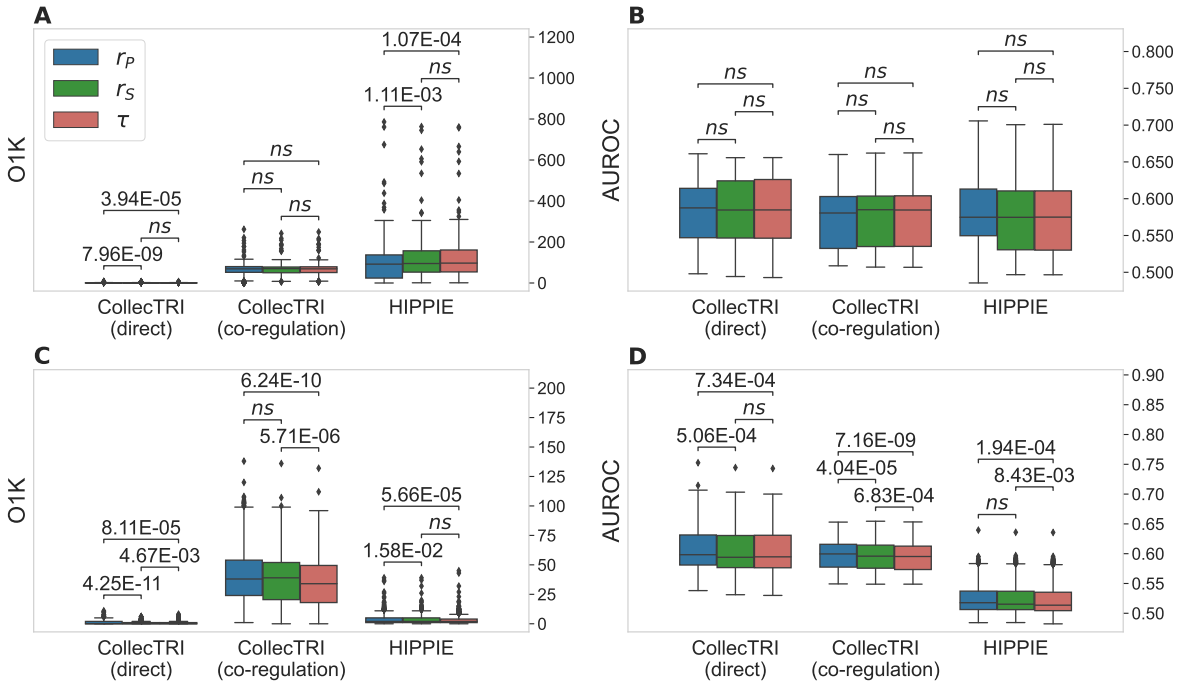
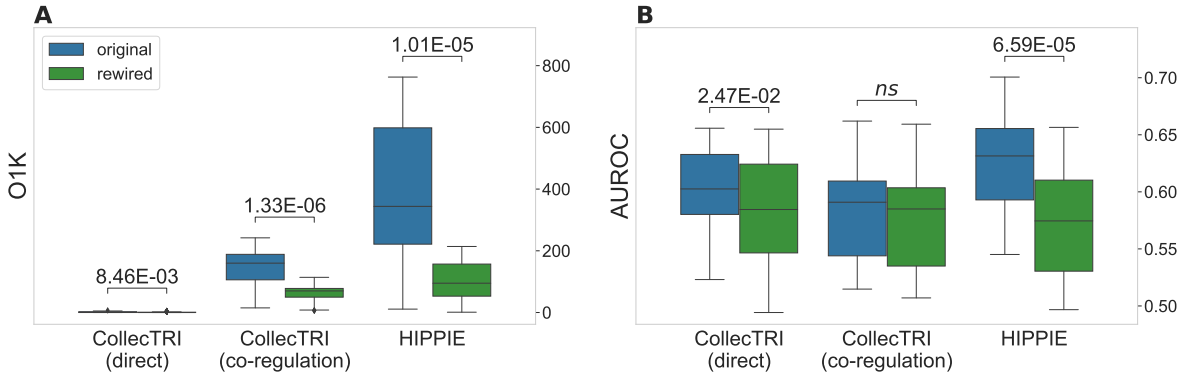
Gene co-expression poorly reflects expert-curated transcription factor-target gene regulation

Suryadipto Sarkar, David B. Blumenthal

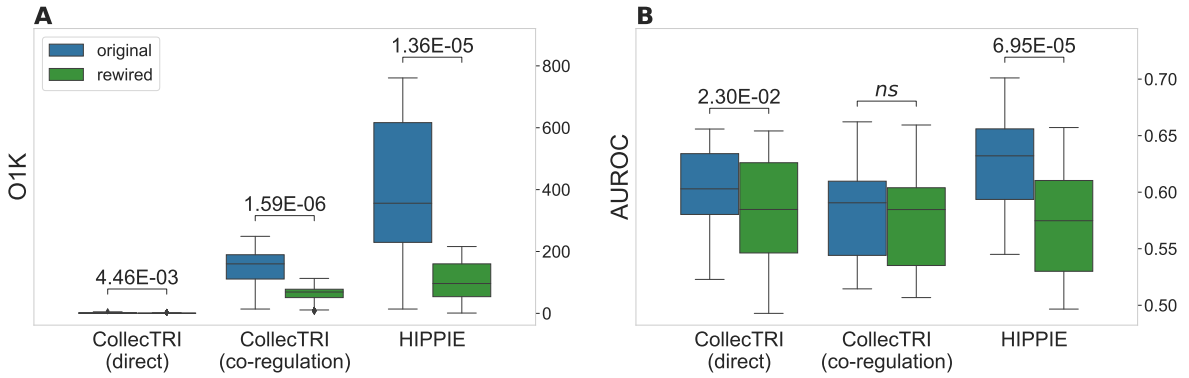
Supplementary information



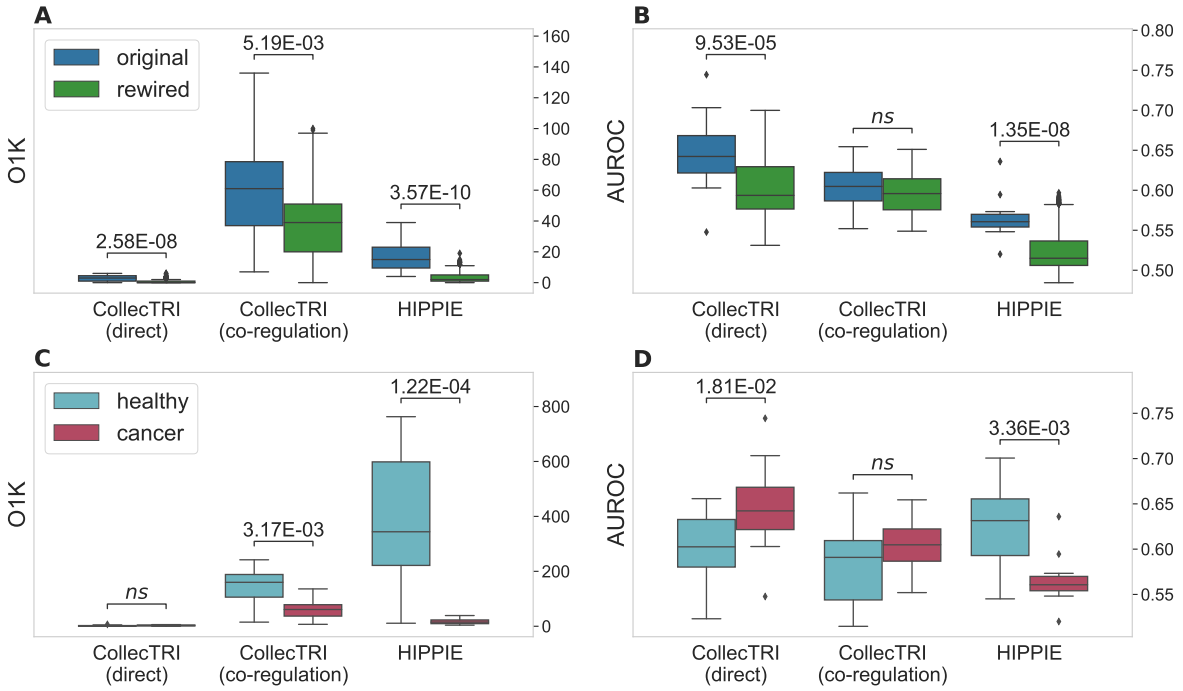
Supplementary Figure 1. Concordance between expert-curated networks and GCNs based on, respectively, Pearson correlation, Spearman's rank correlation, and Kendall's rank correlation. (A, B) O1K and AUROC scores obtained for comparison between expert-curated networks and GCNs inferred from healthy tissue gene expression data. (C, D) O1K and AUROC scores obtained for comparison between expert-curated networks and GCNs inferred from cancer gene expression data. Statistical significance was determined via a two-sided Wilcoxon signed-rank test with significance threshold 0.05.



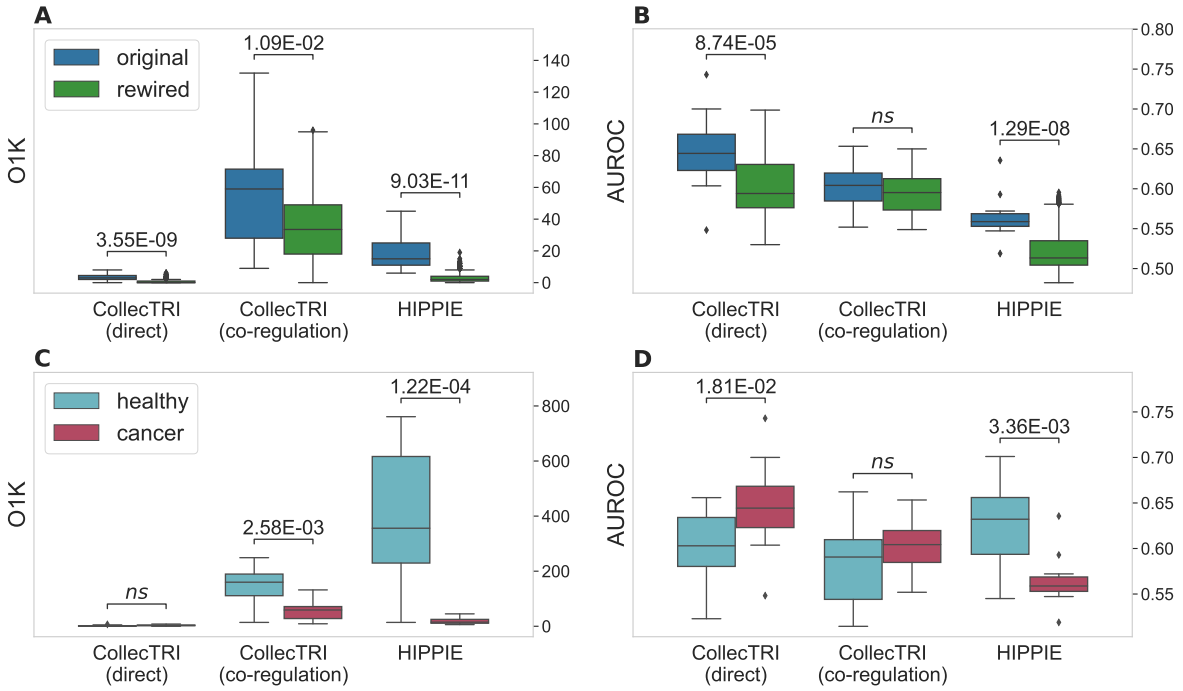
Supplementary Figure 2. Distributions of O1K scores (A) and AUROC scores (B) obtained when comparing r_S-based GCNs inferred from 16 healthy tissue gene expression datasets to expert-curated networks and their randomized counterparts. Statistical significance was determined via a one-sided Wilcoxon rank-sum test (alternative hypothesis: score distributions are larger for original than for rewired expert-curated networks) with significance threshold 0.05.



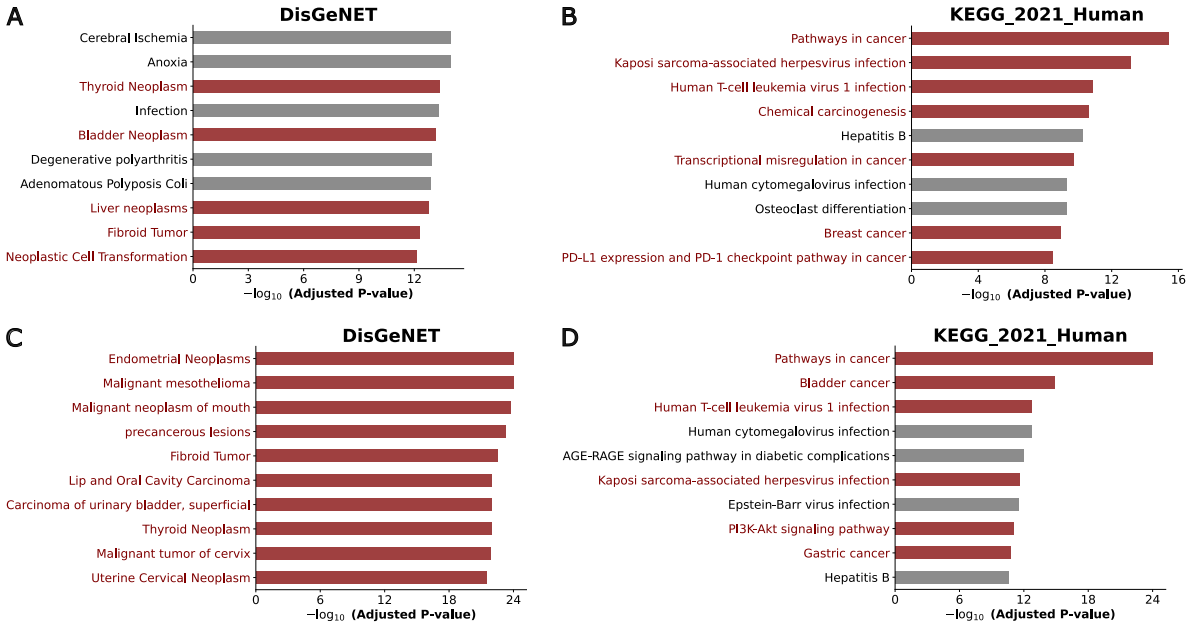
Supplementary Figure 3. Distributions of O1K scores (A) and AUROC scores (B) obtained when comparing τ-based GCNs inferred from 16 healthy tissue gene expression datasets to expert-curated networks and their randomized counterparts. Statistical significance was determined via a one-sided Wilcoxon signed-rank test (alternative hypothesis: score distributions are larger for original than for rewired expert-curated networks) with significance threshold 0.05.



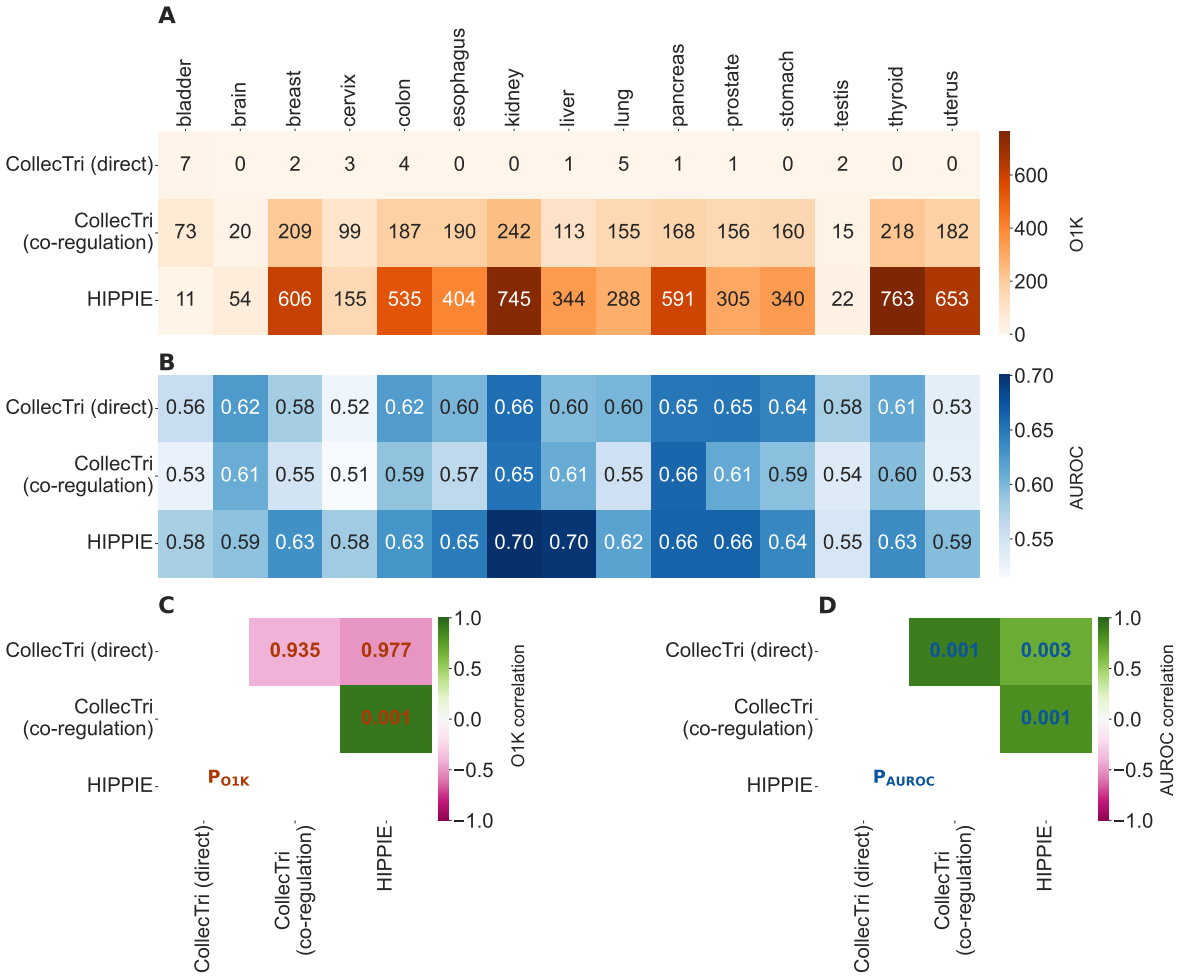
Supplementary Figure 4. (A, B) Distributions of O1K scores (A) and AUROC scores (B) obtained when comparing r_S-based GCNs inferred from 16 cancer tissue gene expression datasets to expert-curated networks and their randomized counterparts. Statistical significance was determined via a one-sided Wilcoxon signed-rank test (alternative hypothesis: score distributions are larger for original than for rewired expert-curated networks) with significance threshold 0.05. (C, D) Distributions of O1K (C) and AUROC (D) scores obtained for cancer datasets in comparison to the scores obtained for healthy tissue data. Statistical significance was determined via a two-sided Wilcoxon signed-rank test with significance threshold 0.05.



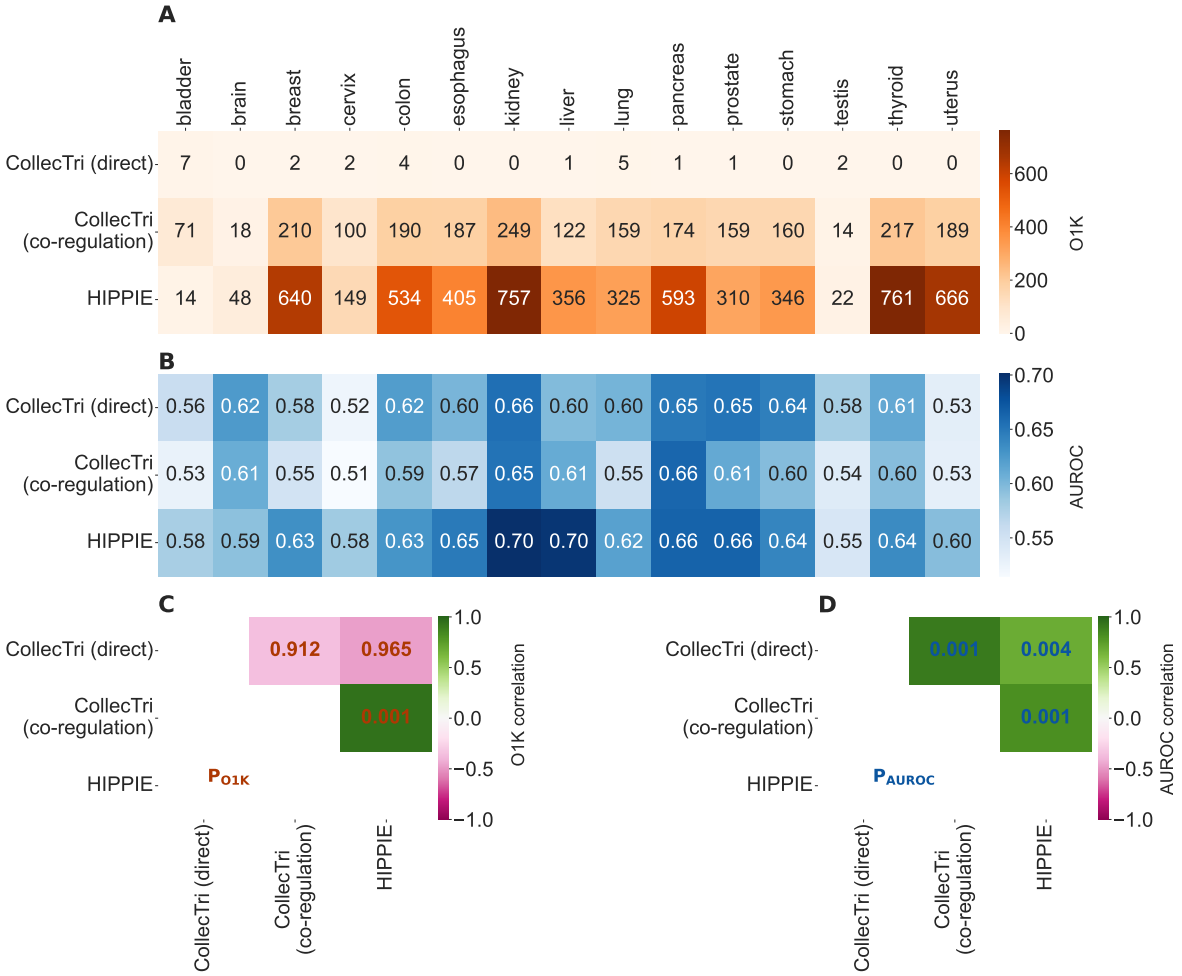
Supplementary Figure 5. (A, B) Distributions of O1K scores (A) and AUROC scores (B) obtained when comparing τ-based GCNs inferred from 16 cancer tissue gene expression datasets to expert-curated networks and their randomized counterparts. Statistical significance was determined via a one-sided Wilcoxon signed-rank test (alternative hypothesis: score distributions are larger for original than for rewired expert-curated networks) with significance threshold 0.05. (C, D) Distributions of O1K (C) and AUROC (D) scores obtained for cancer datasets in comparison to the scores obtained for healthy tissue data. Statistical significance was determined via a two-sided Wilcoxon signed-rank test with significance threshold 0.05.



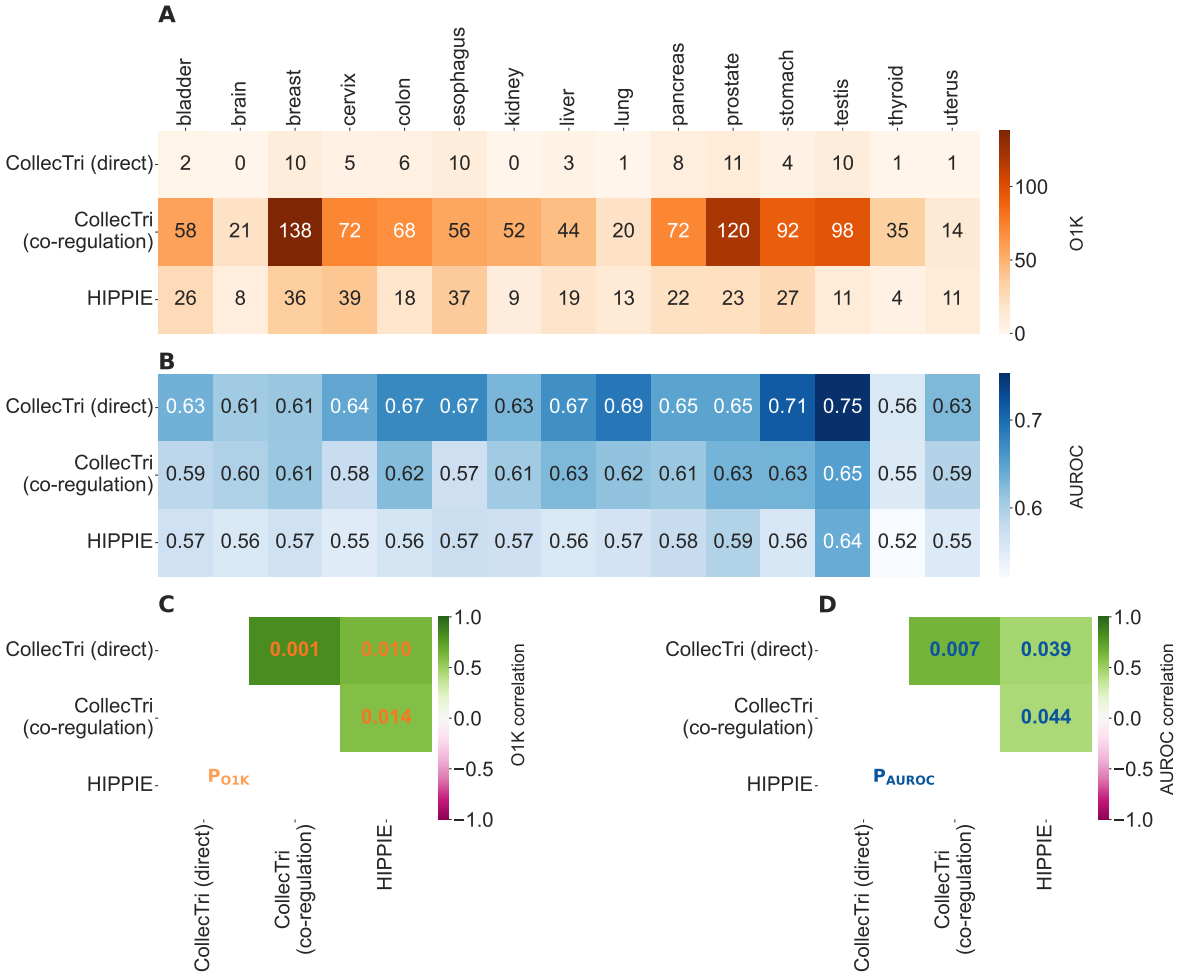
Supplementary Figure 6. Enrichment results for hubs in the CollecTRI GRN. Cancer-associated gene sets are highlighted in red. (A, B) Top 10 most significantly enriched DisGeNET (A) and KEGG (B) gene sets for the top 20 TFs in CollecTRI with the largest numbers of regulated target genes (SP1, TP53, MYC, NFKB, JUN, ESR1, AP1, RELA, AR, CREB1, HIF1A, CEBPB, SP3, STAT3, NFKB1, PPARG, EGR1, SPI1, FOS, CTNNB1). (C, D) Top 10 most significantly enriched DisGeNET (C) and KEGG (D) gene sets for the top 20 target genes in CollecTRI with the largest numbers of in-going regulatory links (CDKN1A, CCND1, TNF, MYC, CDKN2A, VEGFA, TP53, IL6, BCL2, CDH1, INS, ESR1, IFNG, PTGS2, CDKN1B, CXCL8, IL2, AR, JUN, TERT).



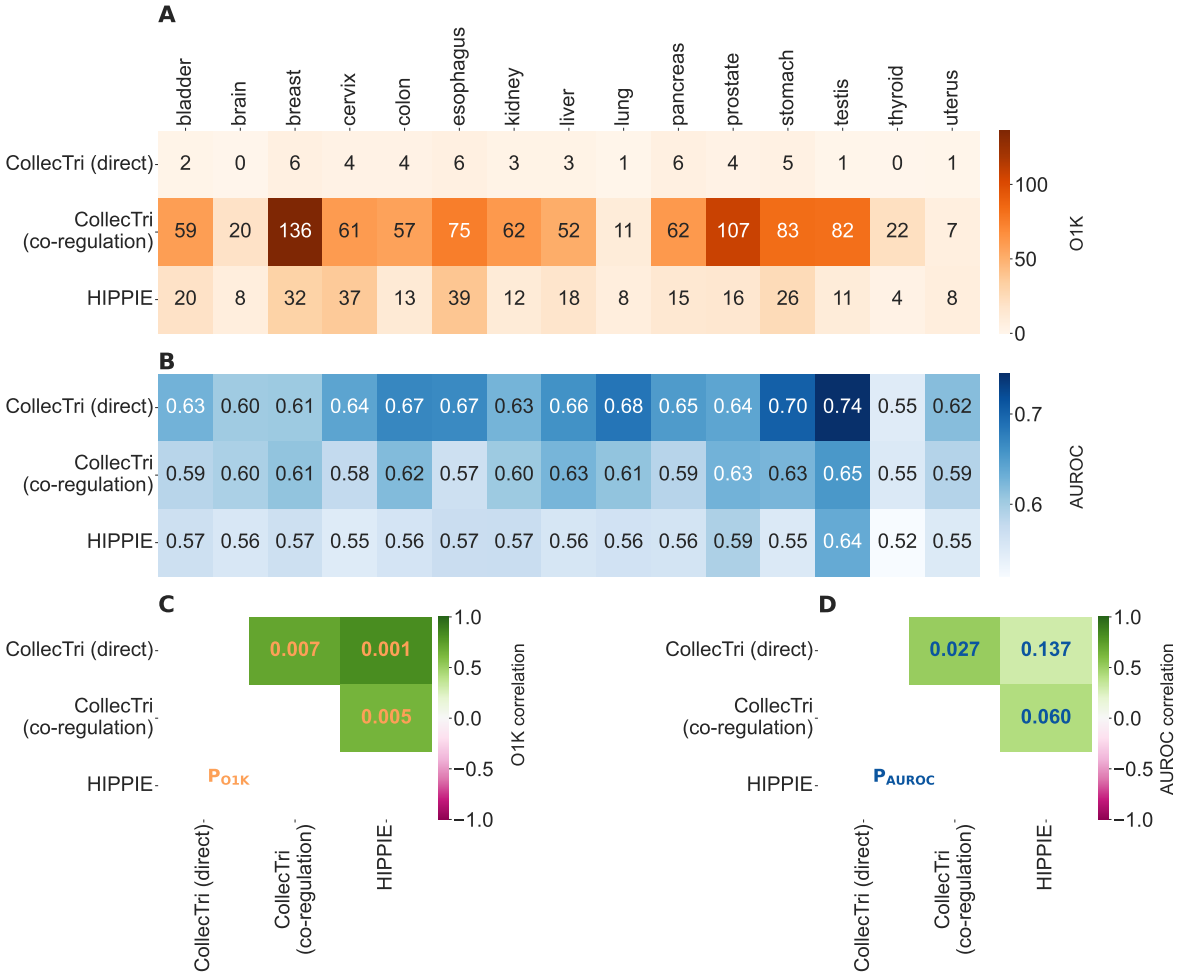
Supplementary Figure 7. (A, B) Individual r_S-based O1K (A) and AUROC (B) scores obtained for the different healthy tissue datasets and the different expert-curated networks. (C) Spearman correlation coefficients r_S between the O1K score distributions shown in the rows of (A), annotated with permutation-based P-values (alternative hypothesis: $r_S > 0$). (D) Spearman correlation coefficients r_S between the AUROC score distributions shown in the rows of (B), annotated with permutation-based P-values P_{AUROC} (alternative hypothesis: $r_S > 0$).



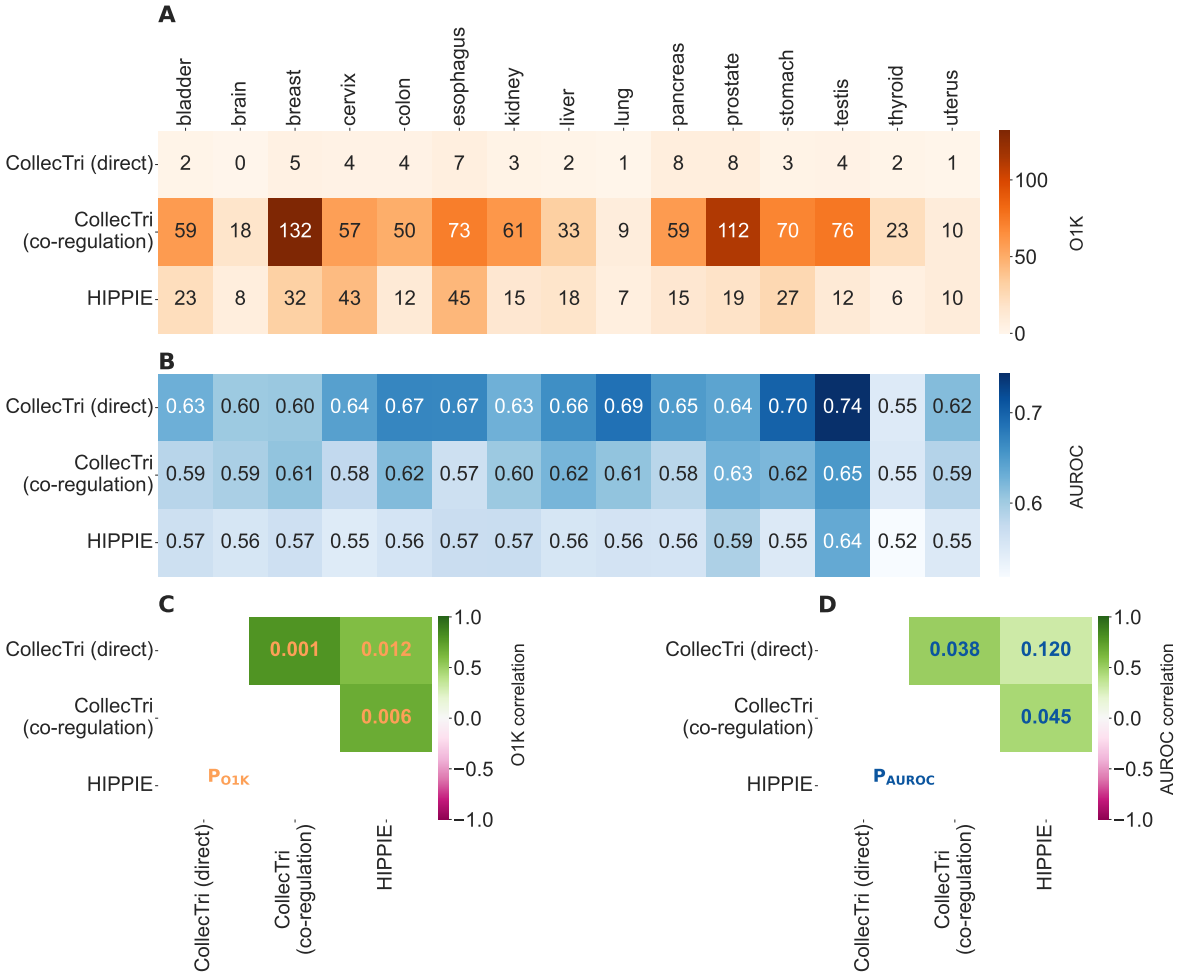
Supplementary Figure 8. (A, B) Individual τ-based O1K (A) and AUROC (B) scores obtained for the different healthy tissue datasets and the different expert-curated networks. (C) Spearman correlation coefficients r_S between the O1K score distributions shown in the rows of (A), annotated with permutation-based P-values P_{O1K} (alternative hypothesis: $r_S > 0$). (D) Spearman correlation coefficients r_S between the AUROC score distributions shown in the rows of (B), annotated with permutation-based P-values P_{AUROC} (alternative hypothesis: $r_S > 0$).



Supplementary Figure 9. (A, B) Individual r_P-based O1K (A) and AUROC (B) scores obtained for the different cancer tissue datasets and the different expert-curated networks. (C) Spearman correlation coefficients r_S between the O1K score distributions shown in the rows of (A), annotated with permutation-based P-values P_{O1K} (alternative hypothesis: $r_S > 0$). (D) Spearman correlation coefficients r_S between the AUROC score distributions shown in the rows of (B), annotated with permutation-based P-values P_{AUROC} (alternative hypothesis: $r_S > 0$).



Supplementary Figure 10. (A, B) Individual r_S-based O1K (A) and AUROC (B) scores obtained for the different cancer tissue datasets and the different expert-curated networks. (C) Spearman correlation coefficients r_S between the O1K score distributions shown in the rows of (A), annotated with permutation-based P-values P_{O1K} (alternative hypothesis: $r_S > 0$). (D) Spearman correlation coefficients r_S between the AUROC score distributions shown in the rows of (B), annotated with permutation-based P-values P_{AUROC} (alternative hypothesis: $r_S > 0$).



Supplementary Figure 11. (A, B) Individual τ-based O1K (A) and AUROC (B) scores obtained for the different cancer tissue datasets and the different expert-curated networks. (C) Spearman correlation coefficients r_S between the O1K score distributions shown in the rows of (A), annotated with permutation-based P-values P_{O1K} (alternative hypothesis: $r_S > 0$). (D) Spearman correlation coefficients r_S between the AUROC score distributions shown in the rows of (B), annotated with permutation-based P-values P_{AUROC} (alternative hypothesis: $r_S > 0$).