Supplementary Information

Data-driven structural analysis of Small Cell Lung Cancer transcription factor network suggests potential subtype regulators and transition pathways

Mustafa Ozen[1,2] and Carlos F. Lopez[1,2*]

[1]Dept. of Biochemistry, Vanderbilt University, Nashville, TN 37212, USA

[2]Present address: Multiscale Modeling Group, SI3, Altos Labs, Redwood City, CA 94065, USA

* To whom correspondence should be addressed: Email: clopez@altoslabs.com

Supplementary Notes

1. DST versus MDST: Comparing interactions remaining on the found DSTs and MDSTs

The dense spanning trees (DSTs) of the SCLC TF network are the substructures that emphasize some TFs as the hubs while preserving minimum total distances between the TFs and hence the maximum influence on each other. On the other hand, MDSTs of the weighted SCLC TF network are the substructures that still emphasize some TFs as the hubs and preserve the maximum influence between the TFs while minimizing the total weights assigned to the edges, that is for each edge e_i, the weight $w_i = 1 - P(e_i \text{ exists})$. Once we solved the associated optimization problems (Equations (1) and (2) in the main text), we observed 146,143 DSTs and 46 MDSTs all having the same objective values for their associated objective functions. Looking at the average node degrees among all the found DSTs and MDSTs, we have seen that most of the found hubs overlap between both analyses.

Here, we compare the interactions remaining in the DSTs and MDSTs. To do so, we computed the probability of an interaction remaining in the found DSTs (Supplementary Figure 3A) and MDSTs (Supplementary Figure 3B). As seen in the figure, some edges always remain in the found DSTs and MDSTs. For instance, the interaction between FLI1 and MITF always remains in the found DSTs. Similarly, the interaction between MITF and EBF1 always remains in the found MDSTs. Upon comparing all the interactions that always remain in the found DSTs and MDSTs, i.e., $P(e_i \text{ remaining in DST and MDST}) = 1$, we have seen that the interactions ASCL1–FLI1, GATA4–FLI1, ISL1–FLI1, MYCN–FLI1, NEUROD1–FLI1, NEUROD2–FLI1, RARG–FLI1, RCOR2–FLI1, SOX11–FLI1, STAT6–FLI1, and TCF3–FLI1 are common. This means that to observe the minimum total distance and maximum influence between the TFs network, these interactions should be kept in the DSTs and MDSTs, which shows their structural importance.

Additionally, the interactions having a high probability of remaining in the found DSTs and MDSTs might help to identify the possible important pathways between the hubs. For example, the interaction between the ASCL1–FLI1 always remains in both DSTs and MDSTs. Also, the MITF–ASCL1 connection has a probability of 1 for DSTs and 0.8 for MDSTs, meaning that it is very highly likely to have this connection in both substructures. This means that it is highly probable that the pathway FLI1–ASCL1–MITF also exists in the found DSTs and MDSTs, in which FLI1 (regulator of NE subtype) and MITF (regulator of NON-NE subtype) are two major hubs. Therefore, one can target this pathway both in silico and in experiments to test their potential impact on SCLC subtypes and NON-NE to NE subtype transitions as done in the main text.

2. Comparing various hub definitions and their results on the SCLC TF network

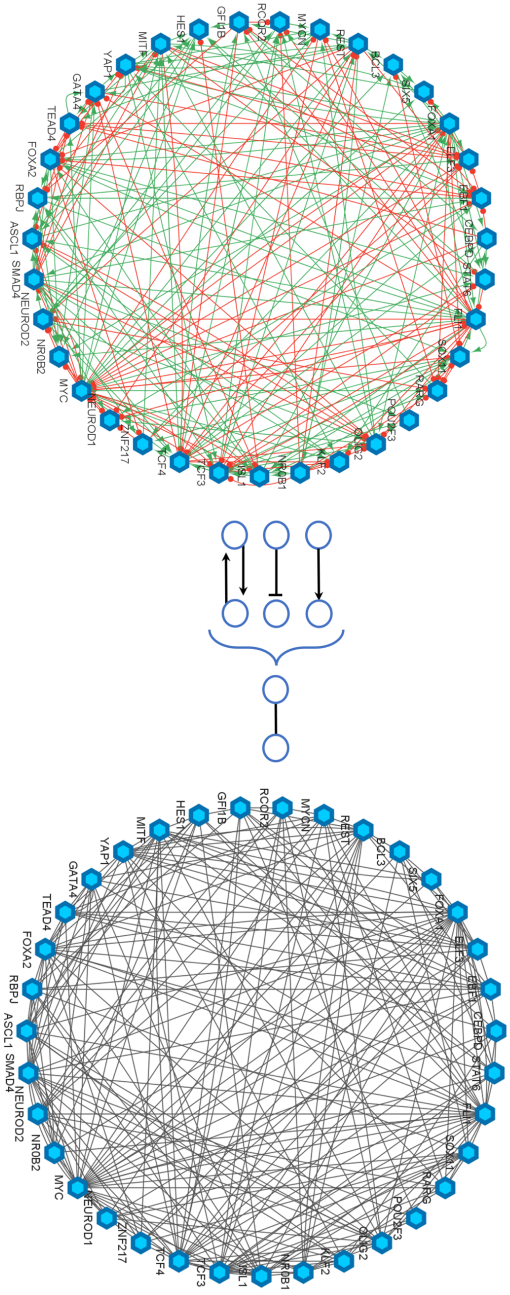
There are different ways to define and identify the hubs for a given network. However, given that this is a network structure-based analysis, different definitions of the hubs as well as the ways of their identification using methods focusing on various structural properties of the network may yield different results. For example, for the SCLC TF network, if one defines a hub as the node that has the most connection in the network and ranks the nodes based on their degrees, the top five TFs would be MYC (degree = 31), FLI1 (degree = 27), FOXA1 (degree = 25), TFC4 (degree = 22), TFC3 (degree = 21) regardless of the edge weights. Similarly, when the nodes are ranked based on their clustering coefficients (cc), the top five nodes will be SIX5 (cc = 0.7), RARG (cc = 0.5833), RCOR2 (cc = 0.5714), ASCL1 (cc = 0.5606), CEBPD (cc = 0.5556). We tried other metrics and summarized the results in Supplementary Table 1. As seen in the table, different methods yield different rankings because they rank the nodes based on different structural properties of the network. Nonetheless, we believe they are not very well suited for biological applications as they are purely structural concepts and don't concern about the closeness, i.e., the influence of the nodes with each other. We focused on the dense spanning trees because these substructures not only focus on the high individual node connectivity but also concerns how close all the nodes in the observed subnetworks are to preserve the maximum influence of the nodes with each other, which is biologically more relevant as discussed in the main text.

Supplementary Tables

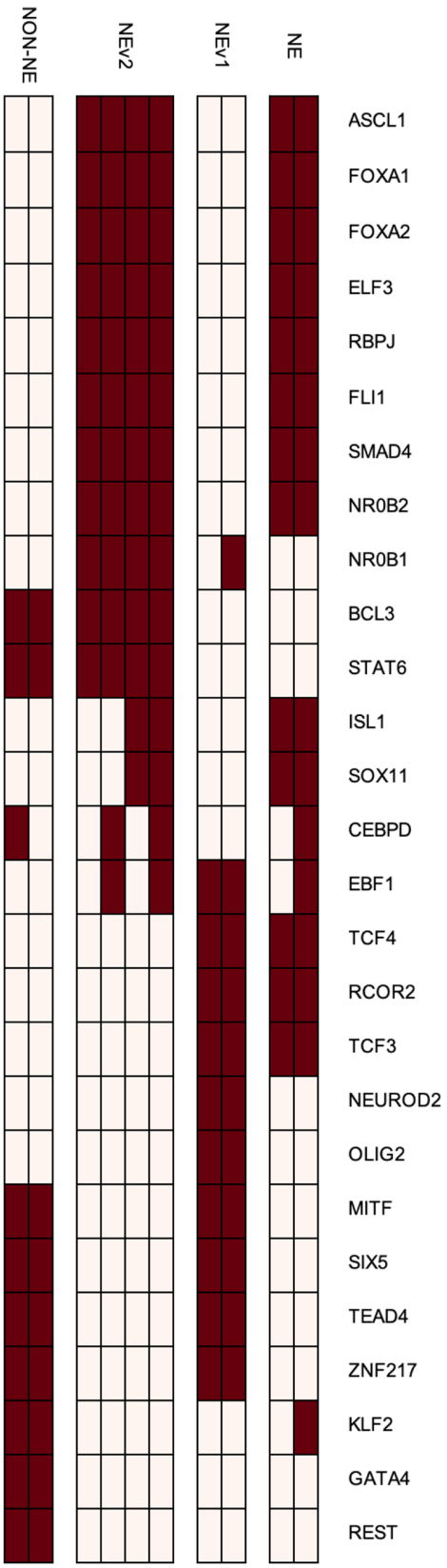
Supplementary Table 1: Top 5 transcription factors identified as hubs using different structural analysis methods.

Method	**Top 5 TFs**
Node degree (nd)	MYC (nd = 31), FLI1 (nd = 27), FOXA1 (nd = 25), TFC4 (nd = 22), TFC3 (nd = 21)
Clustering coefficients (cc)	SIX5 (cc = 0.7), RARG (cc = 0.5833), RCOR2 (cc = 0.5714), ASCL1 (cc = 0.5606), CEBPD (cc = 0.5556)
Neighborhood connectivity (nc)	KLF2 (nc = 16.11), ISL1 (nc = 15.92), STAT6 (nc = 15.8), SIX5 (nc = 15.8), GATA4 (nc = 15.64)
Betweenness centrality (bc)	MYC (bc = 0.1153), FLI1 (bc = 0.06), SMAD4 (bc = 0.053), FOXA1 (bc = 0.052), TFC4 (bc = 0.049)
Topological coefficients (tc)	SIX5 (tc = 0.51), KLF2 (tc = 0.47), ZNF217 (tc = 0.47), ISL1 (tc = 0.46), STAT6 (tc = 0.46)

Supplementary Figures

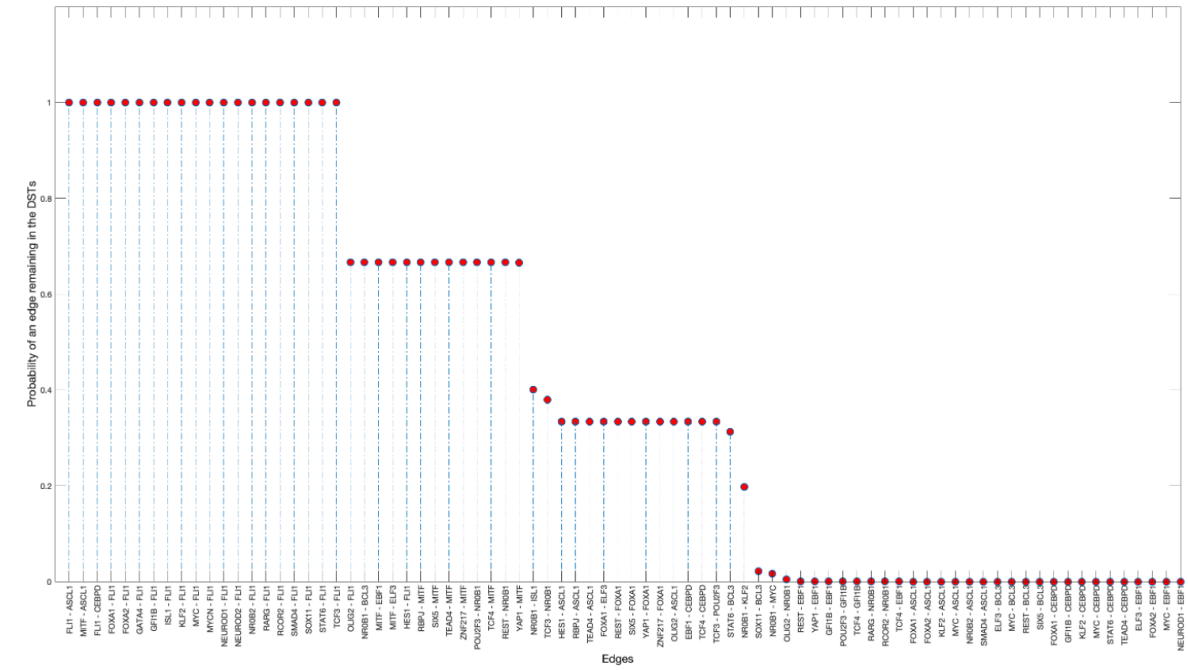


Supplementary Figure 1. Converting directed SCLC TF network into undirected network to observe relatively unbiased network structure. Here, we only care whether there is an interaction between the two TFs and ignore the type of interaction, i.e., activation or inhibition.

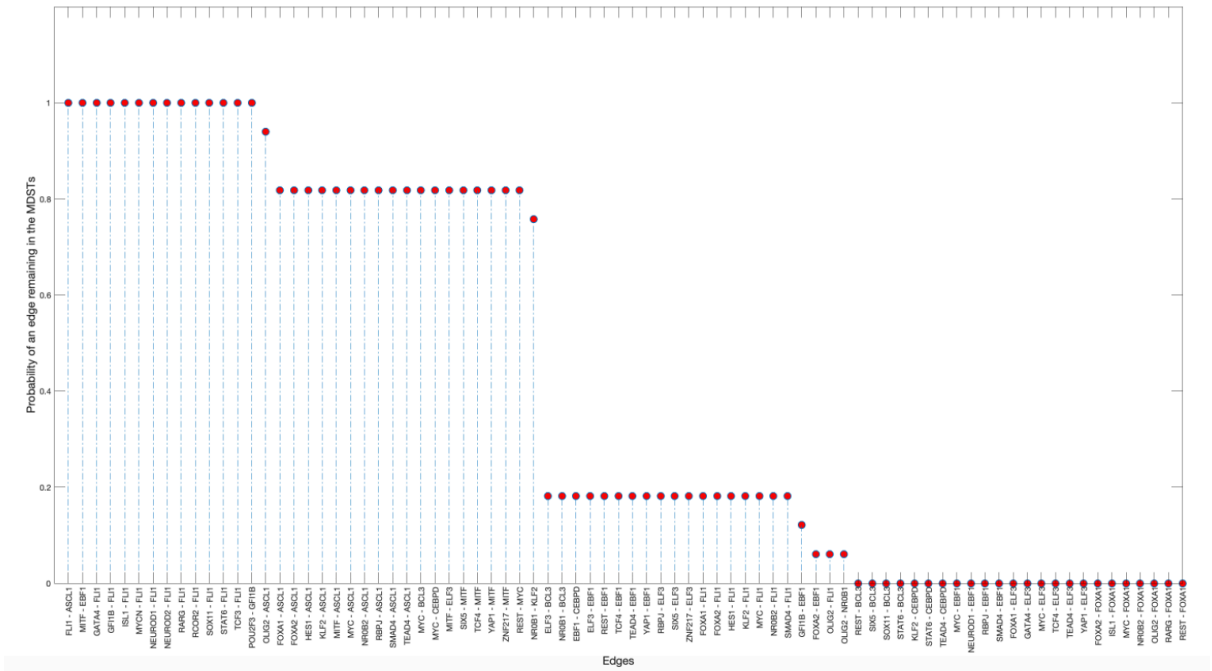


Supplementary Figure 2. Boolean states of each TFs in different SCLC subtypes as identified by Wooten et al. [34].

A



B



Supplementary Figure 3. Probabilities of interactions remaining in the found DSTs (A) and MDSTs (B).