each block to Ψ :



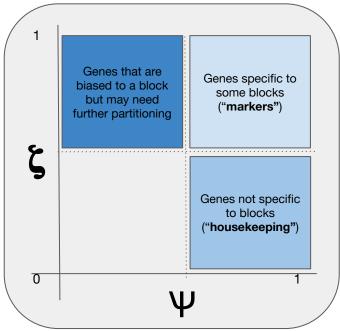


\(\) - Specificity of Information to Partition

The specificity of information to a partition (ζ) is given by:



Comparison of the SIB distribution to the uniform distribution



How to select category and condition

Category - Mouse strain Condition - Sex

| Sample | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|-----------------|------|-----|-------|------|----------|-----|--------|----------|---|-------------|----|----------|----|-----|-----|------|
| Mouse strain | | Str | ain A | | Strain B | | | Strain C | | | | Strain D | | | | |
| Sex | Male | е | Fen | nale | Ma | ale | Female | | | Male Female | | | Ma | ale | Fen | nale |

Number of unique draws

= (Number of replicates per category-condition group) $^{(Number of category-condition groups)}$ = $2^8 = 256$

One example draw:

| Sample | 1 | 3 | 5 | 7 | 9 | 11 | 13 | 15 | |
|--------------|------|--------|------|----------|------|--------|----------|--------|--|
| Mouse strain | Str | ain A | Stra | Strain B | | in C | Strain D | | |
| Sex | Male | Female | Male | Female | Male | Female | Male | Female | |

How to select category and condition

Category - Cell line Condition - Gene perturbation

| Sample | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|----------------------|-------|------|--------|------------|------|-------|-------------|------|--------|----------|------|-------|
| Cell line | | | Ce | ell line A | | | Cell line B | | | | | |
| Gene perturbation | Wildt | зуре | Overex | oression | Knoo | ckout | Wild | type | Overex | oression | Knoo | ckout |

Number of unique draws

= (Number of replicates per category-condition group) $^{(Number of category-condition groups)}$ = $2^6 = 64$

One example draw:

| Sample | 1 3 | | 5 | 7 | 9 | 11 | | |
|----------------------|-----|-----------|----|-------------|----|----|--|--|
| Cell line | | Cell line | A | Cell line B | | | | |
| Gene perturbation | WT | OE | КО | WT | OE | ко | | |

How to select category and condition

Category - Mouse strain Condition - Sex

| Sample | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|-----------------|------|-----|--------|---|----|-------------|---|---|----------|------|----|-----|------|-----|--------|--------|
| Mouse strain | | Str | ain A | | | Strain B | | | Strain C | | | | | | Strain | D |
| Sex | Male | | Female | | Ma | Male Female | | | | Male | | Fen | nale | Mal | е | Female |

Number of unique draws

= \prod (Number of replicates per category-condition group) = 1*3*2*2*3*2*1 = 144

One example draw:

| Sample | 1 | 2 | 5 | 7 | 9 | 12 | 14 | 16 | |
|--------------|------|--------|------|--------|------|--------|----------|--------|--|
| Mouse strain | Str | ain A | Stra | in B | Stra | in C | Strain D | | |
| Sex | Male | Female | Male | Female | Male | Female | Male | Female | |