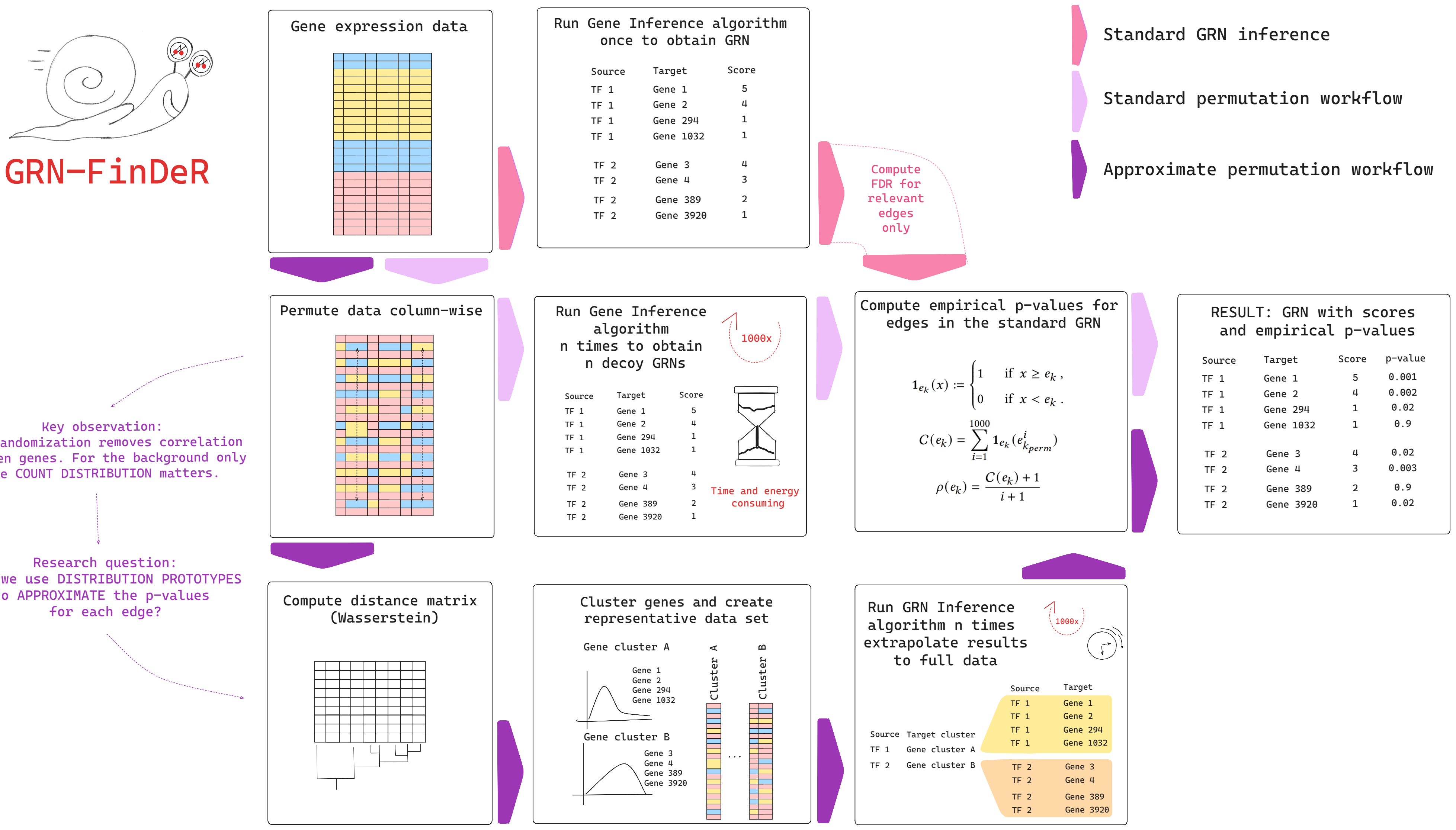
GRN-FinDeR
Standard GRN inference
Standard permutation workflow
Approximate permutation workflow
Gene expression data
Run Gene Inference algorithm once to obtain GRN
Compute FDR for relevant edges only
Permute data column-wise
Run Gene Inference algorithm n times to obtain n decoy GRNs
1000x
Time and energy consuming
Compute empirical p-values for edges in the standard GRN
RESULT: GRN with scores and empirical p-values
Key observation: The randomization removes correlation between genes. For the background only the COUNT DISTRIBUTION matters.
Research question: Can we use DISTRIBUTION PROTOTYPES to APPROXIMATE the p-values for each edge?
Compute distance matrix (Wasserstein)
Cluster genes and create representative data set
Gene cluster A
Gene 1
Gene 2
Gene 294
Gene 1032
Gene cluster B
Gene 3
Gene 4
Gene 389
Gene 3920
Cluster A
Cluster B
...
Run GRN Inference algorithm n times extrapolate results to full data
1000x

Run Gene Inference algorithm once to obtain GRN

| Source | Target | Score |
|---|---|---|
| TF 1 | Gene 1 | 5 |
| TF 1 | Gene 2 | 4 |
| TF 1 | Gene 294 | 1 |
| TF 1 | Gene 1032 | 1 |
| TF 2 | Gene 3 | 4 |
| TF 2 | Gene 4 | 3 |
| TF 2 | Gene 389 | 2 |
| TF 2 | Gene 3920 | 1 |

Run Gene Inference algorithm n times to obtain n decoy GRNs

| Source | Target | Score |
|---|---|---|
| TF 1 | Gene 1 | 5 |
| TF 1 | Gene 2 | 4 |
| TF 1 | Gene 294 | 1 |
| TF 1 | Gene 1032 | 1 |
| TF 2 | Gene 3 | 4 |
| TF 2 | Gene 4 | 3 |
| TF 2 | Gene 389 | 2 |
| TF 2 | Gene 3920 | 1 |

Compute empirical p-values for edges in the standard GRN

$$\mathbf{1}_{e_k}(x) := \begin{cases} 1 & \text{if } x \geq e_k\,, \\ 0 & \text{if } x < e_k\,. \end{cases}$$

$$C(e_k) = \sum_{i=1}^{1000} \mathbf{1}_{e_k}(e^i_{k_{perm}})$$

$$\rho(e_k) = \frac{C(e_k)+1}{i+1}$$

RESULT: GRN with scores and empirical p-values

| Source | Target | Score | p-value |
|---|---|---|---|
| TF 1 | Gene 1 | 5 | 0.001 |
| TF 1 | Gene 2 | 4 | 0.002 |
| TF 1 | Gene 294 | 1 | 0.02 |
| TF 1 | Gene 1032 | 1 | 0.9 |
| TF 2 | Gene 3 | 4 | 0.02 |
| TF 2 | Gene 4 | 3 | 0.003 |
| TF 2 | Gene 389 | 2 | 0.9 |
| TF 2 | Gene 3920 | 1 | 0.02 |

Run GRN Inference algorithm n times extrapolate results to full data

| Source | Target cluster |
|---|---|
| TF 1 | Gene cluster A |
| TF 2 | Gene cluster B |

| Source | Target |
|---|---|
| TF 1 | Gene 1 |
| TF 1 | Gene 2 |
| TF 1 | Gene 294 |
| TF 1 | Gene 1032 |
| TF 2 | Gene 3 |
| TF 2 | Gene 4 |
| TF 2 | Gene 389 |
| TF 2 | Gene 3920 |