Intratumor Heterogeneity Through the Lens of Gene Regulatory Networks

Narein Rao[1,†], Nicolae Sapoval[1,†], Hamim Zafar[2,3,4,*], and Luay Nakhleh[1,5,*]

[1]Department of Computer Science, Rice University, Houston TX 77005, USA
[2]Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, India
[3]Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, India
[4]Mehta Family Center for Engineering in Medicine, Indian Institute of Technology Kanpur, India
[5]Department of BioSciences, Rice University, Houston TX 77005, USA
[*]Corresponding authors: hamim@iitk.ac.in, nakhleh@rice.edu
[†]These authors have contributed equally to this work.

April 1, 2025

Abstract

Tumors exhibit highly heterogeneous genomic and transcriptomic landscapes. Characterization of this intratumor heterogeneity (ITH) and its implications for the cancer therapies has been one of the focal points of computational cancer biology in the past years. However, while ITH at the level of genomic changes, such as copy number aberrations, has been well characterized, the downstream impact of ITH on the gene regulatory landscape remains underexplored. Disruptions in gene regulatory programs can lead to emergence of more aggressive or drug-resistant tumor clones. In order to better understand whether the evidence for ITH at the gene regulatory level can be inferred from scRNA-seq data we have conducted a series of experiments that explore the levels of gene regulatory network (GRN) discrepancies between the clonal subpopulations of tumor cells. In particular, we quantify the extent to which GRNs inferred from different tumor clones diverge from each other in three settings: triple negative breast cancer, colorectal cancer, and a longitudinal dataset from a patient undergoing chemotherapy for breast cancer. Our analyses indicate that a substantial (15-25%) proportion of GRN edges can be attributed to clone-specific activity. Furthermore, we observe differential network patterns across outlier transcription factors implicated in tumorogenesis and tumor progression. Our findings suggest that differential network patterns between clonal subpopulations can be reliably identified, and that such patterns offer a complementary insight that is not captured by differential gene analyses.

1 Introduction

Cancer is a complex disease characterized by high genomic and transcriptomic diversity [9] both between different cancers and patients, and within a single tumor. The latter diversity is called intratumor heterogeneity (ITH) and presents a major obstacle in design and delivery of efficient and personalized cancer therapies [9, 39, 14]. A common way to characterize ITH relies on the clonal cancer evolution hypothesis [42] and consists of identifying clonal subpopulations of cells typically characterized by their somatic single-nucleotide variant (SNV) [62, 20], copy-number aberration (CNA) profiles [15] or the combination of both [49, 13]. However, while the inference of clonal subpopulations from single-cell DNA/genomic data (scDNA) [61, 51, 63] and single-cell RNA/transcriptomic data (scRNA) [1, 23] has become common in the modern studies of tumor evolution, assessment of the clone-specific changes to gene regulatory networks (GRNs) remains limited. Prior work in this area includes: general characterization of the GRNs [31] or quantification of the ITH [45], annotation of the protein-interaction networks with scRNA-seq expression data [66], studies targeting specific cancer hallmark pathways [30, 54] and comparisons of tumor-derived GRNs with the normal cell-derived GRNs [16]. However, the heterogeneity of transcriptional regulatory networks across the tumor clones within the same tumor or across different time points remains largely unexplored.

In this work, we test the hypothesis that clone-specific changes in GRNs can be inferred and quantified from scRNA-seq datasets. In particular, we analyze three single-cell cancer datasets from colorectal and breast cancer, and validate that our hypothesis holds for: (a) copy-number clones inferred from scRNA-seq data [23], (2) copy-number clones inferred from scDNA-seq data with clone annotations mapped to the scRNA-seq data [6], and (3) longitudinal scRNA-seq data corresponding to stage IV Breast Cancer pretreatment and posttreatment samples [44]. We observe that around 15-25% of the edges in the clone-specific GRNs are unique to that clone, and these edges are well supported in the subsampling based bootstraps. We could also identify transcription factors that are only present in one of the inferred clone-specific GRNs. At the level of the local differences between the transcription factors shared by the clone-specific GRNs, we are able to identify significant outliers with different inferred connectivity patterns. Overall our analyses suggest that we can reliably identify differential GRN patterns between clonal cell populations and provide complementary information on ITH that is not revealed by differential gene expression analyses.

2 Methods

2.1 Datasets

In our study we have analyzed three single-cell cancer datasets from patients with triple-negative breast cancer (TNBC) [23], colorectal cancer (CRC) [6], and stage IV breast cancer (BC) [44]. The TNBC dataset analyzed in this work consists of two scRNA-seq samples TNBC1 and TNBC2 from the original study with tumor clone annotations based on CopyKat copy-number profiles inferred by the authors of the original study [23]. The CRC dataset consists of a single scTrio-seq sample, the clone annotations were based on DNA copy-number profiles from the original study and were carried over to the single-cell transcriptomic data via MaCroDNA [19]. The BC dataset consists of three longitudinal scRNA-seq samples: one pretreatment biopsy, and two post-treatment samples collected from the surgically removed tumor [44]. For the BC dataset, normal cells were identified using SingleR [4] according to the protocol provided in the original publication [44], and the tumor clones were inferred based on the copy-number profiles obtained with inferCNV [1] package.

2.2 Gene regulatory network inference

We have originally considered three state-of-the-art GRN inference methods for this study: CellOracle [33], GRNboost2 [40] as a part of the SCENIC [3] pipeline, and Inferelator [52]. We have constrained ourselves to GRN inference methods built for scRNA-seq data for two key reasons. First, scRNA-seq data is common in the cancer studies, with several studies analyzing paired [6] or unpaired RNA-seq and DNA-seq data [19]. In contrast, there are currently fewer studies that attempt to integrate other data modalities such as ATAC-seq or CITE-seq into the inference and analysis of the clonal subpopulations within cancer. Second, the methods we considered are commonly employed in the field, and hence our evaluations can provide a useful insight for the practitioners. The aforementioned methods were executed using their default settings. Due to the absence of prior information required for running Inferelator, we utilized the output of GRNBoost2 as a prior network matrix for Inferelator. We observed that for the TNBC and CRC data the GRN inferred from all tumor cells had few common edges with the clone-specific GRNs for GRNboost2 and Inferelator (Supp. Fig. 9). Thus, we have only focused on the GRNs inferred by CellOracle for the subsequent analyses.

2.3 Bootstrap analysis

Bootstrap analysis was performed to assess the confidence in the inferred GRN edges. For each clone, 100 independent subsamples were drawn without replacement consisting of a fraction $f \in \{0.75, 0.9, 0.95\}$ of all cells in the clone. For each subsample, a GRN was inferred similarly to the main analysis. Edges appearing in the clone-specific GRN inferred from the complete data were then annotated with their support values based on the number of times the edge appeared in the corresponding bootstrap experiment.

An additional bootstrapping experiment was performed for TNBC1 and TNBC2 samples. We randomly shuffled clone annotation labels on the tumor cells in the sample creating 100 independent samples. We then inferred the GRN specific to each partitioning of the data, thus creating a null model for arbitrary cell partitioning.

2.4 Graph similarity measures

In order to evaluate the similarity of the inferred gene regulatory networks, we employed both global and local measures of graph similarity. For a global measure of the graph similarity, we employed the Jaccard similarity between the edge sets of two inferred GRNs. More formally, for two directed graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, the Jaccard similarity of their edge sets is defined as $J_{sim}(G_1, G_2) = \frac{|E_1 \cap E_2|}{|E_1 \cup E_2}$. A Jaccard similarity value of 1 would indicate two identical edge sets, and a Jaccard similarity value of 0 would indicate no shared edges between the two networks. We also use the Hamming distance between edge membership vectors, which mathematically is simply defined as $H_{dist} = 1 - J_{sim}$ in this case.

The local measure of graph similarity we employ also computes the Jaccard similarity, but between the edge sets originating from ($J^{succ}(v)$) or incident to ($J^{pred}(v)$) a fixed node $v \in V_1 \cap V_2$. Thus, if we let $N_i^{succ}(v) = \{u \in V_i | (v, u) \in E_i\}$ and $N_i^{pred}(v) = \{u \in V_i | (u, v) \in E_i\}$ denote the successors and predecessors of a node v in the GRN G_i, then $J^{succ}(v) = \frac{|N_1^{succ}(v) \cap N_2^{succ}(v)|}{|N_1^{succ}(v) \cup N_2^{succ}(v)|}$ and $J^{pred}(v) = \frac{|N_1^{pred}(v) \cap N_2^{pred}(v)|}{|N_1^{pred}(v) \cup N_2^{pred}(v)|}$.

2.5 Centrality measures

In order to identify "key" nodes within the inferred GRNs, we computed node betweenness centrality [21] — a commonly used measure of importance of a node in a network [31], based on the number of shortest paths that pass through that node. We utilized the NetworkX Python package (v2.8.8) [26] to calculate the centrality measures in this study.

3 Results

3.1 Triple negative breast cancer

First, we evaluated the overlaps and differences between edges in GRNs inferred from the two major clonal cell populations in triple negative breast cancer (TNBC) tumor samples TNBC1 and TNBC2. We observe that the majority of the edges (74.68% of clone 1 and 75.34% of clone 2 edges in TNBC1; 74.49% of clone 1 and 74.23% of clone 2 edges in TNBC2) are shared between the major clones in TNBC1 and TNBC2 samples (Fig. 1A,C). In contrast, the tumor-specific GRNs and the GRN inferred from the normal cells shared a much smaller number of overlapping edges (Fig. 1A,C). Additionally, the similarity between the inferred clone-specific GRNs measured as the Jaccard similarity of the corresponding sets of edges (TNBC1: 0.6, TNBC2: 0.592), is more than 6 standard deviations (TNBC1: 0.0035, TNBC2: 0.008) below the mean Jaccard similarity of GRNs (TNBC1: 0.684, TNBC2: 0.657) inferred from random shuffles of the tumor cells data (Supp. Fig. 10). Furthermore, we note that the distributions of the bootstrap support values for all inferred edges within clone-specific GRNs match closely the distributions of the support values for the edges that are unique to those clone specific GRNs (Fig. 1B,D).

To further assess dissimilarities between the clone-specific GRNs, we evaluated the Hamming distances between edge vectors (i.e. edge presence/absence) for bootstrap subsamples (subsampling fraction 95%) drawn from the same major clones and the ones for the subsamples coming from different clones (Fig. 2). We observe that in both cases the distances between the subsamples drawn from the same clone are significantly smaller (t-test p-values for all comparisons $< 10^{-5}$) than those between the bootstrap GRNs inferred from different clones (Fig. 2).

In order to better understand the genes driving the observed differences in the clone-specific GRNs, we analyzed the degree distributions for GRN nodes in the clone-specific GRNs (Fig. 3). For the majority of genes the degree of the corresponding node in the GRN is stable between the two major clones (Fig. 3, $R^2 = 0.9876$ for TNBC1 and $R^2 = 0.9391$ for TNBC2). However, we have identified several outliers that have high (> 100) degrees in one of the clone-specific GRNs, while not appearing at all in the other one. In the TNBC1 sample *GLI1* [27, 7, 29], *LHX8*, *ATOH1 [50, 25]*, and *OTP* [48] genes are such outliers for clone 1, and *EOMES* [28] and *PAX2* [53] are the corresponding outliers for clone 2 (Fig. 3A). These outliers are further supported by the bootstrap GRNs (subsampling fraction 95%), with *GLI1* appearing in 98 clone 1 bootstraps and 0 clone 2 bootstraps, analogously *LHX8* has 95:0 support, *ATOH1* has 96:0 support, *OTP* has 95:0 support, *EOMES* has 0:96 support, and *PAX2* has 0:97 support. In the TNBC2 sample *GATA1* [47, 8, 22] is the outlier for clone 1, and *SP5* [11] is the corresponding outlier for clone 2 (Fig. 3B). The *GATA1* outlier has mixed support in bootstrap experiments appearing in 96 subsamples from clone 1 and 66 subsamples from clone 2, while *SP5* appears in 0 subsamples from clone 1 and 90 subsamples from clone 2. Notably, out of the 8 genes identified above none are marked as significantly differentially expressed between the two clones by the standard differential expression analyses.

Since the inferred networks have directed edges, we also quantified nodes that have significantly

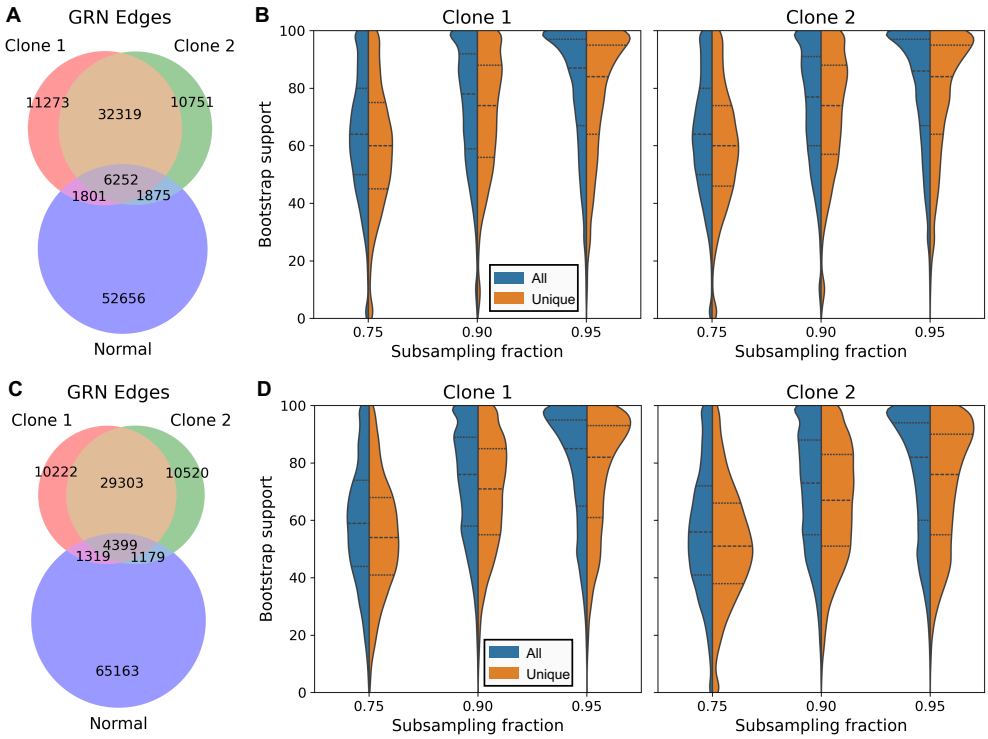

Figure 1: **(A)** Venn diagram showing unique and shared edges between GRNs inferred from tumor clones 1 and 2, and normal cells in the TNBC1 sample. **(B)** Violin plots showing bootstrap support values for edges in GRNs inferred from clone 1 and clone 2 cells in the TNBC1 sample. Bootstrap values are based on 100 subsampling runs retaining 75%, 90%, and 95% of cells in the clone per subsample respectively (x-axis). Colors indicate distributions of support values for all edges in the GRN (blue) and edges unique to the clone's GRN (orange). **(C)** Venn diagram showing unique and shared edges between GRNs inferred from tumor clones 1 and 2, and normal cells in the TNBC2 sample. **(D)** Violin plots showing bootstrap support values for edges in GRNs inferred from clone 1 and clone 2 cells in the TNBC2 sample.

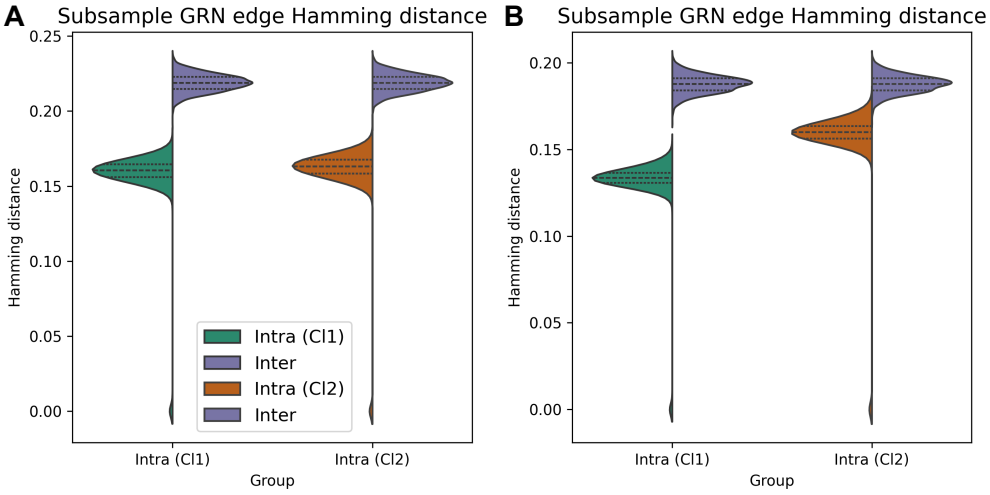

Figure 2: **(A)** Distribution of Hamming distances between edge vectors of the bootstrap GRNs inferred from TNBC1 sample. Intra-clone 1 (green) and intra-clone 2 (orange) distances are measured between all pairs of bootstrap subsamples derived from the same clone, inter-clone distances are measured between all pairs of bootstrap subsamples derived from different clones. **(B)** Distribution of Hamming distances between edge vectors of the bootstrap GRNs inferred from TNBC2 sample. Intra-clone 1 (green) and intra-clone 2 (orange) distances are measured between all pairs of bootstrap subsamples derived from the same clone, inter-clone distances are measured between all pairs of bootstrap subsamples derived from different clones.

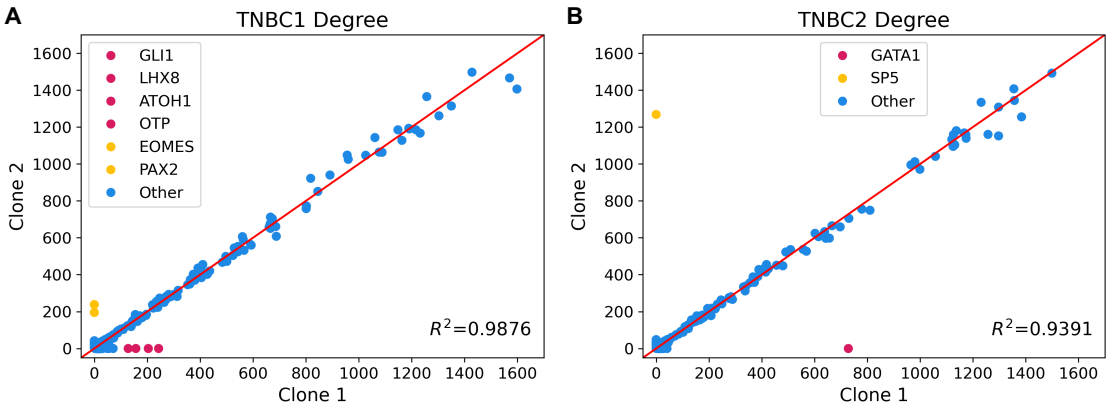

Figure 3: **(A)** Vertex degrees based on clone 1 (x-axis) and clone 2 (y-axis) GRNs inferred from the TNBC1 sample. Coefficient of determination (R^2) for $\hat{y} = x$ is reported. Outliers that have degree 0 in one of the GRNs and degree > 100 in another one are reported. **(B)** Vertex degrees based on clone 1 (x-axis) and clone 2 (y-axis) GRNs inferred from the TNBC1 sample. Coefficient of determination (R^2) for $\hat{y} = x$ is reported. Outliers that have degree 0 in one of the GRNs and degree > 100 in another one are reported.

different successor and predecessor connections between the clone-specific GRNs. Namely, we have computed the Jaccard similarity between the sets of successors in clone-specific GRNs for each node and identified outliers with the Jaccard similarity at least 3 standard deviations below the mean (Fig. 4). For the TNBC1 sample there is one outlier gene based on the successor Jaccard similarity *HERPUD1* [41], and 8 outlier genes based on the predecessor Jaccard similarity *TSPAN12*, *GEM*, *ITGA11*, *CDH19*, *DSCC1*, *AREG*, *BATF*, *ATP1B4*. For the TNBC2 sample there is one outlier gene based on the successor Jaccard similarity *ARID5A* [43], and 7 outlier genes based on the predecessor Jaccard similarity *SFTPB*, *FCRL3*, *LINC00111*, *KCNV2*, *PCDHB5*, *TNFRSF1B*, *NDP*.

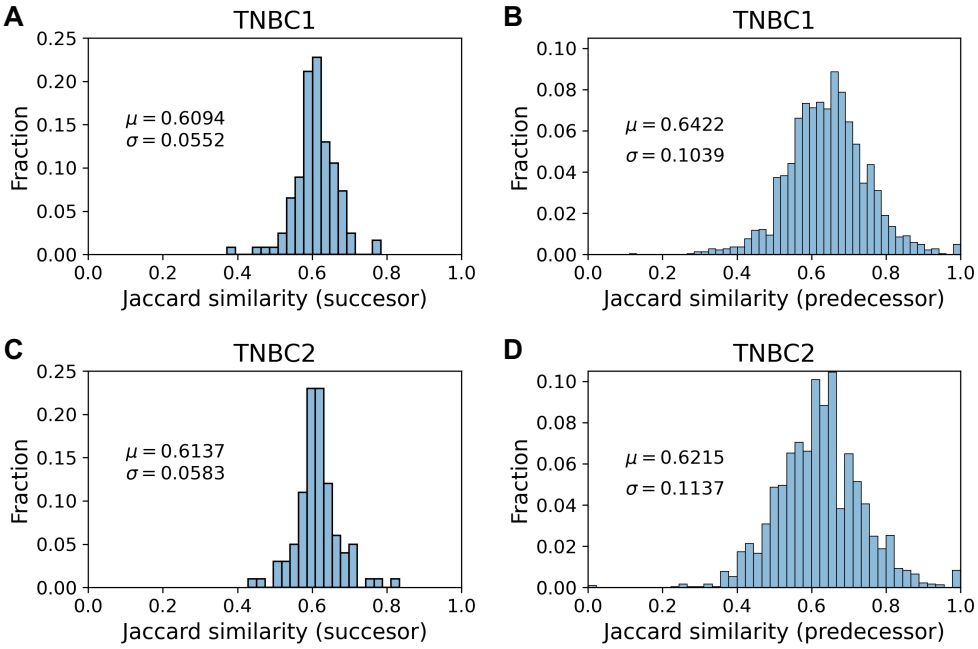

Figure 4: Distributions of Jaccard similarities for the successor **(A, C)** and predecessor **(B, D)** node sets between clone-specific GRNs in: **(A, B)** TNBC1 and **(C, D)** TNBC2 samples.

3.2 Colorectal cancer

Similarly to the TNBC data, we started by evaluating the overlaps and differences between edges in GRNs inferred from the two major clonal cell populations in colorectal cancer (CRC) sample. We observe that the majority of the edges (88.04% of clone 1 and 86.88% of clone 2) edges are shared between the major clones in the CRC sample (Fig. 5A). We also observe that the distances between the subsamples drawn from the same clone are significantly smaller (t-test p-values for all comparisons $< 10^{-5}$) than those between the bootstrap GRNs inferred from different clones (Fig. 5B). Additionally, the similarity between the inferred clone-specific GRNs measured as the Jaccard similarity of the corresponding sets of edges is 0.777, and is more than 30 standard deviations (0.0017) below the mean Jaccard similarity 0.829 of the GRNs inferred from random shuffles of the tumor cells data. Furthermore, we note that the distributions of the bootstrap support values for all inferred edges within clone-specific GRNs match closely the distributions of the support values for the edges that are unique to those clone specific GRNs (Fig. 5C).

Unlike for the TNBC samples, the CRC data did not have any nodes that had degree 0 in one of the clones and not the other, and the overall agreement between node degrees in the two

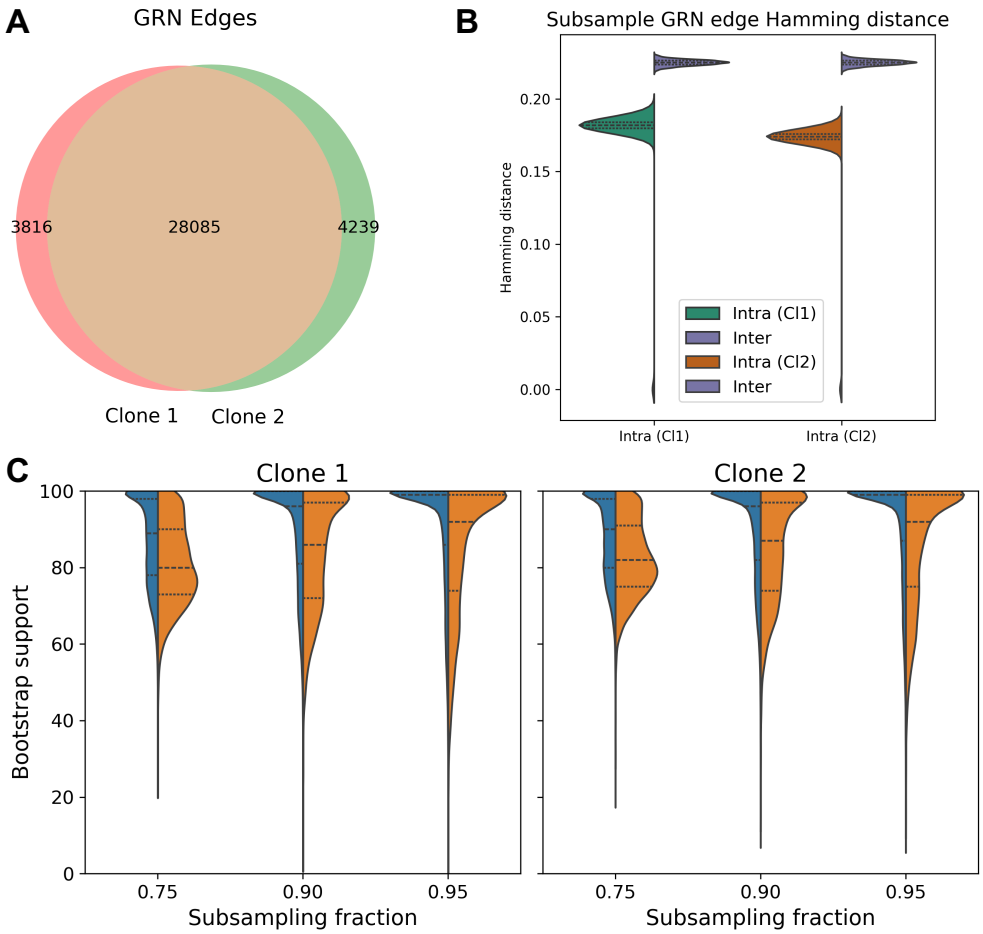

Figure 5: **(A)** Venn diagram showing unique and shared edges between GRNs inferred from tumor clones 1 and 2 in the CRC sample. **(B)** Distribution of Hamming distances between edge vectors of the bootstrap GRNs inferred from the CRC sample. Intra-clone 1 (green) and intra-clone 2 (orange) distances are measured between all pairs of bootstrap subsamples derived from the same clone, inter-clone distances are measured between all pairs of bootstrap subsamples derived from different clones.. **(C)** Violin plots showing bootstrap support values for edges in GRNs inferred from clone 1 and clone 2 cells in the CRC sample. Bootstrap values are based on 100 subsampling runs retaining 75%, 90%, and 95% of cells in the clone per subsample respectively (x-axis). Colors indicate distributions of support values for all edges in the GRN (blue) and edges unique to the clone's GRN (orange).

clone-specific GRNs was high ($R^2 = 0.9984$, Supp. Fig. 11). In order to explore which edges drive the differences between the two inferred GRNs, we have computed the Jaccard similarity between the sets of successors in clone-specific GRNs for each node and identified outliers with the Jaccard similarity at least 3 standard deviations below the mean (Fig. 6). For the CRC sample there were no outlier genes based on the successor Jaccard similarity, and 7 outlier genes based on the predecessor Jaccard similarity *TMEM38A*, *KYNU* [65], *RAB26* [67], *PTOV1* [37, 57], *MTG1*, *C19orf47*, *ZNF507* [36].

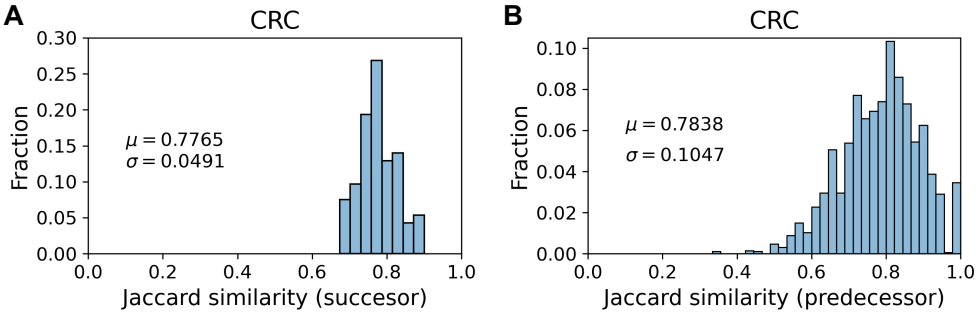

Figure 6: Distributions of Jaccard similarities for the successor **(A)** and predecessor **(B)** node sets between clone-specific GRNs in the CRC sample.

3.3 Longitudinal data from breast cancer

For the longitudinal data from a breast cancer patient we have focused on GRN differences between the pretreatment tumor sample (Pre), and two posttreatment tumor samples: one originating from the vicinity of the central necrotic tissue (Post1), and one containing viable cancer cells collected near tumor margins (Post2) [44]. Unlike for the clonal subpopulations within the same sample, the differences in edges between the inferred GRNs for the three samples were large: out of 56,290 edges inferred for the Pre tumor GRN only 10.45% (5,882) were shared with Post1 edges, and 17.99% (10,128) with Post2 edges. In total there are 2,812 edges common to all three tumor-specific GRNs, and 7,841 edges shared between Post1 and Post2. Due to the large differences in the inferred GRNs there are many nodes that are specific to a single GRN and have a high degree (Supp. Fig. 12). Among those nodes *ETS1* [35], *KLF5* [46, 32], and *KLF3* [34] have the highest total sum of their degrees in Post1 and Post2 GRNs while not apprearing in the Pre GRN. Additionally, these nodes are well supported by the subsampling bootstrap values, having support in Pre:Post1:Post2 subsamples of $0 : 100 : 68$, $0 : 100 : 100$, and $0 : 85 : 100$, respectively. Similarly, *PAX5* [12, 10] and *PRDM1* [68] only appear in the Pre GRN and are supported by the bootstraps at $95 : 0 : 0$ and $100 : 0 : 0$, respectively.

Given the large discrepancies between the inferred GRNs, we focused on the nodes with high betweenness centrality for further evaluation of the potential drivers of key differences between the three samples (Fig. 7). *ATF3* transcription factor has the highest betweenness centrality in the Pre and Post2, and is ranked at the 11th position by betweenness centrality in the GRN inferred from Post1. On the other hand, *TEAD2* [17, 24] gene has high centrality (ranked 2nd) in Post1, but is absent as a GRN node in Pre and Post2. Similarly, *PLAGL1* [2, 27] and *NFE2* [64, 56] genes have high centrality (ranked 2nd and 3rd, respectively) in the Post2, but do not appear as nodes in the GRNs inferred from Pre and Post1 tumor data.

Next, we explored local differences between the inferred GRNs based on the connectivity patterns

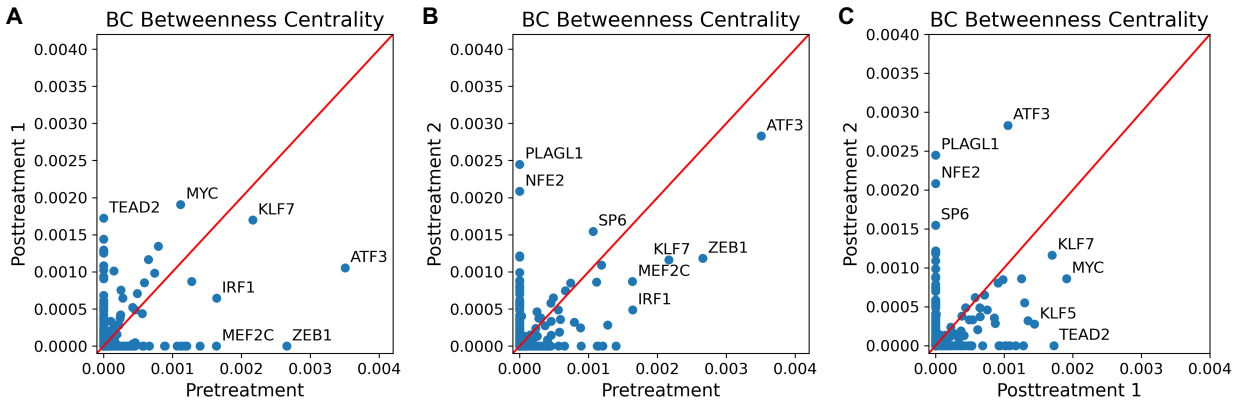

Figure 7: Betweenness centrality for the nodes present in GRNs inferred from the BC data. **(A)** Centrality based on pretreatment sample (x-axis) and posttreatment sample 1 (y-axis) GRNs inferred from the BC dataset. **(B)** Centrality based on pretreatment sample (x-axis) and posttreatment sample 2 (y-axis) GRNs inferred from the BC dataset. **(C)** Centrality based on posttreatment sample 1 (x-axis) and posttreatment sample 2 (y-axis) GRNs inferred from the BC dataset.

of the shared nodes. First, we note that unlike in the case of clonal subpopulations in TNBC1, TNBC2, and CRC the mean Jaccard similarity between the neighbor node sets for shared nodes between the tumor samples is low: mean similarities for successor node sets are 0.1497, 0.1709, and 0.1521 for Pre vs Post1, Pre vs Post2, and Post1 vs Post2 respectively, and the mean similarities for the predecessor nodes sets are 0.2205, 0.2846, and 0.2619 (Fig. 8). As the result of these large uniform discrepancies between the GRNs, only *SOX4* [5] and *HERPUD1* [41] were identified as genes with Jaccard similarity of the successor node sets in Pre and Post2 GRNs at least 3 standard deviations below the mean. Similarly, for the Jaccard similarity of the predecessor node sets we identified *CA2* [55], *ASB2*, *RPL32* [58], *SNAPC1*, *MIA*, and *CLEC2D* [38] as significantly differential for Pre vs Post2 GRNs, and *COL3A1* [60, 59, 18], *CA2* [55], *ASB2*, *CDH19*, and *SLA2* as significantly differential for Post1 vs Post2 GRNs.

4 Discussion

In this work we explored the impact ITH can have on the GRN structure in tumors. Specifically, we assessed the differences between clone-specific GRNs inferred from copy-number based tumor clones.

Our results indicate that analyzing discrepancies between the clone-specific GRNs can offer a complementary characterization of the tumor clones. In particular, we note that the *HERPUD1* gene, identified by our analysis of the TNBC1 sample, has been shown to inhibit apoptosis in ovarian cancer [41], and the *ARID5A* gene, identified by our analysis of the TNBC2 sample, has been previously linked to breast cancer invasion and metastasis [43]. Additionally, *PAX2* [53], *SP5* [11], *GLI1* [27, 7, 29], *ATOH1* [50, 25], *GATA1* [47, 8, 22] genes, that we have identified as the nodes with differential degrees between clone-specific GRNs, were linked to several oncogenic pathways and functions over multiple prior studies. Notably none of these genes were identified by the differential gene expression analyses of the scRNA-seq data, hence suggesting that the strength of the network based analysis lies in uncovering indirect interactions.

Furthermore while we observed a higher degree of concordance between the clone-specific GRNs

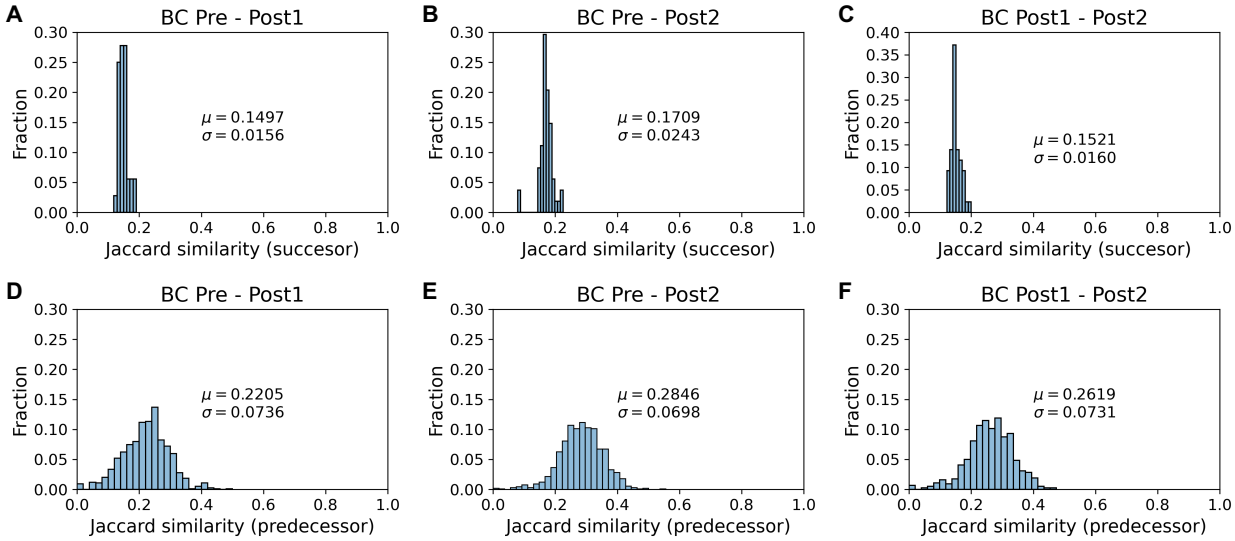

Figure 8: Distributions of Jaccard similarities for the successor **(A-C)** and predecessor **(D-F)** node sets between pretreatment, posttreament sample 1 and posttreatment sample 2 tumor-specific GRNs in the BC dataset.

in the CRC data, we were still able to identify several genes with differential regulatory patterns between the two tumor clones. Notably, among these genes are *PTOV1* [37, 57], *KYNU* [65], *RAB26* [67], and *ZNF507* [36] which have been linked to colorectal and prostate cancer oncogenesis and progression. Since the clone assignments were inferred based on the pairing of DNA-seq and RNA-seq data by MacRoDNA [19], our finding provides additional validation for the hypothesis that the clonal substructure of the cancer cell subpopulations has an impact on the gene regulation and transcriptional activity patterns.

Finally, we noticed dramatic differences in the GRNs inferred from tumor samples pre- and post-treatment. However, despite the lack of a large conserved sub-network, we were still able to refine the analysis to a smaller set of significantly differential gene regulatory patterns. In particular, similarly to the TNBC1 sample we indetified differential connectivity patterns for the *HERPUD1* gene [41], as well as *COL3A1* [60, 59, 18], and *CLEC2D* [38] which are known to be involved in important cancer pathways.

A current limitation of our analyses includes semi-manual extraction of the features of interest from the inferred GRNs. While bootstrapping provides additional statistical support for the findings, identification of the outlier nodes and edges in the inferred GRNs is complicated by two factors: (1) the inferred networks are large having thousands of nodes and tens of thousands of edges, and (2) the majority of the edges, as well as the overall network structure is highly conserved between the clones. Thus, methods focusing on global network properties and discrepancy measures are ineffective at properly quantifying differences of interest. In the future work, we plan to extend methods for quantifying local differences between clone-specific GRNs, and automate subsequent feature extraction based on the bootstrap support scores.

5 Acknowledgements

This research was supported in part by the National Science Foundation, grants IIS-1812822 and IIS-2106837 (L.N.) and the DBT/Wellcome Trust India Alliance grant IA/E/21/1/506298 (H.Z.).

References

[1] inferCNV of the Trinity CTAT Project. https://github.com/broadinstitute/inferCNV.

[2] Abbas Abdollahi, Debra Pisarcik, David Roberts, Jillian Weinstein, Paul Cairns, and Thomas C Hamilton. Lot1 (plagl1/zac1), the candidate tumor suppressor gene at chromosome 6q24–25, is epigenetically regulated in cancer. *Journal of Biological Chemistry*, 278(8):6041–6049, 2003.

[3] Sara Aibar, Carmen Bravo González-Blas, Thomas Moerman, Vân Anh Huynh-Thu, Hana Imrichova, Gert Hulselmans, Florian Rambow, Jean-Christophe Marine, Pierre Geurts, Jan Aerts, et al. Scenic: single-cell regulatory network inference and clustering. *Nature Methods*, 14(11):1083–1086, 2017.

[4] Dvir Aran, Agnieszka P Looney, Leqian Liu, Esther Wu, Valerie Fong, Austin Hsu, Suzanna Chak, Ram P Naikawadi, Paul J Wolters, Adam R Abate, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature immunology*, 20(2):163–172, 2019.

[5] Archis Bagati, Sushil Kumar, Peng Jiang, Jason Pyrdol, Angela E Zou, Anze Godicelj, Nathan D Mathewson, Adam NR Cartwright, Paloma Cejas, Myles Brown, et al. Integrin αvβ6–tgfβ–sox4 pathway drives immune evasion in triple-negative breast cancer. *Cancer cell*, 39(1):54–67, 2021.

[6] Shuhui Bian, Yu Hou, Xin Zhou, Xianlong Li, Jun Yong, Yicheng Wang, Wendong Wang, Jia Yan, Boqiang Hu, Hongshan Guo, et al. Single-cell multiomics sequencing and analyses of human colorectal cancer. *Science*, 362(6418):1060–1063, 2018.

[7] Rebecca LH Bigelow, Nikhil S Chari, Anne Birgitte Undén, Kevin B Spurgers, Sangjun Lee, Dennis R Roop, Rune Toftgard, and Timothy J McDonnell. Transcriptional regulation of bcl-2 mediated by the sonic hedgehog signaling pathway through gli-1. *Journal of Biological Chemistry*, 279(2):1197–1205, 2004.

[8] Pavel Burda, Nikola Curik, Juraj Kokavec, Petra Basova, Dana Mikulenkova, Arthur I Skoultchi, Jiri Zavadil, and Tomas Stopka. Pu.1 activation relieves gata-1–mediated repression of cebpa and cbfb during leukemia differentiation. *Molecular Cancer Research*, 7(10):1693–1703, 2009.

[9] Rebecca A Burrell, Nicholas McGranahan, Jiri Bartek, and Charles Swanton. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, 501(7467):338–345, 2013.

[10] Xueyan Chen, Weilin Wu, Ji-Hak Jeong, Matjaz Rokavec, Rui Wei, Shaolong Feng, Werner Schroth, Hiltrud Brauch, Shangwei Zhong, and Jun-Li Luo. Cytokines-activated nuclear ikkα-fat10 pathway induces breast cancer tamoxifen-resistance. *Science China Life Sciences*, 67(7):1413–1426, 2024.

[11] Yongxin Chen, Yingqiu Guo, Xijin Ge, Hirotaka Itoh, Akira Watanabe, Takeshi Fujiwara, Tatsuhiko Kodama, and Hiroyuki Aburatani. Elevated expression and potential roles of human sp5, a member of sp transcription factor family, in human cancers. *Biochemical and biophysical research communications*, 340(3):758–766, 2006.

[12] Zhao-Hui Chen, Yi-Bo Chen, Hao-Ran Yue, Xue-Jie Zhou, Hai-Yan Ma, Xin Wang, Xu-Chen Cao, and Yue Yu. Pax5-mir-142 feedback loop promotes breast cancer proliferation by regulating dnmt1 and zeb1. *Molecular Medicine*, 29(1):89, 2023.

[13] Ziwei Chen, Fuzhou Gong, Lin Wan, and Liang Ma. Bitsc 2: Bayesian inference of tumor clonal tree by joint analysis of single-cell snv and cna data. *Briefings in Bioinformatics*, 23(3):bbac092, 2022.

[14] Feixiong Cheng, Huixiao Hong, Shengyong Yang, and Yuquan Wei. Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Briefings in bioinformatics*, 18(4):682–697, 2017.

[15] Salim Akhter Chowdhury, E Michael Gertz, Darawalee Wangsa, Kerstin Heselmeyer-Haddad, Thomas Ried, Alejandro A Schäffer, and Russell Schwartz. Inferring models of multiscale copy number evolution for single-tumor phylogenetics. *Bioinformatics*, 31(12):i258–i267, 2015.

[16] Madison Dautle, Shaoqiang Zhang, and Yong Chen. sctiger: A deep-learning method for inferring gene regulatory networks from case versus control scrna-seq datasets. *International Journal of Molecular Sciences*, 24(17), 2023.

[17] Maren Diepenbruck, Lorenz Waldmeier, Robert Ivanek, Philipp Berninger, Phil Arnold, Erik van Nimwegen, and Gerhard Christofori. Tead2 expression levels control the subcellular distribution of yap and taz, zyxin expression and epithelial–mesenchymal transition. *Journal of cell science*, 127(7):1523–1536, 2014.

[18] Haoyuan Du, Hua Wang, Yuwei Luo, Yang Jiao, Jiajun Wu, Shaowei Dong, and Dong Du. An integrated analysis of bulk and single-cell sequencing data reveals that emp1+/col3a1+ fibroblasts contribute to the bone metastasis process in breast, prostate, and renal cancers. *Frontiers in Immunology*, 14:1313536, 2023.

[19] Mohammadamin Edrisi, Xiru Huang, Huw A Ogilvie, and Luay Nakhleh. Accurate integration of single-cell dna and rna for analyzing intratumor heterogeneity using macrodna. *Nature Communications*, 14(1):8262, 2023.

[20] Mohammadamin Edrisi, Monica V Valecha, Sunkara BV Chowdary, Sergio Robledo, Huw A Ogilvie, David Posada, Hamim Zafar, and Luay Nakhleh. Phylovar: toward scalable phylogeny-aware inference of single-nucleotide variations from single-cell dna sequencing data. *Bioinformatics*, 38(Supplement_1):i195–i202, 2022.

[21] LC Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 1977.

[22] A Furuhata, M Murakami, H Ito, S Gao, K Yoshida, S Sobue, R Kikuchi, T Iwasaki, A Takagi, T Kojima, et al. Gata-1 and gata-2 binding to 3' enhancer of wt1 gene is essential for its transcription in acute leukemia and solid tumor cell lines. *Leukemia*, 23(7):1270–1277, 2009.

[23] Ruli Gao, Shanshan Bai, Ying C Henderson, Yiyun Lin, Aislyn Schalck, Yun Yan, Tapsi Kumar, Min Hu, Emi Sei, Alexander Davis, et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nature Biotechnology*, 39(5):599–608, 2021.

[24] Paula González-Alonso, Sandra Zazo, Ester Martín-Aparicio, Melani Luque, Cristina Chamizo, Marta Sanz-Álvarez, Pablo Minguez, Gonzalo Gómez-López, Ion Cristóbal, Cristina Caramés, et al. The hippo pathway transducers yap1/tead induce acquired resistance to trastuzumab in her2-positive breast cancer. *Cancers*, 12(5):1108, 2020.

[25] Saija Haapa-Paananen, S Kiviluoto, M Waltari, Marjut Puputti, John Patrick Mpindi, Pekka Kohonen, Olli Tynninen, H Haapasalo, Heikki Joensuu, Merja Perälä, et al. Hes6 gene is

selectively overexpressed in glioma and represents an important transcriptional regulator of glioma proliferation. *Oncogene*, 31(10):1299–1310, 2012.

[26] Aric Hagberg, Pieter J Swart, and Daniel A Schult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2008.

[27] Lillianne G Harris, Lewis K Pannell, Seema Singh, Rajeev S Samant, and Lalita A Shevde. Increased vascularity and spontaneous metastasis of breast cancer by hedgehog signaling mediated upregulation of cyr61. *Oncogene*, 31(28):3370–3380, 2012.

[28] Hongwei He, Yong Yi, Xiaoyan Cai, Jiaxing Wang, Xiaochun Ni, Yipeng Fu, and Shuangjian Qiu. Down-regulation of eomes drives t-cell exhaustion via abolishing eomes-mediated repression of inhibitory receptors of t cells in liver cancer. *Journal of Cellular and Molecular Medicine*, 25(1):161–169, 2021.

[29] Ganapati V Hegde, Katie J Peterson, Katy Emanuel, Amit K Mittal, Avadhut D Joshi, John D Dickinson, Gayathri J Kollessery, Robert G Bociek, Philip Bierman, Julie M Vose, et al. Hedgehog-induced survival of b-cell chronic lymphocytic leukemia cells in a stromal cell microenvironment: a potential new therapeutic target. *Molecular Cancer Research*, 6(12):1928–1936, 2008.

[30] Jui-Hung Hung, Troy W Whitfield, Tun-Hsiang Yang, Zhenjun Hu, Zhiping Weng, and Charles DeLisi. Identification of functional modules that correlate with phenotypic difference: the influence of network topology. *Genome biology*, 11:1–16, 2010.

[31] Giovanni Iacono, Ramon Massoni-Badosa, and Holger Heyn. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biology*, 20(1):110, Jun 2019.

[32] Dewei Jiang, Ting Qiu, Junjiang Peng, Siyuan Li, Tala, Wenlong Ren, Chuanyu Yang, Yi Wen, Chuan-Huizi Chen, Jian Sun, et al. Yb-1 is a positive regulator of klf5 transcription factor in basal-like breast cancer. *Cell Death & Differentiation*, 29(6):1283–1295, 2022.

[33] Kenji Kamimoto, Blerta Stringa, Christy M Hoffmann, Kunal Jindal, Lilianna Solnica-Krezel, and Samantha A Morris. Dissecting cell identity via network inference and in silico gene perturbation. *Nature*, 614(7949):742–751, 2023.

[34] Khushbukhat Khan, Sadia Safi, Asma Abbas, Yasmin Badshah, Erum Dilshad, Mehak Rafiq, Kainat Zahra, and Maria Shabbir. Unravelling structure, localization, and genetic crosstalk of klf3 in human breast cancer. *BioMed research international*, 2020(1):1354381, 2020.

[35] Gi-Cheon Kim, Choong-Gu Lee, Ravi Verma, Dipayan Rudra, Taemook Kim, Keunsoo Kang, Jong Hee Nam, Young Kim, Sin-Hyeog Im, and Ho-Keun Kwon. Ets1 suppresses tumorigenesis of human breast cancer via trans-activation of canonical tumor suppressor genes. *Frontiers in oncology*, 10:642, 2020.

[36] Wookbong Kwon, Seong-Kyoon Choi, Daehwan Kim, Hyeon-Gyeom Kim, Jin-Kyu Park, Jee Eun Han, Gil-Jae Cho, Sungho Yun, Wookyung Yu, Se-Hyeon Han, et al. Znf507 affects tgf-β signaling via tgfbr1 and map3k8 activation in the progression of prostate cancer to an aggressive state. *Journal of Experimental & Clinical Cancer Research*, 40:1–17, 2021.

[37] Fashun Liu, Songlin Wan, Yue Li, Zhenxiong Ye, Daojiang Li, and Zhen Li. Pan-cancer analysis confirms the prognostic and immunological effects of prostate tumor overexpressed-1 in human cancers. *Current Cancer Drug Targets*, 24(1):28–45, 2024.

[38] Armando M Marrufo, Stephen O Mathew, Pankaj Chaudhary, Joseph D Malaer, Jamboor K Vishwanatha, and Porunelloor A Mathew. Blocking llt1 (clec2d, ocil)-nkrp1a (cd161) interaction enhances natural killer cell-mediated lysis of triple-negative breast cancer cells. *American journal of cancer research*, 8(6):1050, 2018.

[39] Nicholas McGranahan and Charles Swanton. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer cell*, 27(1):15–26, 2015.

[40] Thomas Moerman, Sara Aibar Santos, Carmen Bravo González-Blas, Jaak Simm, Yves Moreau, Jan Aerts, and Stein Aerts. Grnboost2 and arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics*, 35(12):2159–2161, 2019.

[41] Xin Nie, Dawo Liu, Mingjun Zheng, Xiao Li, Ouxuan Liu, Qian Guo, Liancheng Zhu, and Bei Lin. Herpud1 promotes ovarian cancer cell survival by sustaining autophagy and inhibit apoptosis via pi3k/akt/mtor and p38 mapk signaling pathways. *BMC cancer*, 22(1):1338, 2022.

[42] Peter C Nowell. The clonal evolution of tumor cell populations: Acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. *Science*, 194(4260):23–28, 1976.

[43] Kishan Kumar Nyati, Shigeru Hashimoto, Shailendra Kumar Singh, Murat Tekguc, Hozaifa Metwally, Yu-Chen Liu, Daisuke Okuzaki, Yohannes Gemechu, Sujin Kang, and Tadamitsu Kishimoto. The novel long noncoding rna au021063, induced by il-6/arid5a signaling, exacerbates breast cancer invasion and metastasis by stabilizing trib3 and activating the mek/erk pathway. *Cancer Letters*, 520:295–306, 2021.

[44] Kazutaka Otsuji, Yoko Takahashi, Tomo Osako, Takayuki Kobayashi, Toshimi Takano, Sumito Saeki, Liying Yang, Satoko Baba, Kohei Kumegawa, Hiromu Suzuki, et al. Serial single-cell rna sequencing unveils drug resistance and metastatic traits in stage iv breast cancer. *NPJ Precision Oncology*, 8(1):222, 2024.

[45] Youngjune Park, Sangsoo Lim, Jin-Wu Nam, and Sun Kim. Measuring intratumor heterogeneity by network entropy using rna-seq data. *Scientific reports*, 6(1):37767, 2016.

[46] Junying Qin, Zhongmei Zhou, Wenlin Chen, Chunyan Wang, Hailin Zhang, Guangzhe Ge, Ming Shao, Dingyun You, Zhixiang Fan, Houjun Xia, et al. Bap1 promotes breast cancer cell proliferation and metastasis by deubiquitinating klf5. *Nature communications*, 6(1):8471, 2015.

[47] JA Ribeil, Y Zermati, J Vandekerckhove, M Dussiot, J Kersual, and O Hermine. L'érythropoïèse: un paradigme pour l'étude du rôle des caspases dans la mort et la différenciation cellulaire. *Journal de la Société de Biologie*, 199(3):219–231, 2005.

[48] Madhuchhanda Roy, Darya G Buehler, Ranran Zhang, Michael L Schwalbe, Rebecca M Baus, M Shahriar Salamat, Ricardo V Lloyd, and Jason N Rosenbaum. Expression of insulinoma-associated protein 1 (insm1) and orthopedia homeobox (otp) in tumors with neuroendocrine differentiation at rare sites. *Endocrine Pathology*, 30:35–42, 2019.

[49] Gryte Satas, Simone Zaccaria, Geoffrey Mon, and Benjamin J Raphael. Scarlet: single-cell tumor phylogeny inference with copy-number constrained mutation losses. *Cell systems*, 10(4):323–332, 2020.

[50] Deborah Scheffer, Cyrille Sage, David P Corey, and Veronique Pingault. Gene expression profiling identifies hes6 as a transcriptional target of atoh1 in cochlear hair cells. *FEBS letters*, 581(24):4651–4656, 2007.

[51] Jochen Singer, Jack Kuipers, Katharina Jahn, and Niko Beerenwinkel. Single-cell mutation identification via phylogenetic inference. *Nature communications*, 9(1):5144, 2018.

[52] Claudia Skok Gibbs, Christopher A Jackson, Giuseppe-Antonio Saldi, Andreas Tjärnberg, Aashna Shah, Aaron Watters, Nicholas De Veaux, Konstantine Tchourine, Ren Yi, Tymor Hamamsy, et al. High-performance single-cell gene regulatory network inference at scale: the inferelator 3.0. *Bioinformatics*, 38(9):2519–2528, 2022.

[53] Edward T Stuart, Rebecca Haffner, Moshe Oren, and Peter Gruss. Loss of p53 function through pax-mediated transcriptional repression. *The EMBO journal*, 14(22):5638–5645, 1995.

[54] Edwin Wang, Naif Zaman, Shauna Mcgee, Jean-Sébastien Milanese, Ali Masoudi-Nejad, and Maureen O'Connor-McCourt. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. In *Seminars in cancer biology*, volume 30, pages 4–12. Elsevier, 2015.

[55] Elizabeth A Waterman, Neil A Cross, Jenifer M Lippitt, Simon S Cross, Ishtiaq Rehman, Ingunn Holen, Freddie C Hamdy, and Colby L Eaton. The antibody mab8051 directed against osteoprotegerin detects carbonic anhydrase ii: Implications for association studies with human cancers. *International journal of cancer*, 121(9):1958–1966, 2007.

[56] Qing Wu, Xuan Tao, Yang Luo, Shiyao Zheng, Nan Lin, and Xianhe Xie. A novel super-enhancer-related gene signature predicts prognosis and immune microenvironment for breast cancer. *BMC cancer*, 23(1):776, 2023.

[57] Si-An Xie, Wen Zhang, Feng Du, Si Liu, Ting-Ting Ning, Nan Zhang, Shu-Tian Zhang, and Sheng-Tao Zhu. Ptov1 facilitates colorectal cancer cell proliferation through activating akt1 signaling pathway. *Heliyon*, 10(16), 2024.

[58] Lu Xu, Lintao Wang, Chaojun Jiang, Qiannan Zhu, Rui Chen, Jue Wang, and Shui Wang. Biological effect of ribosomal protein l32 on human breast cancer cell behavior. *Molecular medicine reports*, 22(3):2478–2486, 2020.

[59] Fan Yang, Ling Lin, Xiaohua Li, Ronglan Wen, and Xin Zhang. Silencing of col3a1 represses proliferation, migration, invasion, and immune escape of triple negative breast cancer cells via down-regulating pd-l1 expression. *Cell Biology International*, 46(11):1959–1969, 2022.

[60] Guorong Yao, Kaiyue Zhao, Kaikai Bao, and Jing Li. Radiation increases col1a1, col3a1, and col1a2 expression in breast cancer. *Open Medicine*, 17(1):329–340, 2022.

[61] Hamim Zafar, Nicholas Navin, Ken Chen, and Luay Nakhleh. Siclonefit: Bayesian inference of population structure, genotype, and phylogeny of tumor clones from single-cell genome sequencing data. *Genome research*, 29(11):1847–1859, 2019.

[62] Hamim Zafar, Anthony Tzen, Nicholas Navin, Ken Chen, and Luay Nakhleh. Sifit: inferring tumor trees from single-cell sequencing data under finite-sites models. *Genome biology*, 18:1–20, 2017.

[63] Hamim Zafar, Yong Wang, Luay Nakhleh, Nicholas Navin, and Ken Chen. Monovar: single-nucleotide variant detection in single cells. *Nature methods*, 13(6):505–507, 2016.

[64] Di Zhang, Sadahiro Iwabuchi, Tomohisa Baba, Shin-ichi Hashimoto, Naofumi Mukaida, and So-ichiro Sasaki. Involvement of a transcription factor, nfe2, in breast cancer metastasis to bone. *Cancers*, 12(10):3003, 2020.

[65] Long Zhao, Bo Wang, Changjiang Yang, Yilin Lin, Zhen Zhang, Shan Wang, Yingjiang Ye, and Zhanlong Shen. Tdo2 knockdown inhibits colorectal cancer progression via tdo2–kynu–ahr pathway. *Gene*, 792:145736, 2021.

[66] Joseph X Zhou, Roberto Taramelli, Edoardo Pedrini, Theo Knijnenburg, and Sui Huang. Extracting intercellular signaling network of cancer tissues using ligand-receptor expression patterns from whole-tumor and single-cell transcriptomes. *Scientific Reports*, 7(1):8815, 2017.

[67] Junjie Zhou, Wenhao Chen, Qianwen He, Dong Chen, Chunguang Li, Congqing Jiang, Zhao Ding, and Qun Qian. Serbp1 affects the apoptotic level by regulating the expression and alternative splicing of cellular and metabolic process genes in hela cells. *PeerJ*, 10:e14084, 2022.

[68] Zhiyan Zhu, Hao Wang, Yiliang Wei, Fanrong Meng, Zhe Liu, and Zhenfa Zhang. Downregulation of prdm1 promotes cellular invasion and lung cancer metastasis. *Tumor Biology*, 39(4):1010428317695929, 2017.

A Supplementary figures

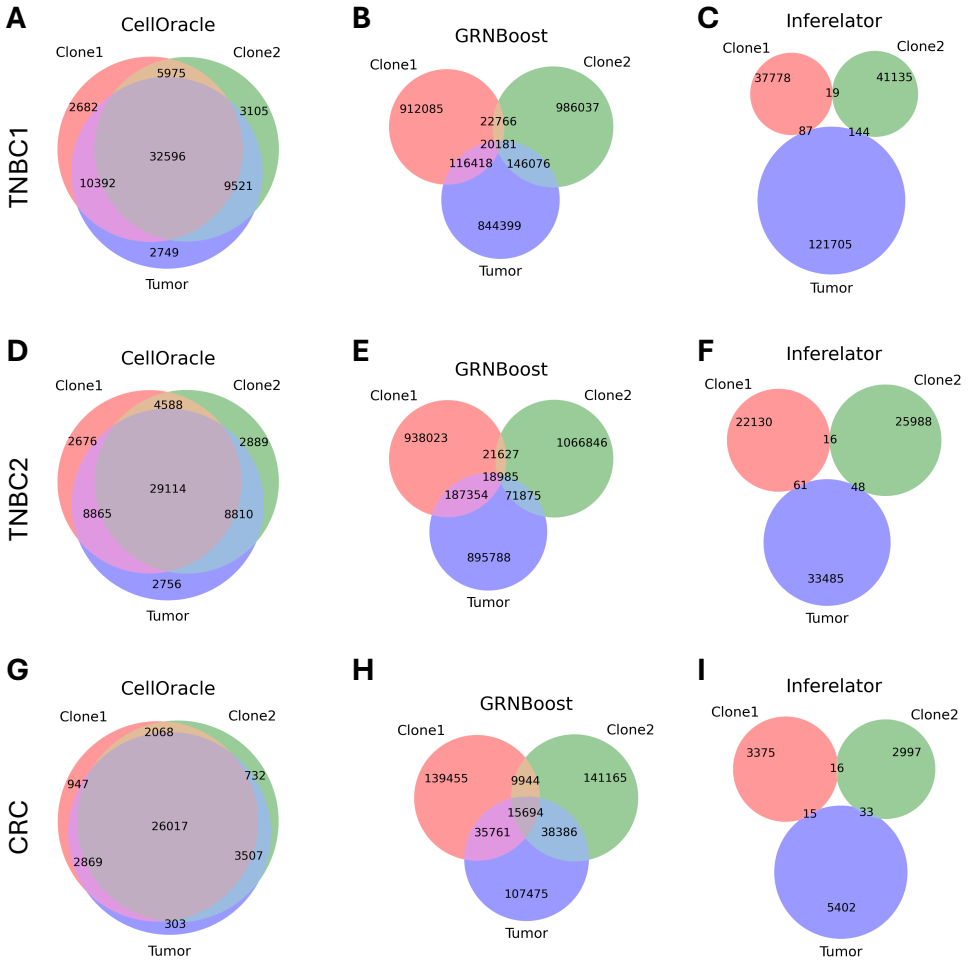

Figure 9: Venn diagrams showing unique and shared edges between GRNs inferred from tumor clones 1 and 2, and all tumor cells jointly. Columns correspond to three GRN inference tools: CellOracle **(A, D, G)**, GRNboost2 **(B, E, H)**, and Inferelator **(C, F, I)**. Rows correspond to different samples: TNBC1 **(A-C)**, TNBC2 **(D-F)**, and CRC **(G-I)**

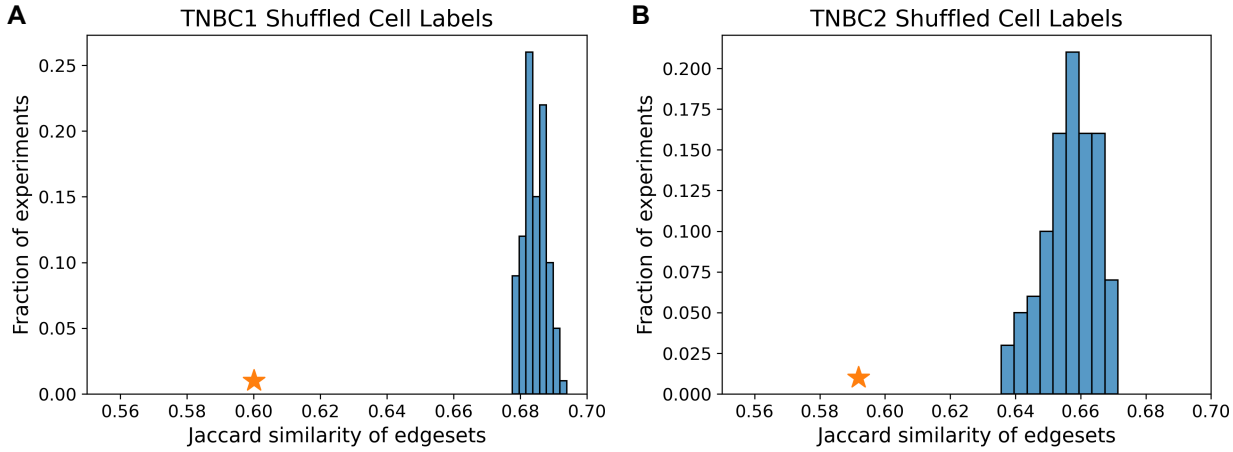

Figure 10: Distribution of the Jaccard similarities between edge sets of two tumor subsets obtained by shuffling the cells. Orange stars represent the Jaccard similarity between the edge sets of the clone-specific GRNs. **(A)** Data for TNBC1 sample. **(B)** Data for TNBC2 sample.

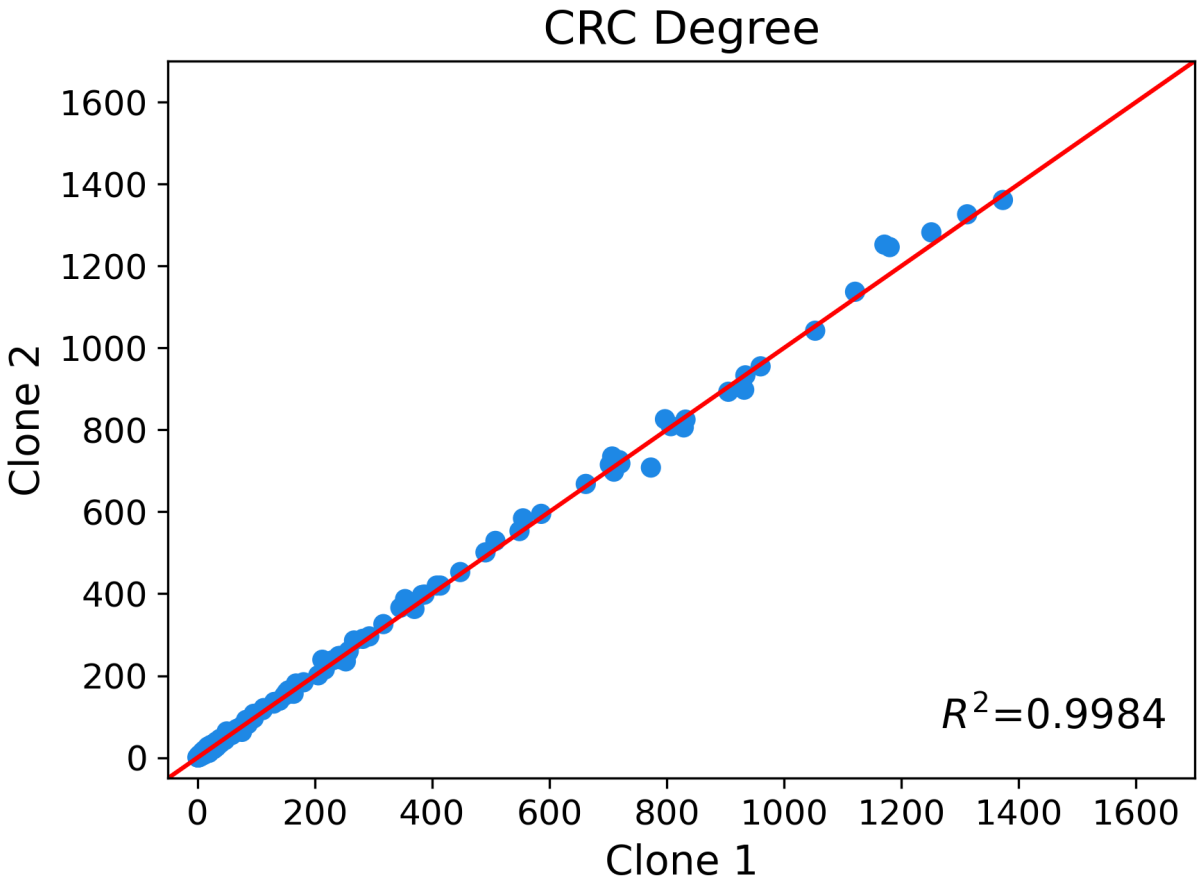

Figure 11: Vertex degrees based on clone 1 (x-axis) and clone 2 (y-axis) GRNs inferred from the CRC sample. Coefficient of determination (R^2) for $\hat{y} = x$ is reported.

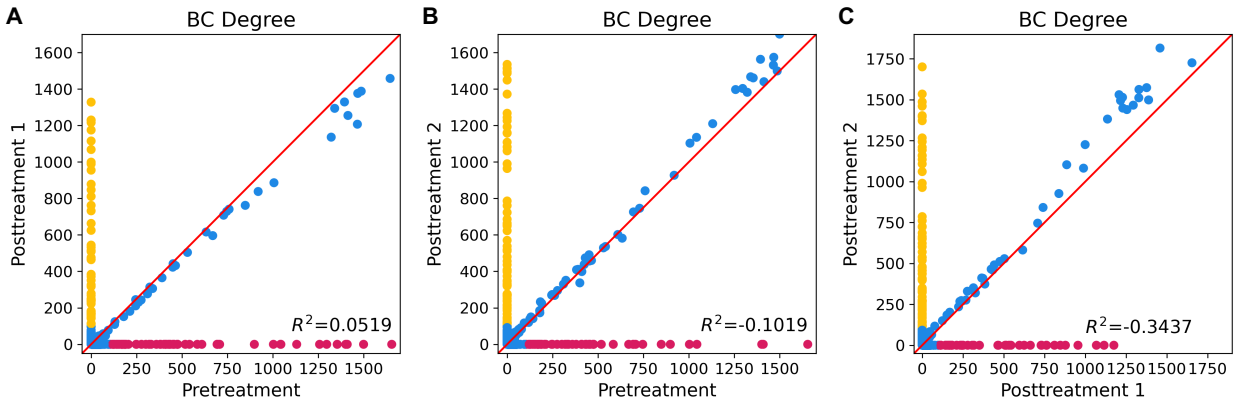

Figure 12: **(A)** Vertex degrees based on Pre (x-axis) and Post1 (y-axis) GRNs inferred from the BC dataset. Coefficient of determination (R^2) for $\hat{y} = x$ is reported. Outliers that have degree 0 in one of the GRNs and degree > 100 in another one are highlighted with color. **(B)** Vertex degrees based on Pre (x-axis) and Post2 (y-axis) GRNs inferred from the BC dataset. Coefficient of determination (R^2) for $\hat{y} = x$ is reported. Outliers that have degree 0 in one of the GRNs and degree > 100 in another one are highlighted with color. **(C)** Vertex degrees based on Post1 (x-axis) and Post2 (y-axis) GRNs inferred from the BC dataset. Coefficient of determination (R^2) for $\hat{y} = x$ is reported. Outliers that have degree 0 in one of the GRNs and degree > 100 in another one are highlighted with color.