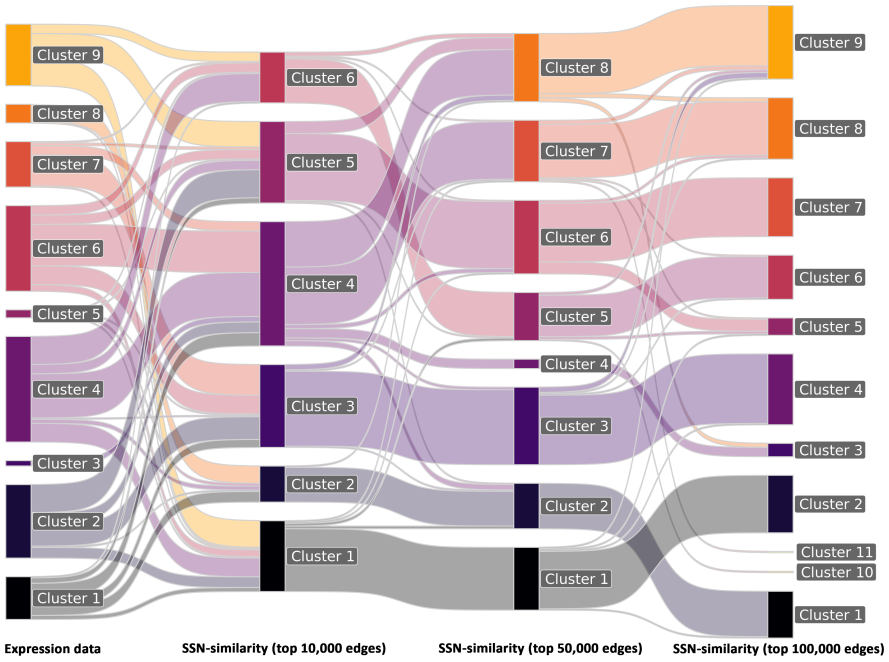
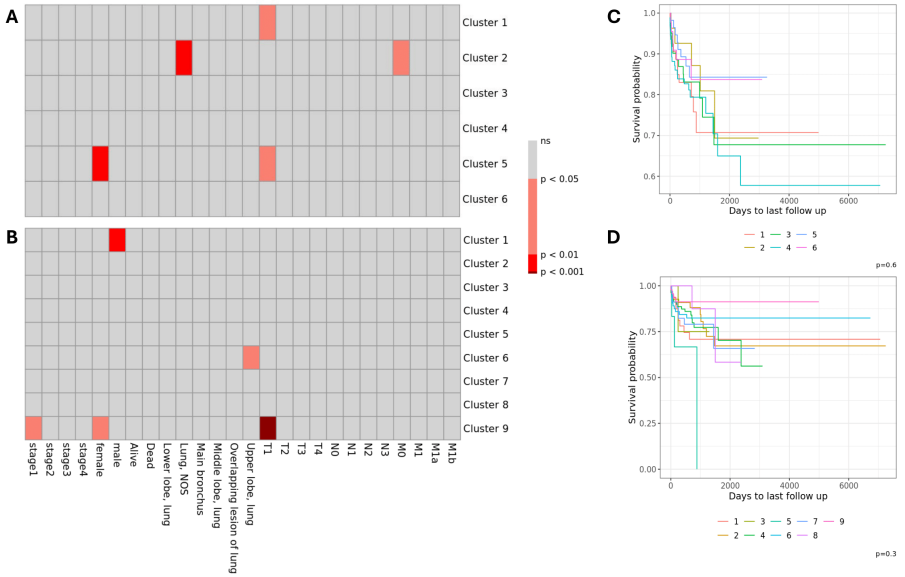
Supplementary Information

Supplementary Table 1: Structural summary statistics of per-cluster consensus networks

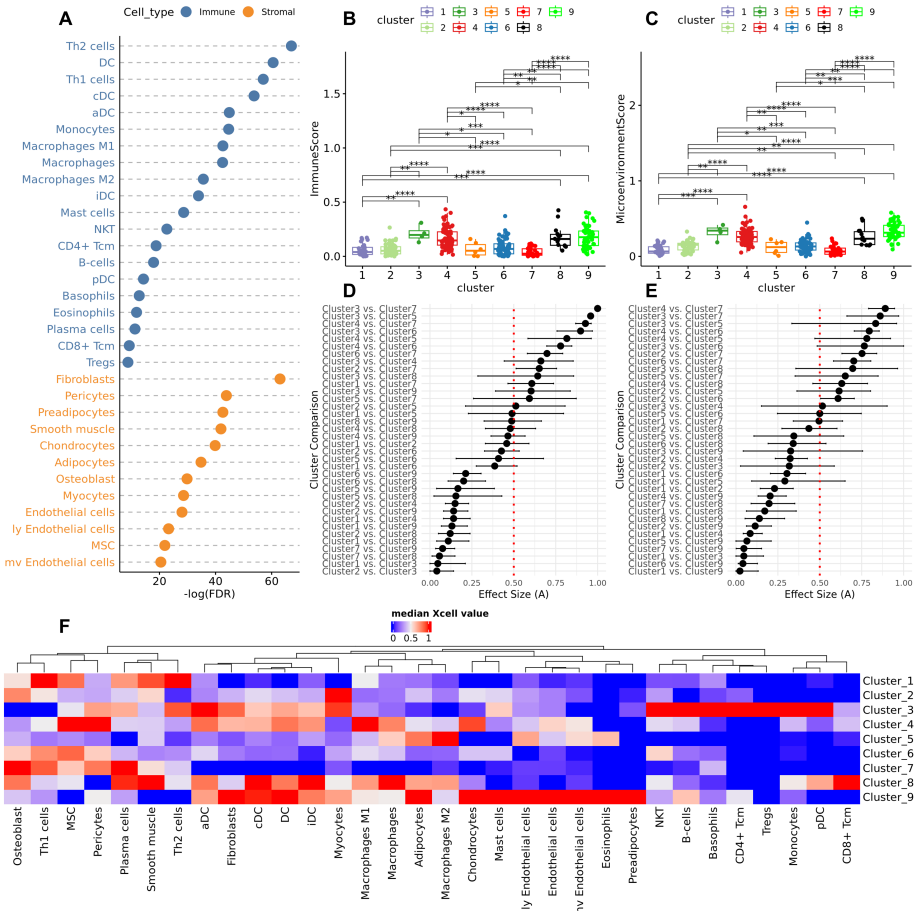
Global Descriptor	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Number of nodes	400	491	774	2257	528	1114
Number of edges	1394	1381	2360	4767	2599	2108
Avg. Number of neighbors	6.9	5.8	6.2	6.2	10.1	3.8
Network diameter	7	9	9	15	6	12
Characteristic path length	3.4	3.6	3.5	4.6	3	4.5
Clustering coefficient	0.056	0.084	0.087	0.089	0.102	0.039
Network density	0.017	0.013	0.008	0.005	0.02	0.004
Network centralization	0.106	0.153	0.101	0.081	0.145	0.048
Number of connected components	1	11	9	347	7	21



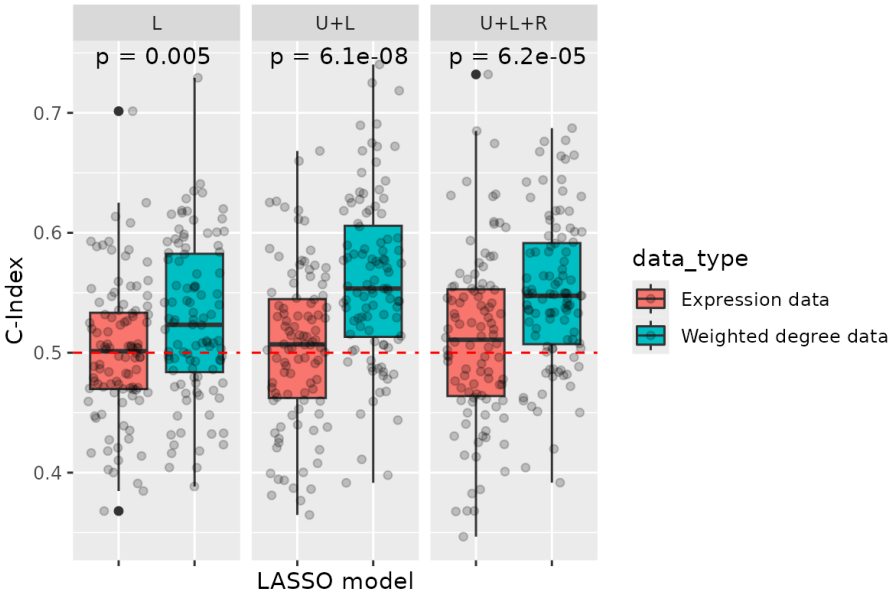
Supplementary Figure 1: Sankey diagram illustrating the agreement between clustering results based on gene expression data and patient-specific network similarity at varying network sizes. Gene expression clustering was performed using k-means, while Louvain community detection algorithm for weighted networks was used for SSN-similarity graphs.



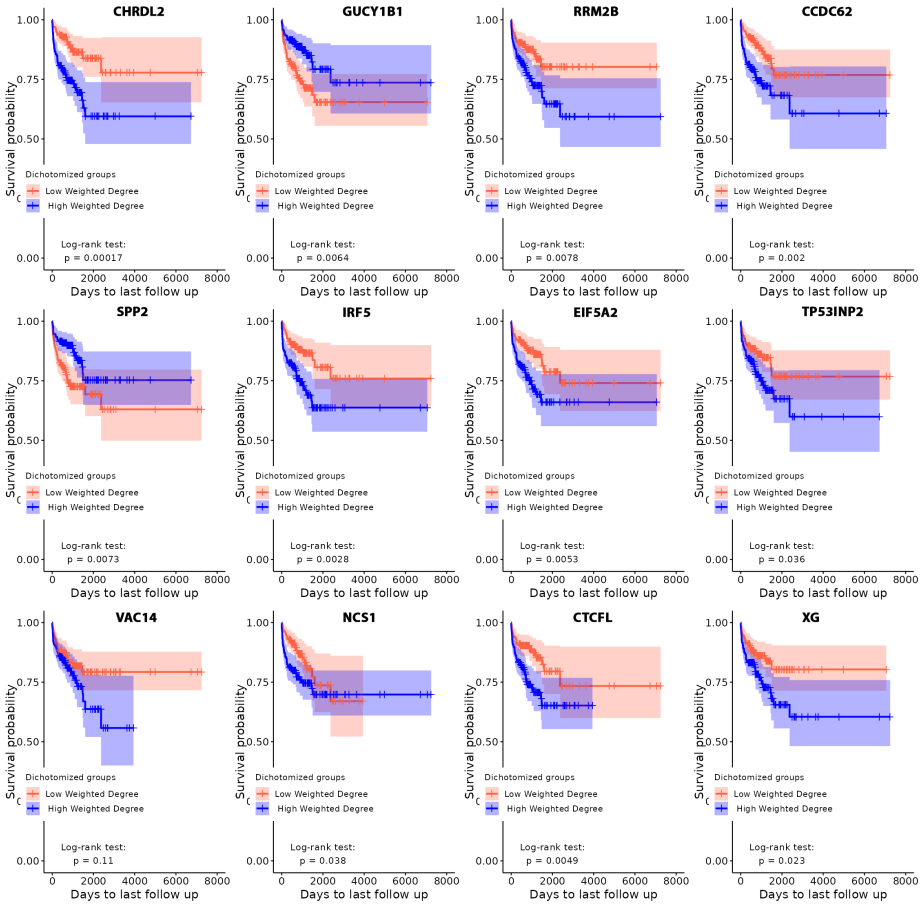
Supplementary Figure 2: Association of clusters with clinical features. **A** Significant enrichment of categorical clinical features in communities obtained from Louvain clustering of SSN-similarity graphs. **B** Significant enrichment of categorical clinical features in k-means clusters based on gene expression. **C** Kaplan-Meier plot of the survival probability distributions of SSN-similarity communities. **D** Kaplan-Meier plot of the survival probability distributions of gene-expression based clusters. Values on **A** and **B** represent the FDR corrected p-values of Fisher's exact test, while log-rank test was done for comparing survival distributions in **C** and **D**.



Supplementary Figure 3: Differences in cell enrichment scores across clusters derived from gene expression data. **A:** Dot-chart of Kruskal-Wallis test results comparing xCell's immune and stromal cell type enrichment scores between gene expression-derived clusters. The significance threshold was set at adjusted p-value < 0.001. **B-C:** Distribution of estimated ImmuneScore and StromalScore values, respectively. Pairwise significant differences were calculated through post-hoc Dunn testing. **D-E:** Forest plot depicting the effect sizes of all pairwise inter-cluster comparisons of ImmuneScore and StromalScore, respectively. Effect sizes were calculated using Vargha and Delaney's A statistic. Dotted red lines indicate the value where there are no differences between groups (i.e. points farther away from the line indicate more extreme differences). Error bars represent the 95% CI obtained by 1999 bootstrap iterations. **F:** Heatmap of median Xcell values for all cell types with significant differences obtained in the Kruskal-Wallis test.



Supplementary Figure 4: Overall predictive performance comparison between models built with gene expression data and those constructed with weighted gene degrees in 100 random training/testing data partitions. 3 modelling frameworks were contrasted: Model L, where lasso is applied to the entire set of genes, Model U+L, where a univariate filter is applied a priori to lasso, and model U+L+R, where the fit obtained on the U+L model is relaxed (known as relaxed lasso). Red dotted lines mark the score in performance where model prediction are indistinct from a random guesser. Performance differences were calculated through Wilcoxon signed rank tests.



Supplementary Figure 5: Kaplan-Meier plots of the 12 stable predictors highlighted in our analysis stratified by their median weighted degrees in the entire cohort.