|  |  |  |
| --- | --- | --- |
| TFs | Regions | Properties (full list below) |
| 4920 factors including:ENCODE-H1, Meissener (2 ChIP datastes), | 1547 temporal MPRA regions | Induction/activity at TP X |
| Fimo with ENCODE and hg19 (2 motif datasets) |  | Consistency/belonging with/to MPRA cluster Y that intersect with K27ac cluster Z |
| Many repetitions of TFs (not sure how to combine this since the binding site can be different) |  | Consistency/belonging with/to MPRA cluster Y that intersect with ATACseq cluster Z |
| Maybe more relevant to use only motif datasets |  | Consistency/belonging with/to MPRA cluster Y that intersect with RNAseq cluster Z |
| TFs\_list.txt  TFs\_to\_Regions\_edges.txt  TFs\_to\_Preoperties\_edges.txt | Regions\_list.txt  Regions\_to\_Preoperties\_edges.txt | Belong to a pathway (or known neuro factor)/close to a gene that belongs in a pathway (or known neuro factor) |

Perturbation MPRA design

**Edges**

TFs 🡪 regions: if a TF binds to that region

TFs 🡪 properties: if a TF satisfies this property (either hypergeometric test cutoff for enrichment or induction)

Region 🡪 properties: if a region satisfies a property

**Constraints**

Number of sequences (~2,500):

200 scrambles

For each TF and region: the unperturbed version should be present in the experiment.

\*All numbers in **purple** can be modified

**We want to choose the minimal number of regions and TFs that satisfy the following:**

For each TF at least **5** **black** edges

For each TF at least **2** **red** edges from each category

For each region at least **2** **blue** edges from each category

Objective function

**Properties list**

1. Induction/activity at TP 0h
2. Induction/activity at TP 3h
3. Induction/activity at TP 6h
4. Induction/activity at TP 12h
5. Induction/activity at TP 24h
6. Induction/activity at TP 48h
7. Induction/activity at TP 72h
8. Consistency/belonging with/to MPRA cluster 1 that intersect with K27ac cluster 1
9. Consistency/belonging with/to MPRA cluster 1 that intersect with K27ac cluster 2
10. Consistency/belonging with/to MPRA cluster 1 that intersect with K27ac cluster 3
11. Consistency/belonging with/to MPRA cluster 1 that intersect with K27ac cluster 4
12. Consistency/belonging with/to MPRA cluster 1 that intersect with K27ac cluster 5
13. Consistency/belonging with/to MPRA cluster 2 that intersect with K27ac cluster 1
14. Consistency/belonging with/to MPRA cluster 2 that intersect with K27ac cluster 2
15. Consistency/belonging with/to MPRA cluster 2 that intersect with K27ac cluster 3
16. Consistency/belonging with/to MPRA cluster 2 that intersect with K27ac cluster 4
17. Consistency/belonging with/to MPRA cluster 2 that intersect with K27ac cluster 5
18. Consistency/belonging with/to MPRA cluster 3 that intersect with K27ac cluster 1
19. Consistency/belonging with/to MPRA cluster 3 that intersect with K27ac cluster 2
20. Consistency/belonging with/to MPRA cluster 3 that intersect with K27ac cluster 3
21. Consistency/belonging with/to MPRA cluster 3 that intersect with K27ac cluster 4
22. Consistency/belonging with/to MPRA cluster 3 that intersect with K27ac cluster 5
23. Consistency/belonging with/to MPRA cluster 4 that intersect with K27ac cluster 1
24. Consistency/belonging with/to MPRA cluster 4 that intersect with K27ac cluster 2
25. Consistency/belonging with/to MPRA cluster 4 that intersect with K27ac cluster 3
26. Consistency/belonging with/to MPRA cluster 4 that intersect with K27ac cluster 4
27. Consistency/belonging with/to MPRA cluster 4 that intersect with K27ac cluster 5
28. Consistency/belonging with/to MPRA cluster 1 that intersect with ATACseq cluster 1
29. Consistency/belonging with/to MPRA cluster 1 that intersect with ATACseq cluster 2
30. Consistency/belonging with/to MPRA cluster 1 that intersect with ATACseq cluster 3
31. Consistency/belonging with/to MPRA cluster 1 that intersect with ATACseq cluster 4
32. Consistency/belonging with/to MPRA cluster 1 that intersect with ATACseq cluster 5
33. Consistency/belonging with/to MPRA cluster 2 that intersect with ATACseq cluster 1
34. Consistency/belonging with/to MPRA cluster 2 that intersect with ATACseq cluster 2
35. Consistency/belonging with/to MPRA cluster 2 that intersect with ATACseq cluster 3
36. Consistency/belonging with/to MPRA cluster 2 that intersect with ATACseq cluster 4
37. Consistency/belonging with/to MPRA cluster 2 that intersect with ATACseq cluster 5
38. Consistency/belonging with/to MPRA cluster 3 that intersect with ATACseq cluster 1
39. Consistency/belonging with/to MPRA cluster 3 that intersect with ATACseq cluster 2
40. Consistency/belonging with/to MPRA cluster 3 that intersect with ATACseq cluster 3
41. Consistency/belonging with/to MPRA cluster 3 that intersect with ATACseq cluster 4
42. Consistency/belonging with/to MPRA cluster 3 that intersect with ATACseq cluster 5
43. Consistency/belonging with/to MPRA cluster 4 that intersect with ATACseq cluster 1
44. Consistency/belonging with/to MPRA cluster 4 that intersect with ATACseq cluster 2
45. Consistency/belonging with/to MPRA cluster 4 that intersect with ATACseq cluster 3
46. Consistency/belonging with/to MPRA cluster 4 that intersect with ATACseq cluster 4
47. Consistency/belonging with/to MPRA cluster 4 that intersect with ATACseq cluster 5
48. Consistency/belonging with/to MPRA cluster 1 that intersect with RNAseq cluster 1
49. Consistency/belonging with/to MPRA cluster 1 that intersect with RNAseq cluster 2
50. Consistency/belonging with/to MPRA cluster 1 that intersect with RNAseq cluster 3
51. Consistency/belonging with/to MPRA cluster 1 that intersect with RNAseq cluster 4
52. Consistency/belonging with/to MPRA cluster 2 that intersect with RNAseq cluster 1
53. Consistency/belonging with/to MPRA cluster 2 that intersect with RNAseq cluster 2
54. Consistency/belonging with/to MPRA cluster 2 that intersect with RNAseq cluster 3
55. Consistency/belonging with/to MPRA cluster 2 that intersect with RNAseq cluster 4
56. Consistency/belonging with/to MPRA cluster 3 that intersect with RNAseq cluster 1
57. Consistency/belonging with/to MPRA cluster 3 that intersect with RNAseq cluster 2
58. Consistency/belonging with/to MPRA cluster 3 that intersect with RNAseq cluster 3
59. Consistency/belonging with/to MPRA cluster 3 that intersect with RNAseq cluster 4
60. Consistency/belonging with/to MPRA cluster 4 that intersect with RNAseq cluster 1
61. Consistency/belonging with/to MPRA cluster 4 that intersect with RNAseq cluster 2
62. Consistency/belonging with/to MPRA cluster 4 that intersect with RNAseq cluster 3
63. Consistency/belonging with/to MPRA cluster 4 that intersect with RNAseq cluster 4
64. Belong to a pathway (or known neuro factor)/close to a gene that belongs in a pathway (or known neuro factor) – **need to make a list of those factors**