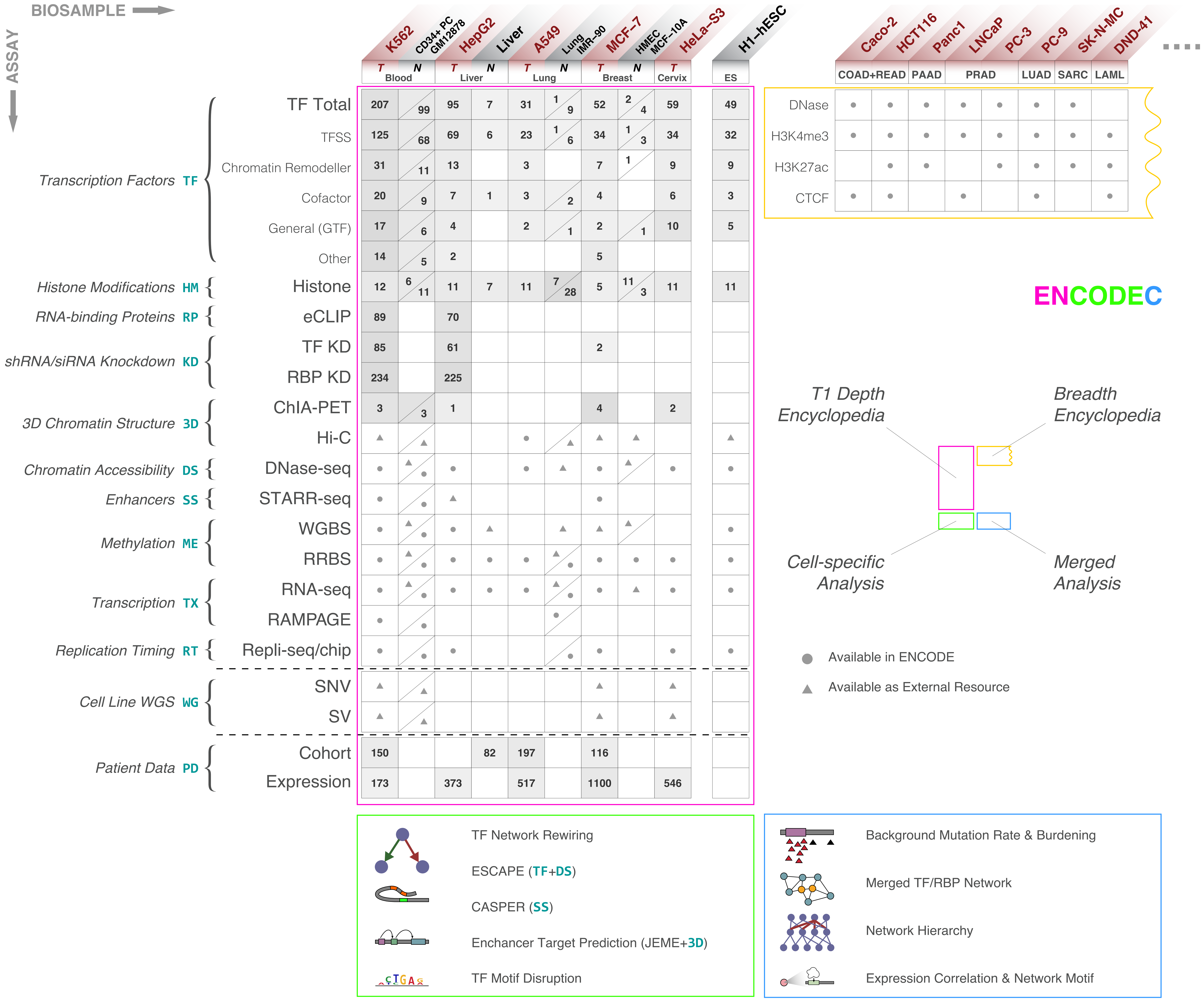
# ENCODEC

v170403
DL



| | | | K562 | CD34+ PC | GM12878 | HepG2 | Liver | A549 | Lung | IMR-90 | MCF-7 | HMEC | MCF-10A | HeLa-S3 | H1-hESC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | T | N | N | T | N | T | N | N | T | N | N | T | |
| | | | Blood | Blood | Blood | Liver | Liver | Lung | Lung | Lung | Breast | Breast | Breast | Cervix | ES |
| Transcription Factors | TF | TF Total | 207 | | 99 | 95 | 7 | 31 | 1 | 9 | 52 | 2 | 4 | 59 | 49 |
| | | TFSS | 125 | | 68 | 69 | 6 | 23 | 1 | 6 | 34 | 1 | 3 | 34 | 32 |
| | | Chromatin Remodeller | 31 | | 11 | 13 | | 3 | | | 7 | 1 | | 9 | 9 |
| | | Cofactor | 20 | | 9 | 7 | 1 | 3 | | 2 | 4 | | | 6 | 3 |
| | | General (GTF) | 17 | | 6 | 4 | | 2 | | 1 | 2 | | 1 | 10 | 5 |
| | | Other | 14 | | 5 | 2 | | | | | 5 | | | | |
| Histone Modifications | HM | Histone | 12 | 6 | 11 | 11 | 7 | 11 | 7 | 28 | 5 | 11 | 3 | 11 | 11 |
| RNA-binding Proteins | RP | eCLIP | 89 | | | 70 | | | | | | | | | |
| shRNA/siRNA Knockdown | KD | TF KD | 85 | | | 61 | | | | | 2 | | | | |
| | | RBP KD | 234 | | | 225 | | | | | | | | | |
| 3D Chromatin Structure | 3D | ChIA-PET | 3 | | 3 | 1 | | | | | 4 | | | 2 | |
| | | Hi-C | ▲ | | ▲ | | | ● | | ▲ | ▲ | ▲ | | | ▲ |
| Chromatin Accessibility | DS | DNase-seq | ● | ▲ | ● | ● | | ● | ▲ | | ● | ▲ | | ● | ● |
| Enhancers | SS | STARR-seq | ● | | ● | ▲ | | | | | ● | | | | |
| Methylation | ME | WGBS | ● | ▲ | ● | ● | ▲ | | ▲ | | ▲ | ▲ | | | ● |
| | | RRBS | ● | ▲ | ● | ● | ● | ● | ▲ | ● | ● | ● | | ● | ● |
| Transcription | TX | RNA-seq | ● | ▲ | ● | ● | ● | ● | ▲ | ● | ● | ▲ | | ● | ● |
| | | RAMPAGE | ● | | ● | | | | ● | | | | | | |
| Replication Timing | RT | Repli-seq/chip | ● | | ● | ● | | | | ● | ● | | | ● | ● |
| Cell Line WGS | WG | SNV | ▲ | | ▲ | | | | | | ▲ | | | ▲ | |
| | | SV | ▲ | | ▲ | | | | | | ▲ | | | ▲ | |
| Patient Data | PD | Cohort | 150 | | | | 82 | 197 | | | 116 | | | | |
| | | Expression | 173 | | | 373 | | 517 | | | 1100 | | | 546 | |

| | Caco-2 | HCT116 | Panc1 | LNCaP | PC-3 | PC-9 | SK-N-MC | DND-41 |
|---|---|---|---|---|---|---|---|---|
| | COAD+READ | COAD+READ | PAAD | PRAD | PRAD | LUAD | SARC | LAML |
| DNase | ● | ● | ● | ● | ● | ● | ● | |
| H3K4me3 | ● | ● | ● | ● | ● | ● | ● | ● |
| H3K27ac | | ● | ● | | ● | ● | ● | ● |
| CTCF | ● | ● | | ● | | ● | | ● |

● Available in ENCODE

▲ Available as External Resource

TF Network Rewiring

ESCAPE (TF+DS)

CASPER (SS)

Enchancer Target Prediction (JEME+3D)

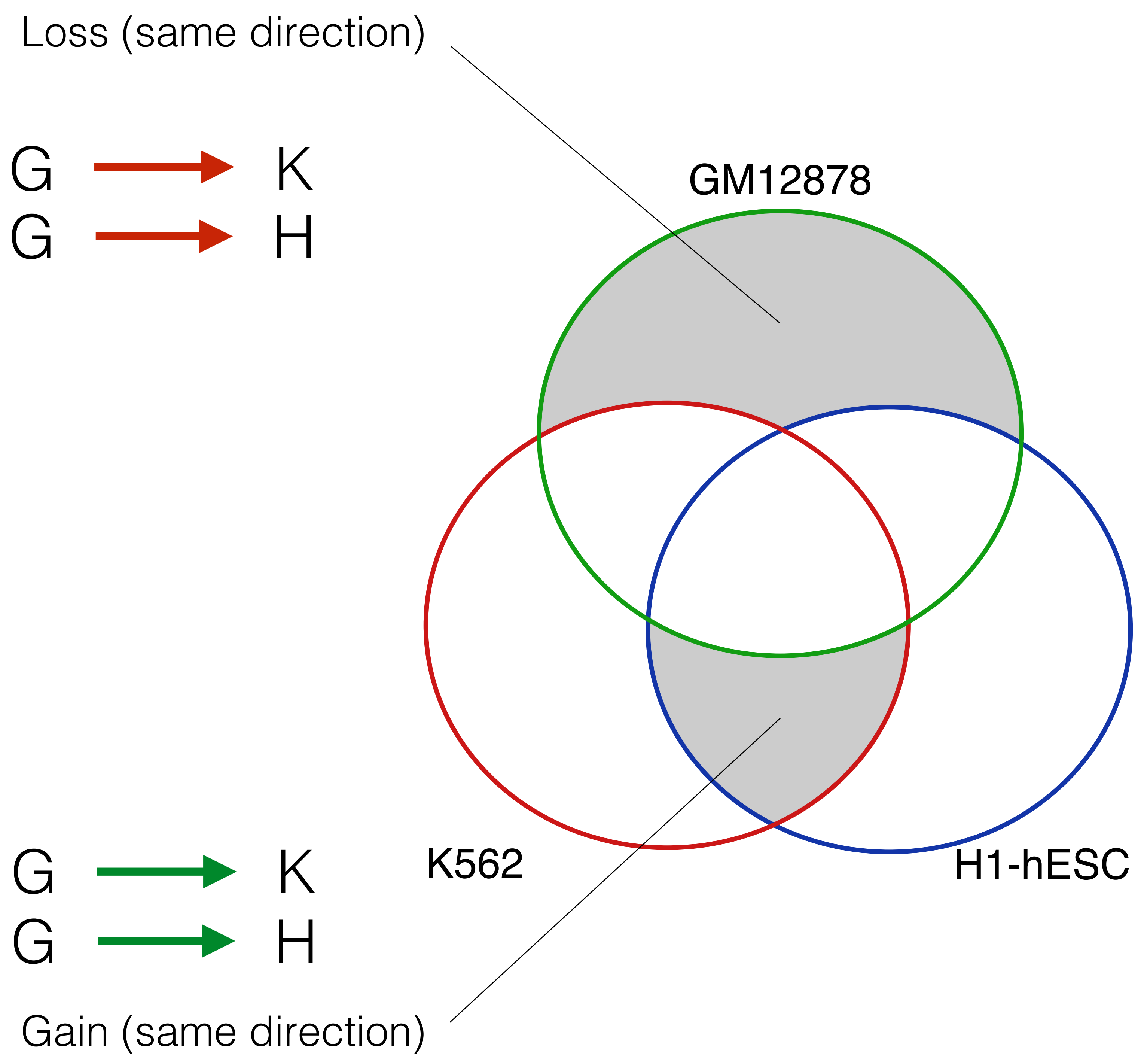
TF Motif Disruption

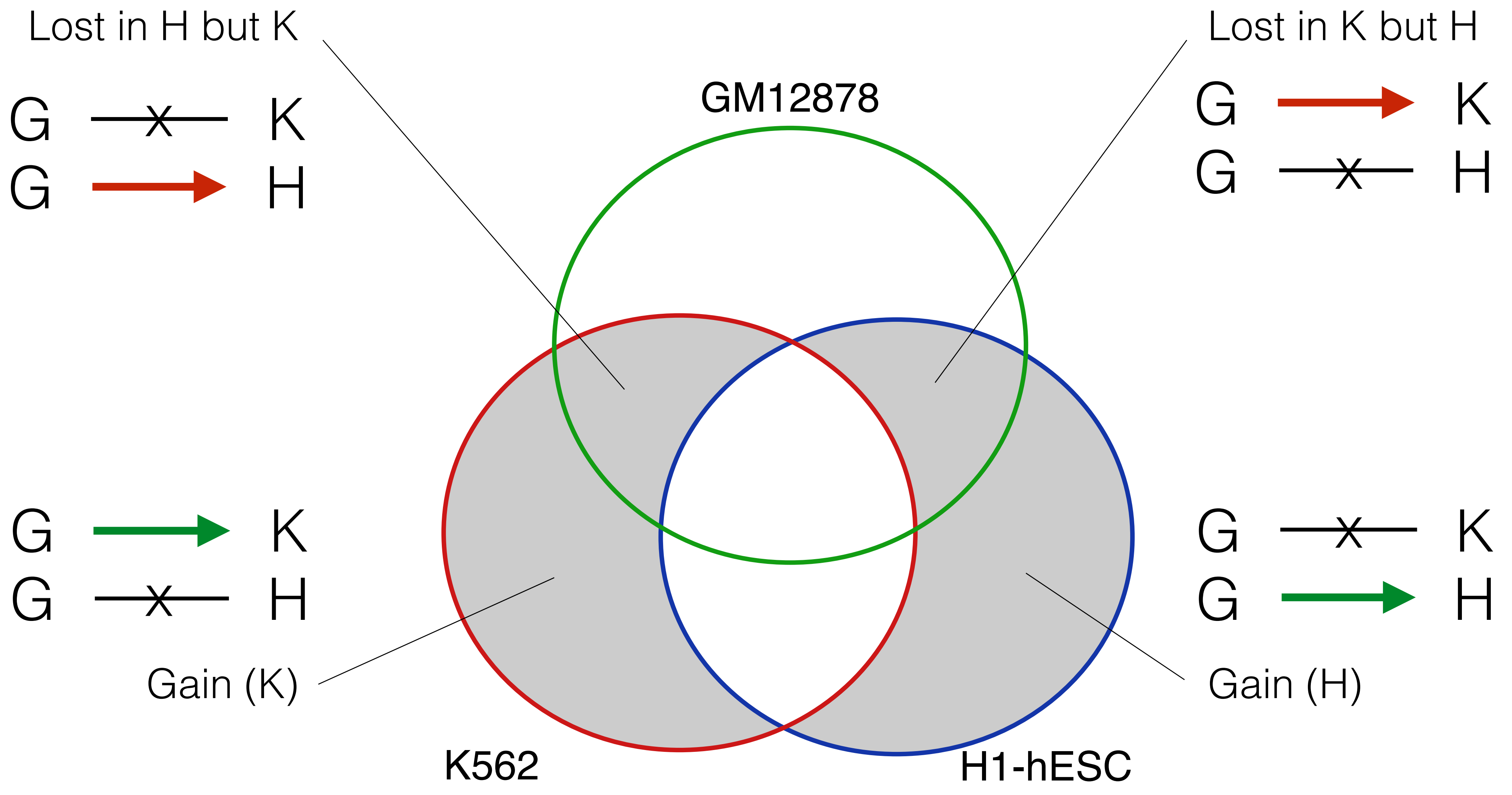
Background Mutation Rate & Burdening

Merged TF/RBP Network

Network Hierarchy

Expression Correlation & Network Motif

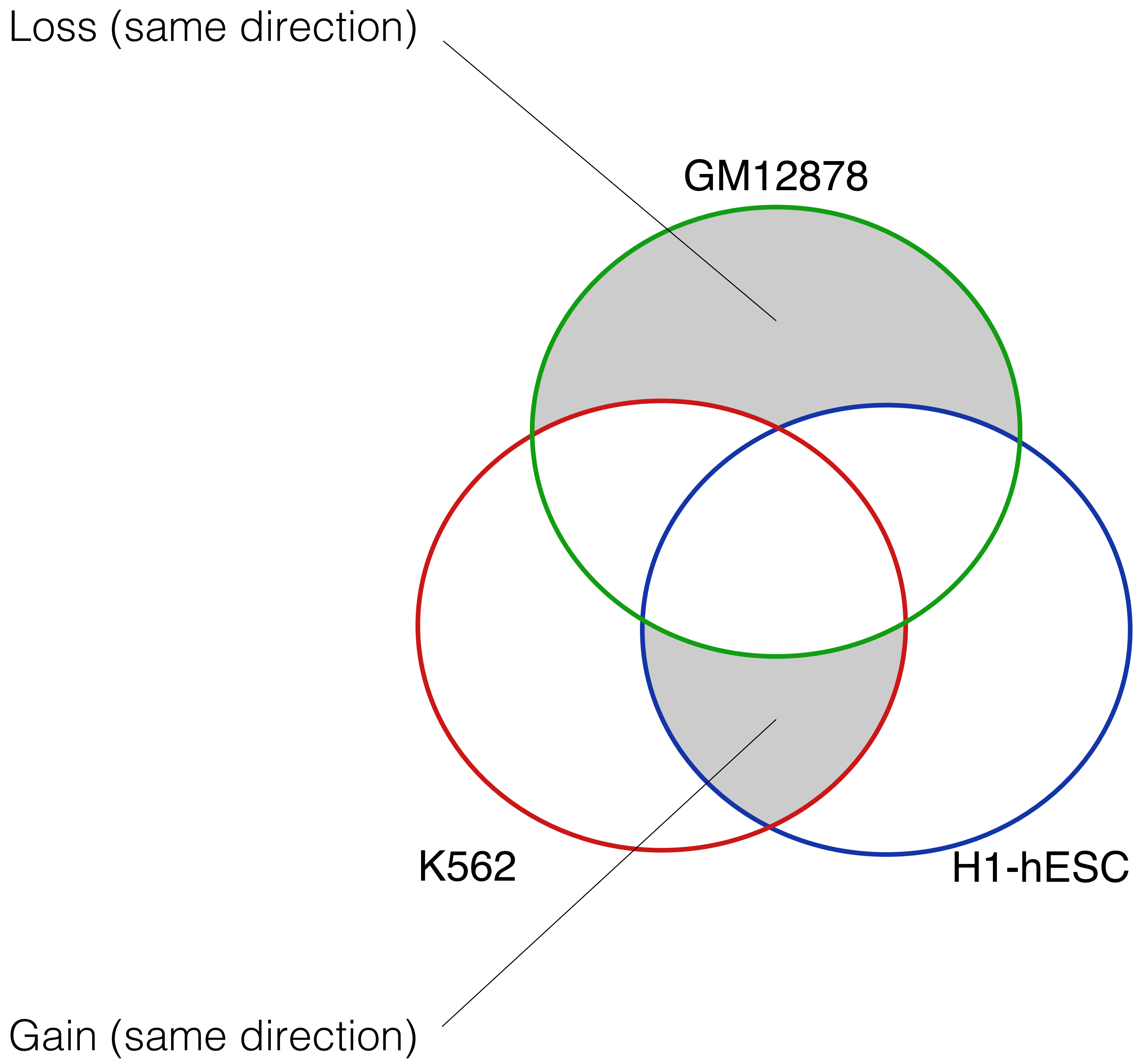


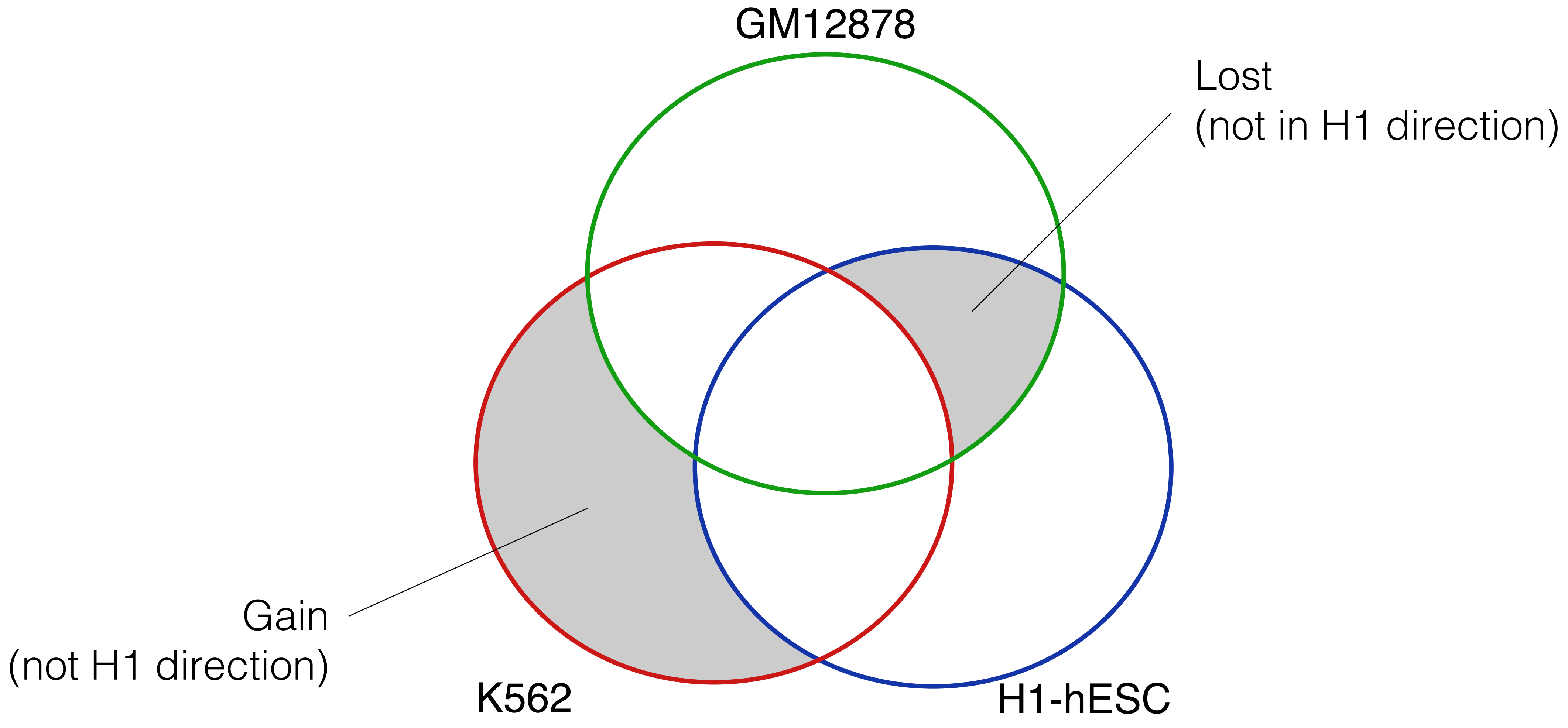


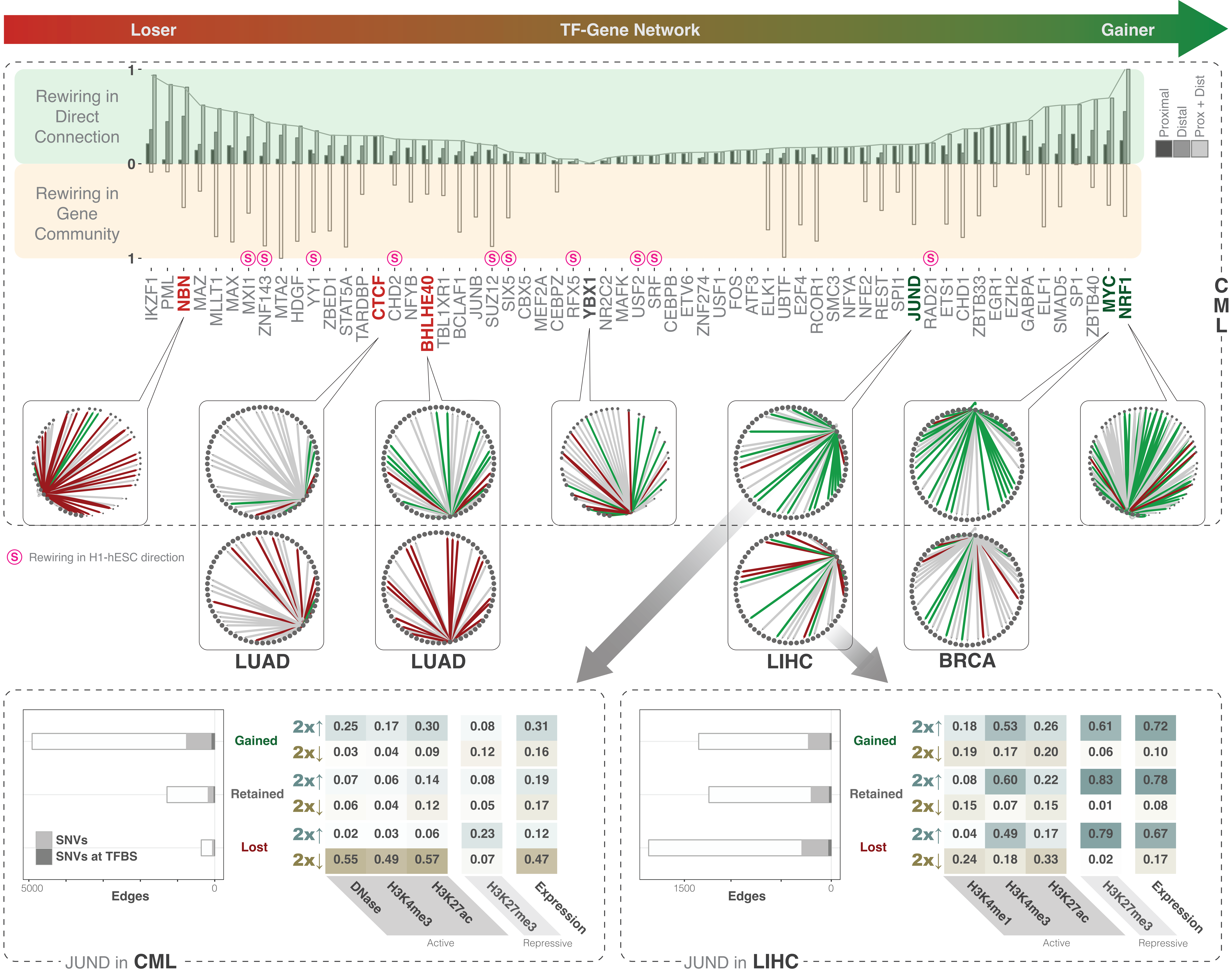
- Edges in the same direction vs. edges in the opposite direction
- TIP+enhancers for “rScore”, TSS only for “H1 similarity”
- More edges in G+K
- Gain dominant

| | **n.edges** |
|---|---|
| **GM12878** | 302,295 |
| **K562** | 572,944 |
| **H1-hESC** | 131,330 |

| **TF** | **RFX5** | **SRF** | **CHD2** | **ZNF143** | **MXI1** | **USF2** | **SUZ12** | **SIX5** | **YY1** | **RAD21** | **CTCF** | **USF1** | **CEBPB** | **SP1** | **ATF3** | **EZH2** | **MAFK** | **CHD1** | **GABPA** | **JUND** | **EGR1** | **MAX** | **MYC** | **REST** | **NRF1** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| same.dir | 1353 | 1541 | 2015 | 2491 | 2478 | 1157 | 1325 | 716 | 1914 | 1193 | 1221 | 1053 | 1167 | 1249 | 883 | 337 | 663 | 982 | 930 | 1248 | 985 | 1055 | 917 | 832 | 313 |
| same.dir x2 | 2706 | 3082 | 4030 | 4982 | 4956 | 2314 | 2650 | 1432 | 3828 | 2386 | 2442 | 2106 | 2334 | 2498 | 1766 | 674 | 1326 | 1964 | 1860 | 2496 | 1970 | 2110 | 1834 | 1664 | 626 |
| opposite.dir | 705 | 1073 | 2081 | 2578 | 2583 | 1264 | 1539 | 963 | 3017 | 2281 | 2640 | 2411 | 3085 | 3369 | 2569 | 1027 | 2178 | 3285 | 3598 | 4891 | 5143 | 5617 | 4916 | 5836 | 5022 |
| overall.dir | 2001 | 2009 | 1949 | 2404 | 2373 | 1050 | 1111 | 469 | 811 | 105 | -198 | -305 | -751 | -871 | -803 | -353 | -852 | -1321 | -1738 | -2395 | -3173 | -3507 | -3082 | -4172 | -4396 |
| % same dir | 0.79 | 0.74 | 0.66 | 0.66 | 0.66 | 0.65 | 0.63 | 0.60 | 0.56 | 0.51 | 0.48 | 0.47 | 0.43 | 0.43 | 0.41 | 0.40 | 0.38 | 0.37 | 0.34 | 0.34 | 0.28 | 0.27 | 0.27 | 0.22 | 0.11 |


Loss (same direction)
GM12878
K562
H1-hESC
Gain (same direction)


GM12878
Lost
(not in H1 direction)
Gain
(not H1 direction)
K562
H1-hESC



JUND in **CML**

| | | DNase | H3K4me3 | H3K27ac | H3K27me3 | Expression |
|---|---|---|---|---|---|---|
| Gained | 2x↑ | 0.25 | 0.17 | 0.30 | 0.08 | 0.31 |
| | 2x↓ | 0.03 | 0.04 | 0.09 | 0.12 | 0.16 |
| Retained | 2x↑ | 0.07 | 0.06 | 0.14 | 0.08 | 0.19 |
| | 2x↓ | 0.06 | 0.04 | 0.12 | 0.05 | 0.17 |
| Lost | 2x↑ | 0.02 | 0.03 | 0.06 | 0.23 | 0.12 |
| | 2x↓ | 0.55 | 0.49 | 0.57 | 0.07 | 0.47 |

JUND in **LIHC**

| | | H3K4me1 | H3K4me3 | H3K27ac | H3K27me3 | Expression |
|---|---|---|---|---|---|---|
| Gained | 2x↑ | 0.18 | 0.53 | 0.26 | 0.61 | 0.72 |
| | 2x↓ | 0.19 | 0.17 | 0.20 | 0.06 | 0.10 |
| Retained | 2x↑ | 0.08 | 0.60 | 0.22 | 0.83 | 0.78 |
| | 2x↓ | 0.15 | 0.07 | 0.15 | 0.01 | 0.08 |
| Lost | 2x↑ | 0.04 | 0.49 | 0.17 | 0.79 | 0.67 |
| | 2x↓ | 0.24 | 0.18 | 0.33 | 0.02 | 0.17 |

# 1 paragraph about merging with the main encyclopedia (cRE)

- Both promoter and enhancer annotations from ENCODEC were carefully consolidated with the main ENCODE Encyclopedia.

- The ENCODE Encyclopedia comprises two levels of annotations. The ground level includes peaks and quantifications produced by uniform processing pipelines for individual data types. The integrative level contains annotations produced by integrating multiple data types. The core of the integrative level is the Registry of candidate Regulatory Elements (cREs). The registry contains approximately ~1.31M human cREs and each cRE has a cell-type non-specific accession number, which then can be browsed from SCREEN (Search Candidate Regulatory Elements by ENCODE, http://screen.umassmed.edu/).

- Annotations from ENCODEX were merged against the Registry of candidate Regulatory Elements (cREs). We assigned cRE accession number to promoters and enhancers when the region had more than 50% overlap. When there are more than one accession numbers matched to the regulatory element, we assigned the accession with the highest overlap.

- 167221 GM12878_EnhancerSeq_rename_sorted_uniq.bed
- x
- 21603 GM12878_SVMfiltered_rename_sorted_uniq_merged.bed
- =
- 6293 GM12878_MFxES.bed

50% from both
2611 41.5%

50% of ours
3043 48.4%

1,310,152 hg19-cREs.bed

- 269855 K562_EnhancerSeq_rename_sorted_uniq.bed
- x
- 22001 K562_SVMfiltered_rename_sorted_uniq_merged.bed
- =
- 12869 K562_MFxES.bed

50 % from both
5483 42.6%

50% of ours
6157 47.8%