

Supplementary Material: Intratumor Heterogeneity Through the Lens of Gene Regulatory Networks

Narein Rao¹, Nicolae Sapoval¹, Luay Nakhleh^{1,2,*} and Hamim Zafar^{3,4,5,*}

¹Department of Computer Science, Rice University, , 77005, TX, USA

²Department of BioSciences, Rice University, , 77005, TX, USA

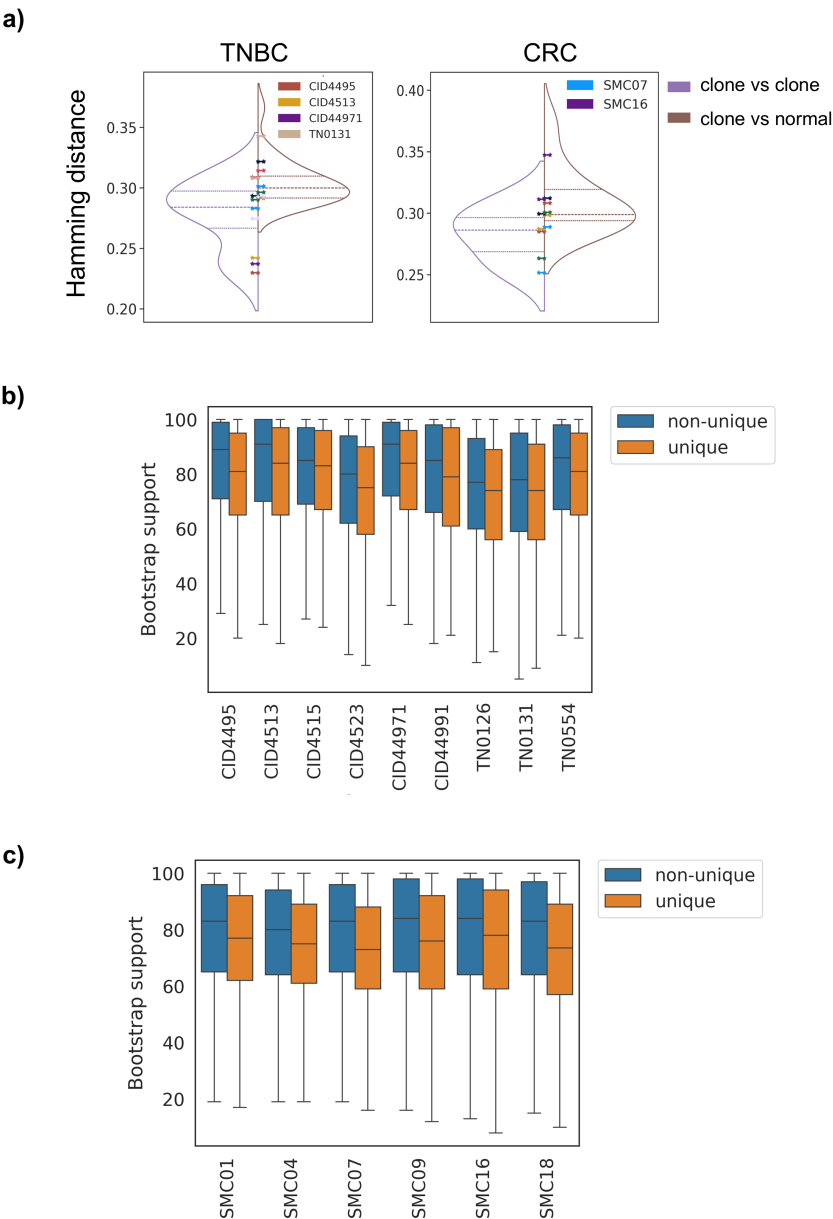
³Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, , , India

⁴Department of Biological Sciences and Bioengineering, Indian Institute of Technology Kanpur, , , India

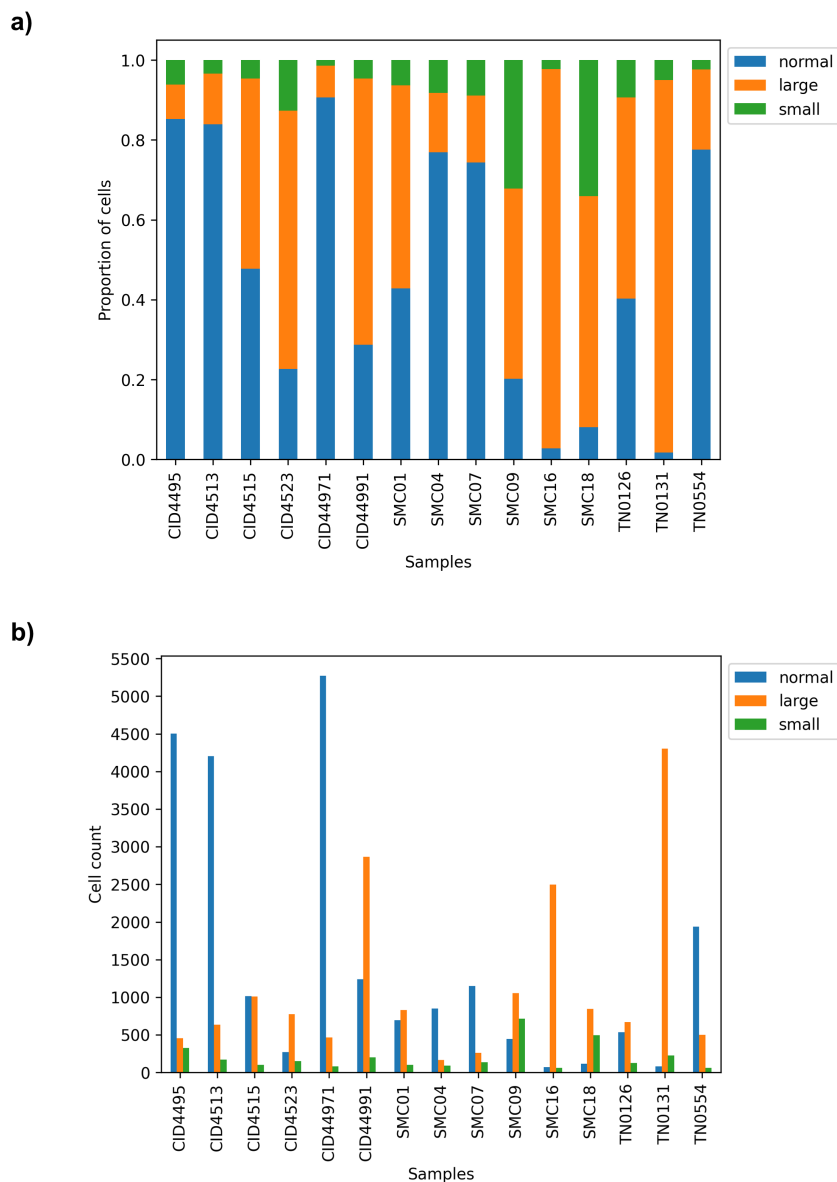
⁵Mehta Family Center for Engineering in Medicine, Indian Institute of Technology Kanpur, , , India

*Corresponding authors. nakhleh@rice.edu, hamim@iitk.ac.in

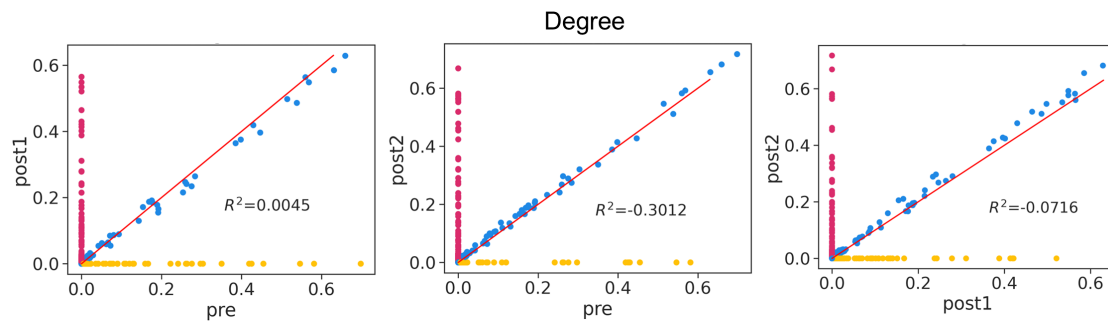
Supplementary Figures



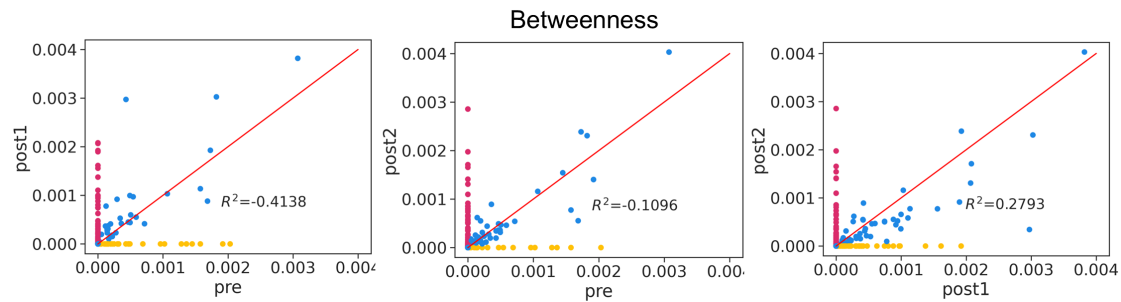
Supplementary Figure S1 (a) Distribution of Hamming distances using binarized edge-presence vectors between clones (purple) and clone and normal subpopulations (brown). Each color peg within the violin plot depicts a specific sample. Mann-Whitney U test set to hypothesis with p-values of 0.0039 and 0.083 for TNBC and CRC, respectively.



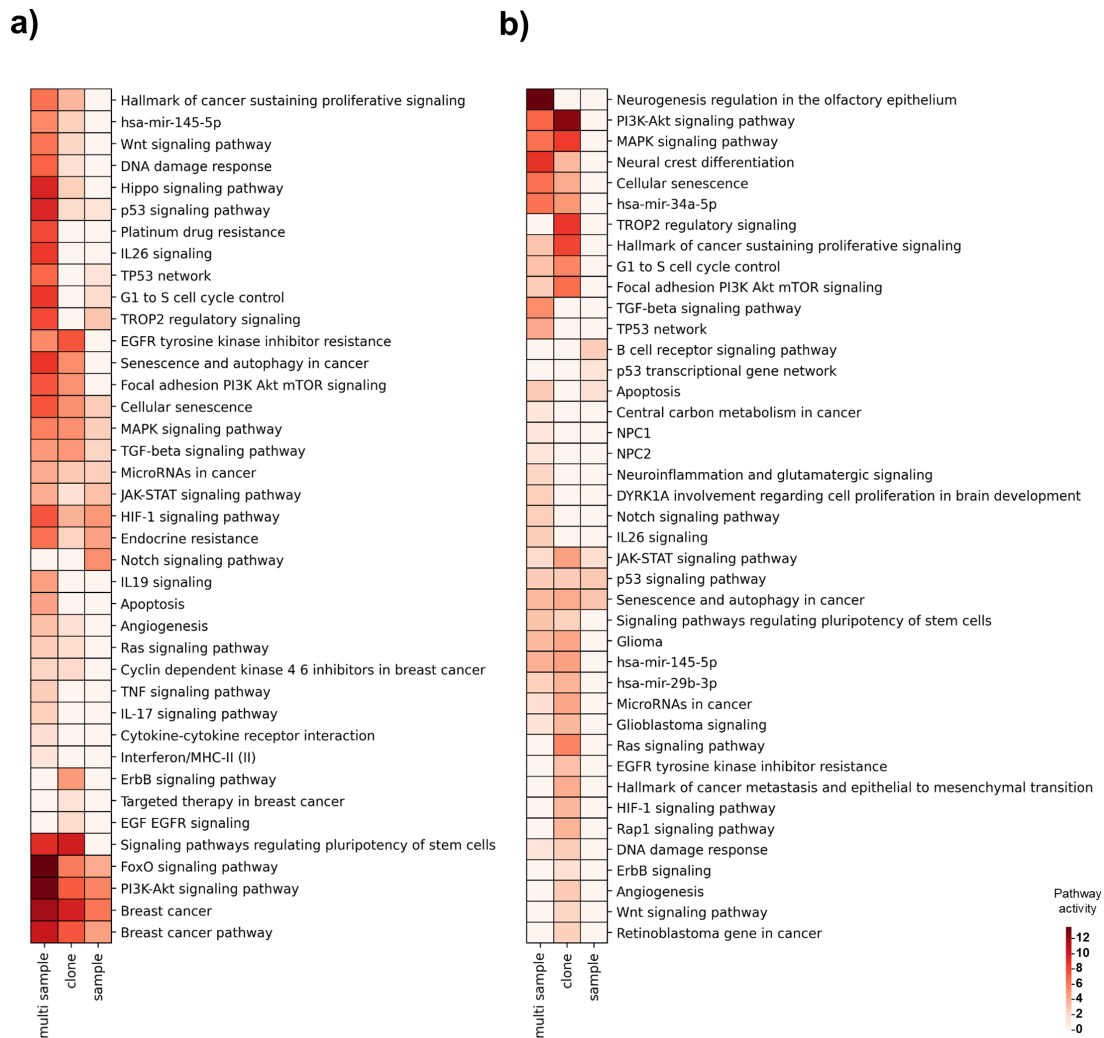
Supplementary Figure S2 (a) Proportion of cells across normal cells, largest and smallest clones in each sample. (b) Absolute counts of cells across normal cells, largest and smallest clones in each sample.



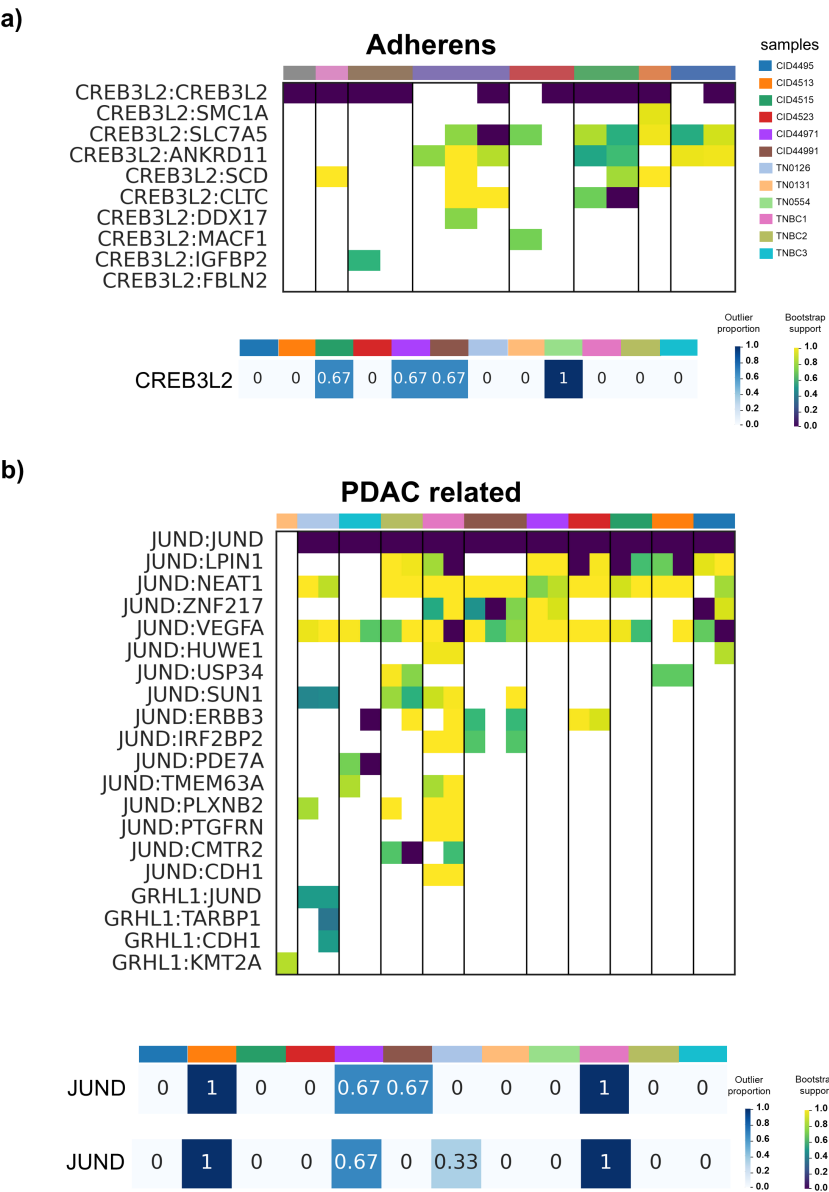
Supplementary Figure S3 Vertex degrees based on ($timepoint_1$, $timepoint_2$) (x- and y-axis, respectively) GRNs inferred from the BC dataset. Coefficient of determination (R^2) for $\hat{y} = x$ is reported. Outliers that have degree 0 in one of the GRNs and degree > 100 in another one are highlighted with color. From left to right: (pre,post1); (pre,post2); (post1,post2).



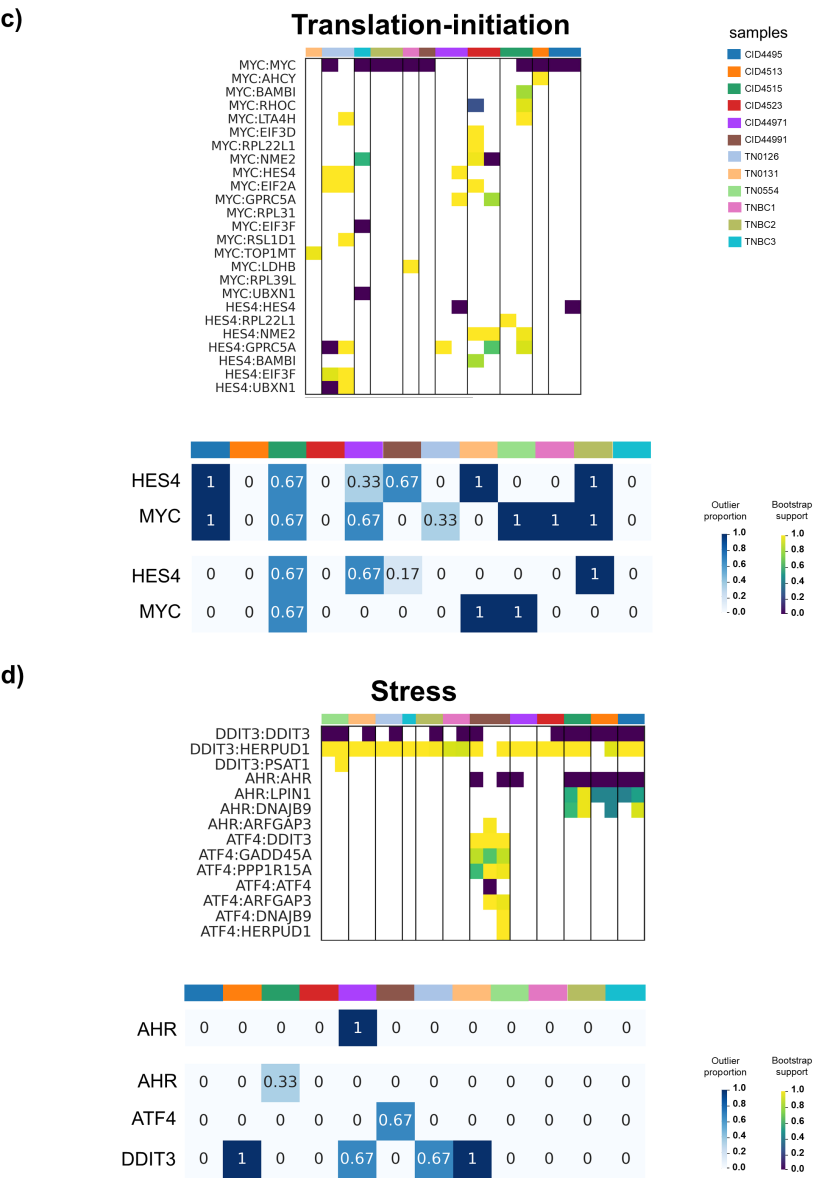
Supplementary Figure S4 Vertex betweenness based on ($timepoint_1$, $timepoint_2$) (x- and y-axis, respectively) GRNs inferred from the BC dataset. Coefficient of determination (R^2) for $\hat{y} = x$ is reported. Outliers that have degree 0 in one of the GRNs and degree > 100 in another one are highlighted with color. From left to right: (pre,post1); (pre,post2); (post1,post2).



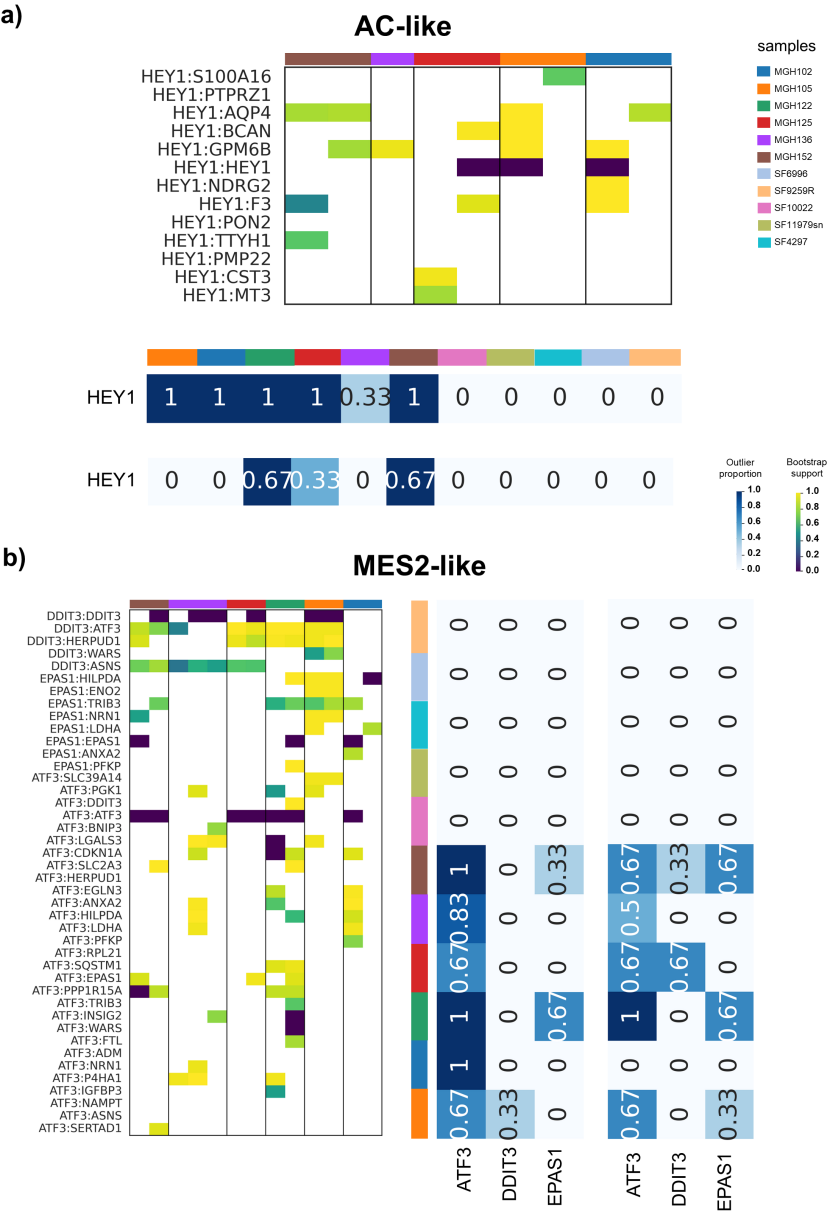
Supplementary Figure S5 (a) Pathways and corresponding pathway enrichment scores specific to each of the three categories in the TNBC analysis. Genesets used for pathway enrichment include the TFs and TGs of every regulon within a specific category. (b) Pathways and corresponding pathway enrichment scores specific to each of the three categories in the Glioblastoma analysis. Genesets used for pathway enrichment include the TFs and TGs of every regulon within a specific category.



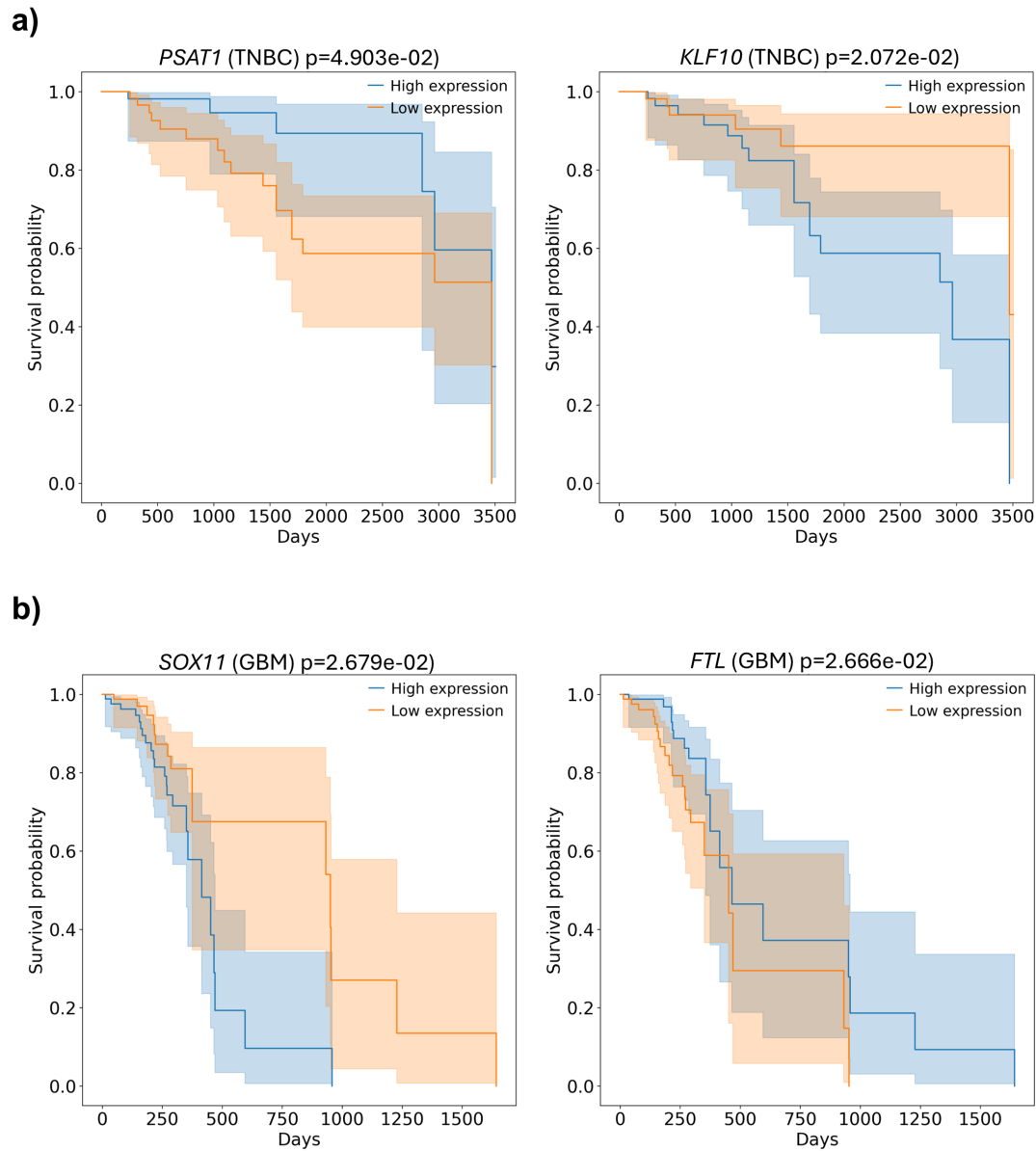
Supplementary Figure S6



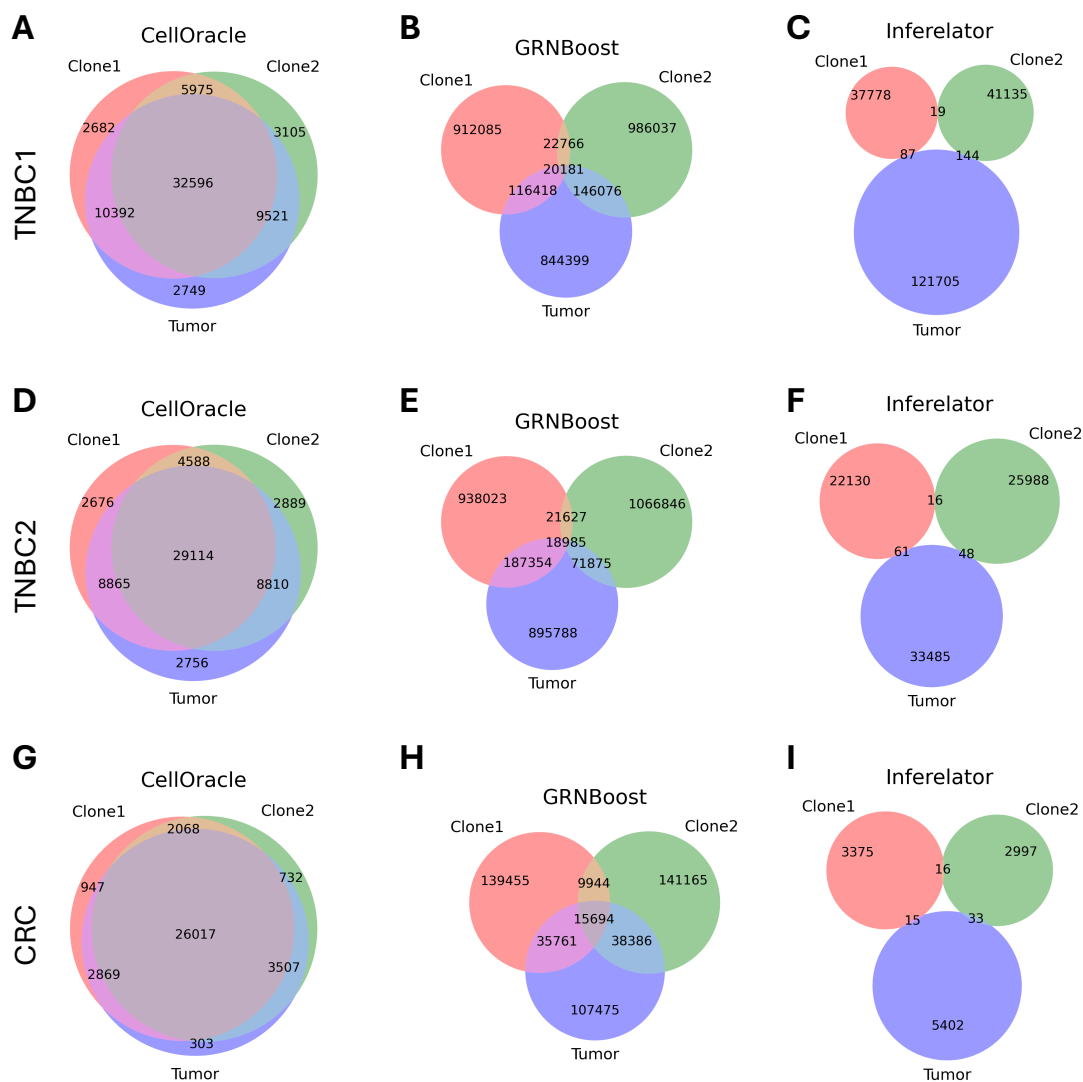
Supplementary Figure S7 (top) Distribution of SCENIC edges across all TNBC samples for meta-programs (a) Adherens (b) PDAC (c) Translation (d) Stress. Each cell represents the corresponding bootstrap value. Sections within each row (sample) represents clones within each sample. (bottom) Distribution of betweenness outliers, across all TNBC samples for meta-programs (a) Adherens (b) PDAC (c) Translation (d) Stress. Each cell represent the fraction of clone pairings a gene has been found to be an outlier in.



Supplementary Figure S8 (a) (top) Distribution of SCENIC edges across all Glioblastoma samples for meta program AC. Each cell represents the corresponding bootstrap value. Sections within each row (sample) represents clones within each sample. (bottom) Distribution of degree (top) and betweenness (bottom) outliers, across all Glioblastoma samples. Each cell represent the fraction of clone pairings a gene has been found to be an outlier in. (b) (top) Distribution of SCENIC edges across all Glioblastoma samples for meta program MES2. Each cell represents the corresponding bootstrap value. Sections within each row (sample) represents clones within each sample. (bottom) Distribution of degree (top) and betweenness (bottom) outliers, across all Glioblastoma samples. Each cell represent the fraction of clone pairings a gene has been found to be an outlier in.

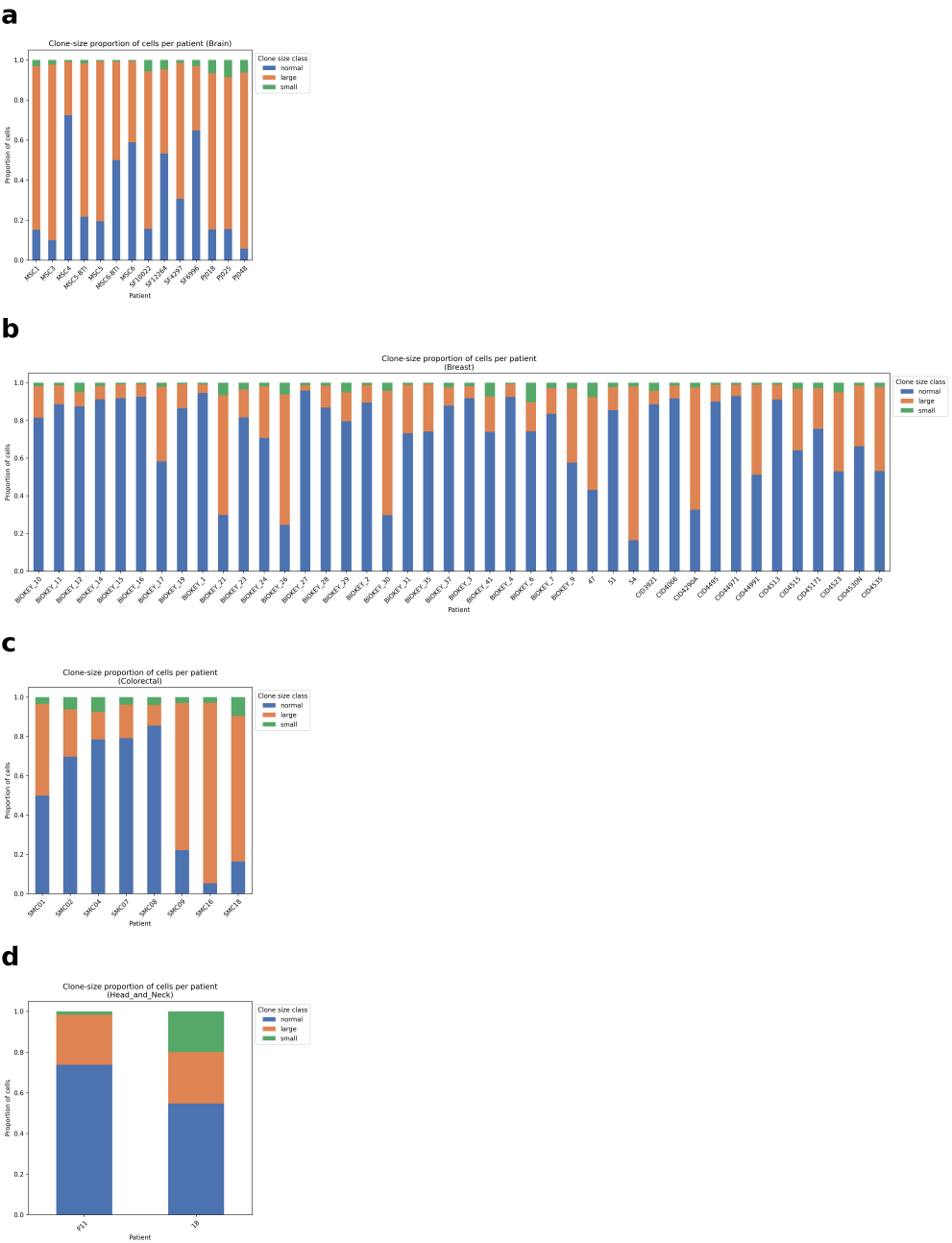


Supplementary Figure S9 (a) Kaplan–Meier survival curves comparing patients stratified by genes *PSAT1* and *KLF10* expression in TNBC patients from TCGA using a median-based threshold across. Patients with expression values greater than or equal to the median were assigned to the high group, and those below the median to the low group. The y-axis represents survival probability, and the x-axis represents time (days). A statistically significant difference in survival between groups was observed (log-rank test, $p < 0.05$). (b) Kaplan–Meier survival curves comparing patients stratified by genes *SOX11* and *FTL* expression in GBM patients from TCGA using a median-based threshold across. Patients with expression values greater than or equal to the median were assigned to the high group, and those below the median to the low group. The y-axis represents survival probability, and the x-axis represents time (days). A statistically significant difference in survival between groups was observed (log-rank test, $p < 0.05$).



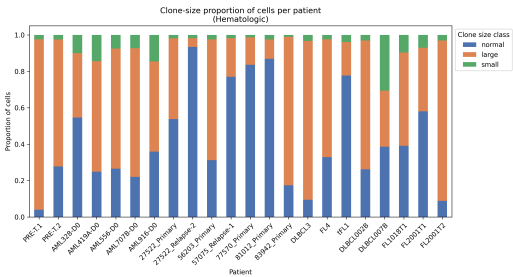
Supplementary Figure S10 Venn diagrams showing unique and shared edges between GRNs inferred from tumor clones 1 and 2, and all tumor cells jointly. Columns correspond to three GRN inference tools: CellOracle (A, D, G), GRNboost2 (B, E, H), and Inferelator (C, F, I). Rows correspond to different samples: TNBC1 (A-C), TNBC2 (D-F), and CRC (G-I).

Supplementary Figures: extended cohort by cancer type

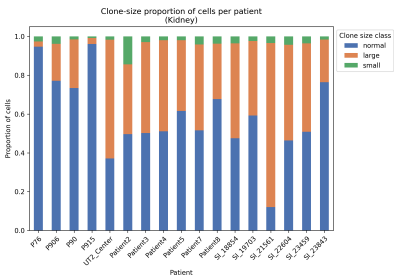


Supplementary Figure S11 Proportion of cells per patient in the normal (blue), large-clone (orange) and small-clone (green) populations, shown separately for each cancer type in the extended cohort: (a) Brain, (b) Breast, (c) Colorectal, (d) Head and neck; each bar is one patient. Part 1 of 3.

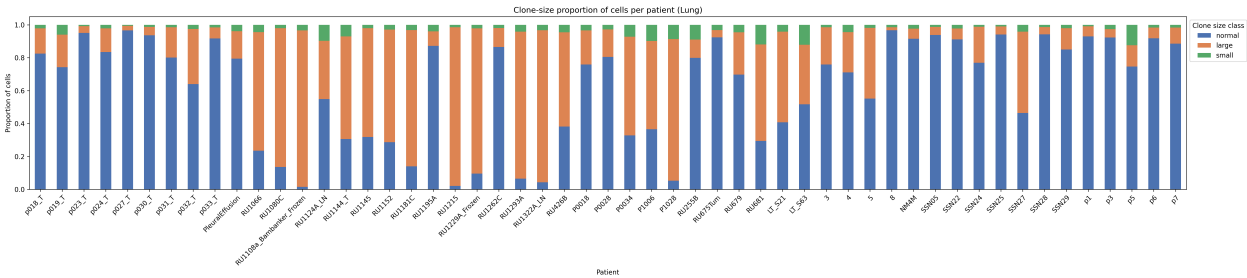
e



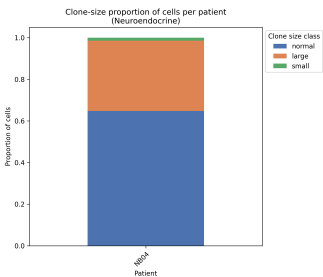
f



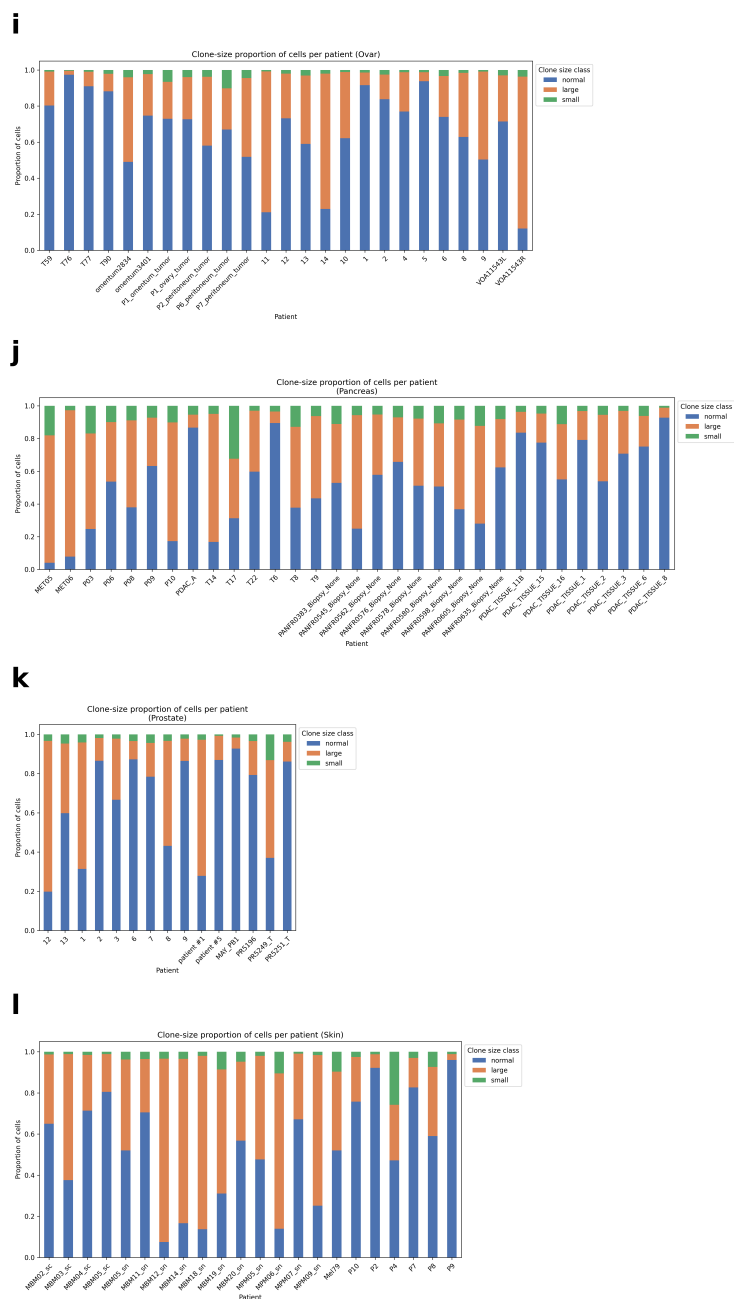
g



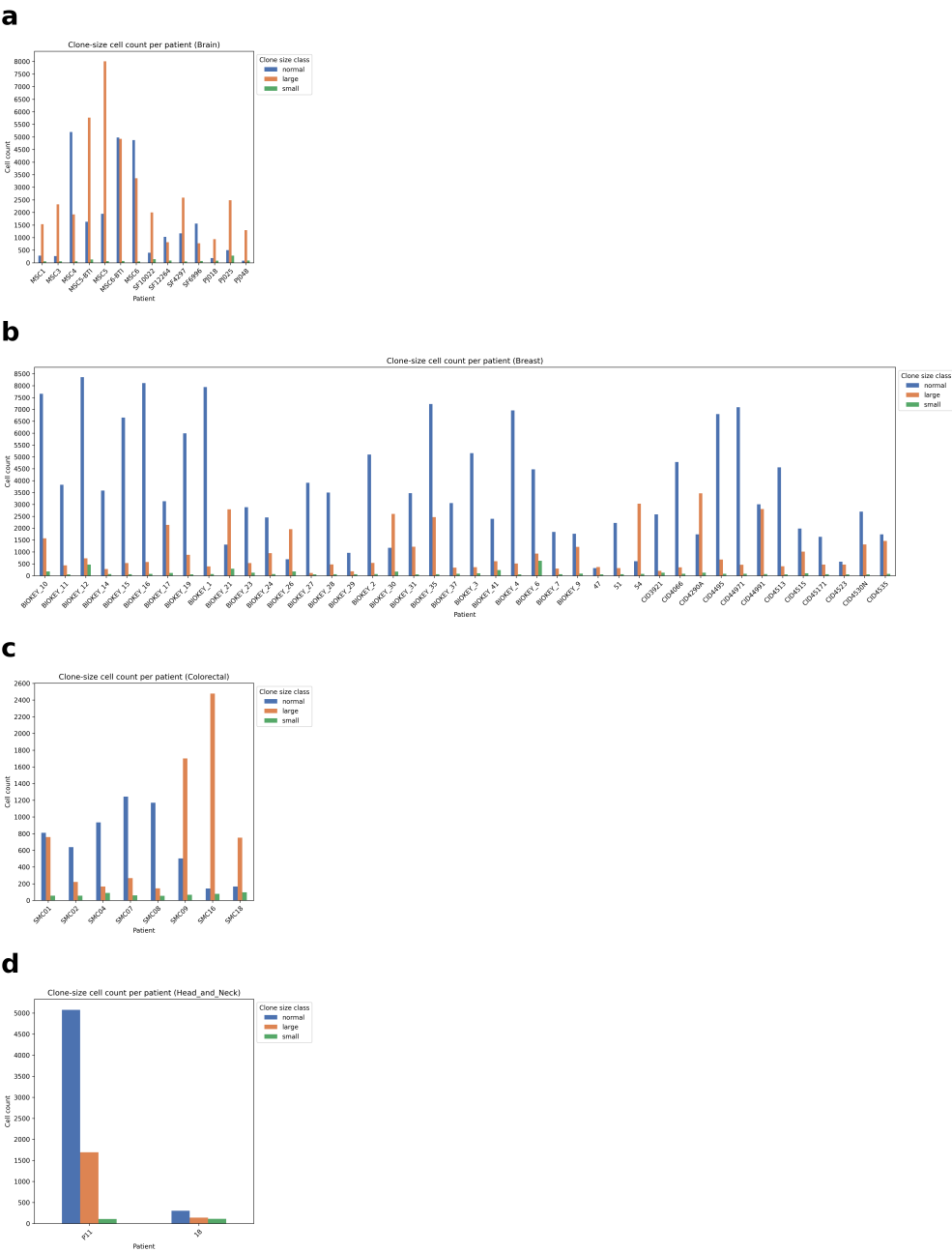
h



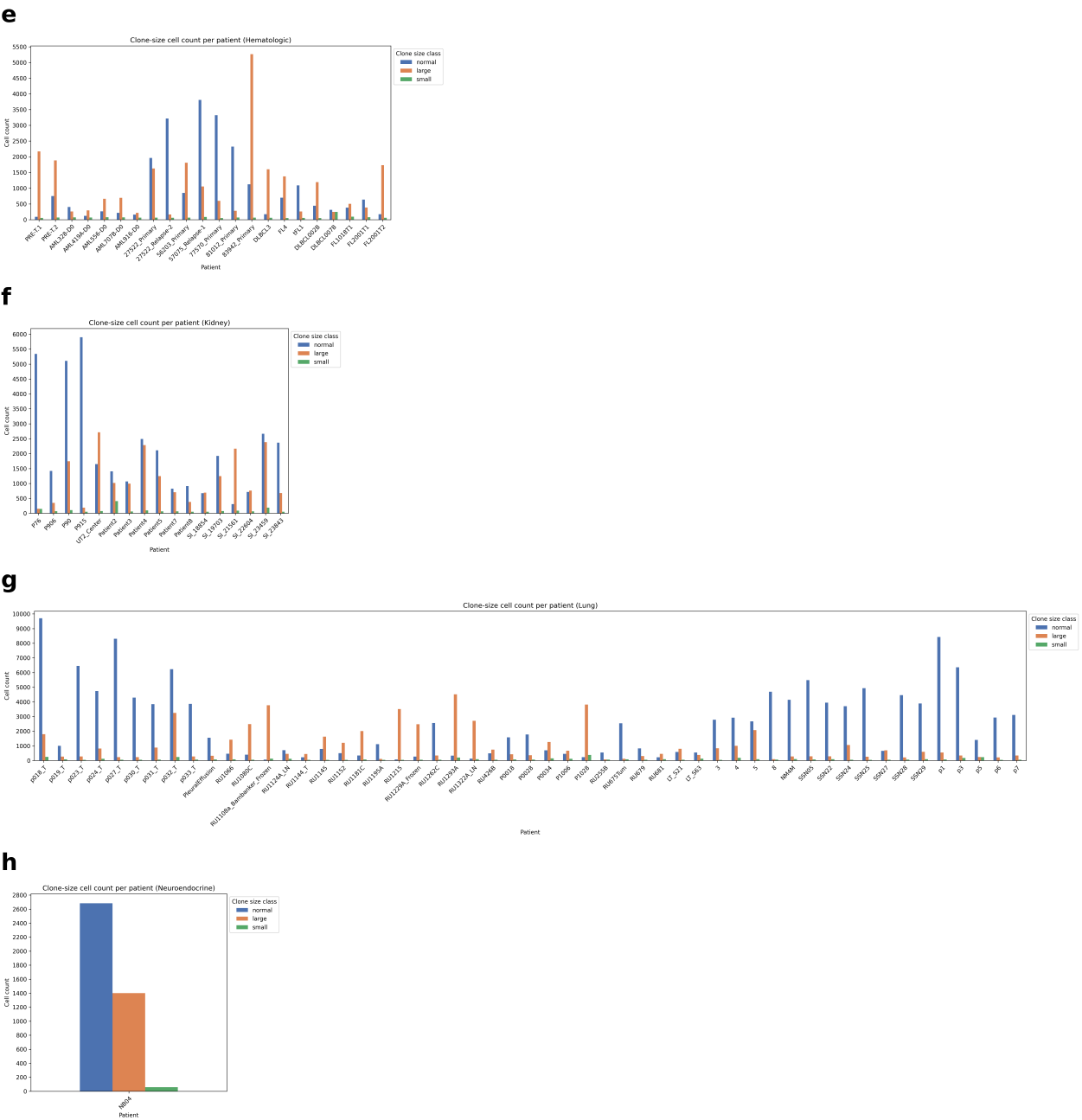
Supplementary Figure S12 Proportion of cells per patient in the normal (blue), large-clone (orange) and small-clone (green) populations, shown separately for each cancer type in the extended cohort: (e) Hematologic, (f) Kidney, (g) Lung, (h) Neuroendocrine; each bar is one patient. Part 2 of 3.



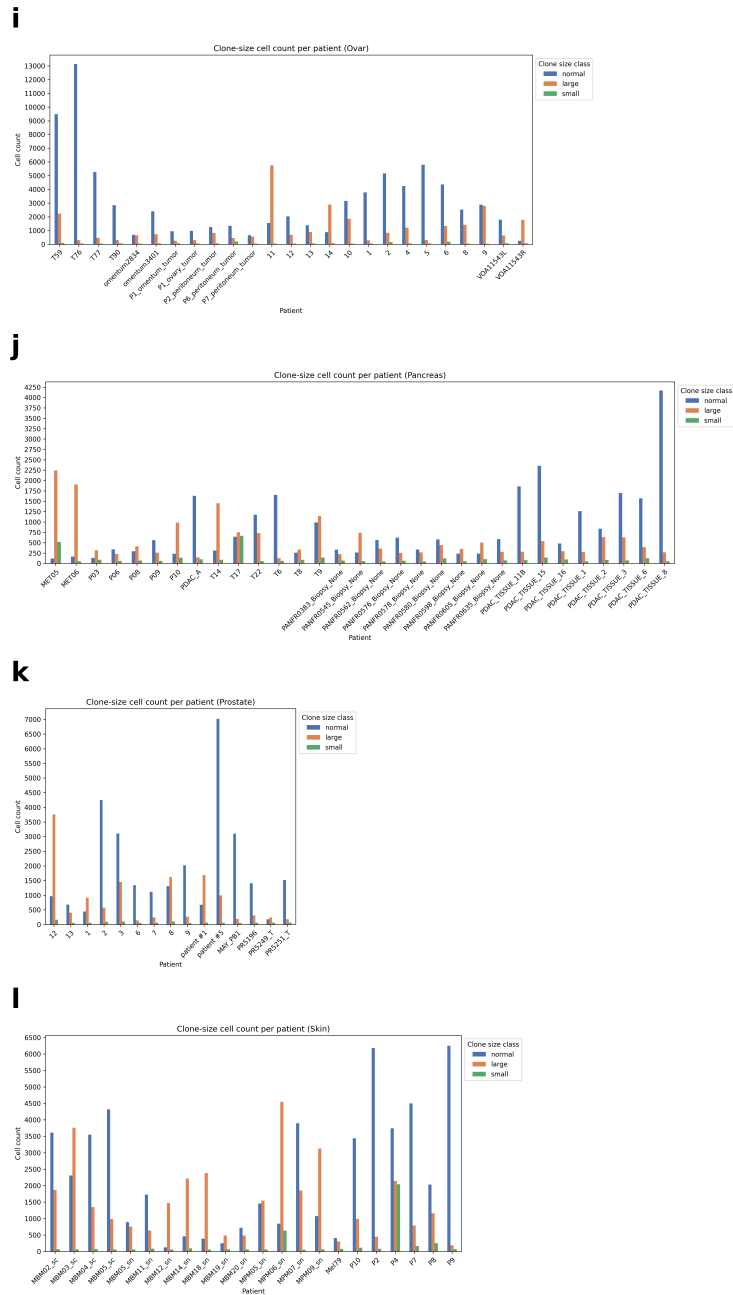
Supplementary Figure S13 Proportion of cells per patient in the normal (blue), large-clone (orange) and small-clone (green) populations, shown separately for each cancer type in the extended cohort: (i) Ovarian, (j) Pancreatic, (k) Prostate, (l) Skin; each bar is one patient. Part 3 of 3.

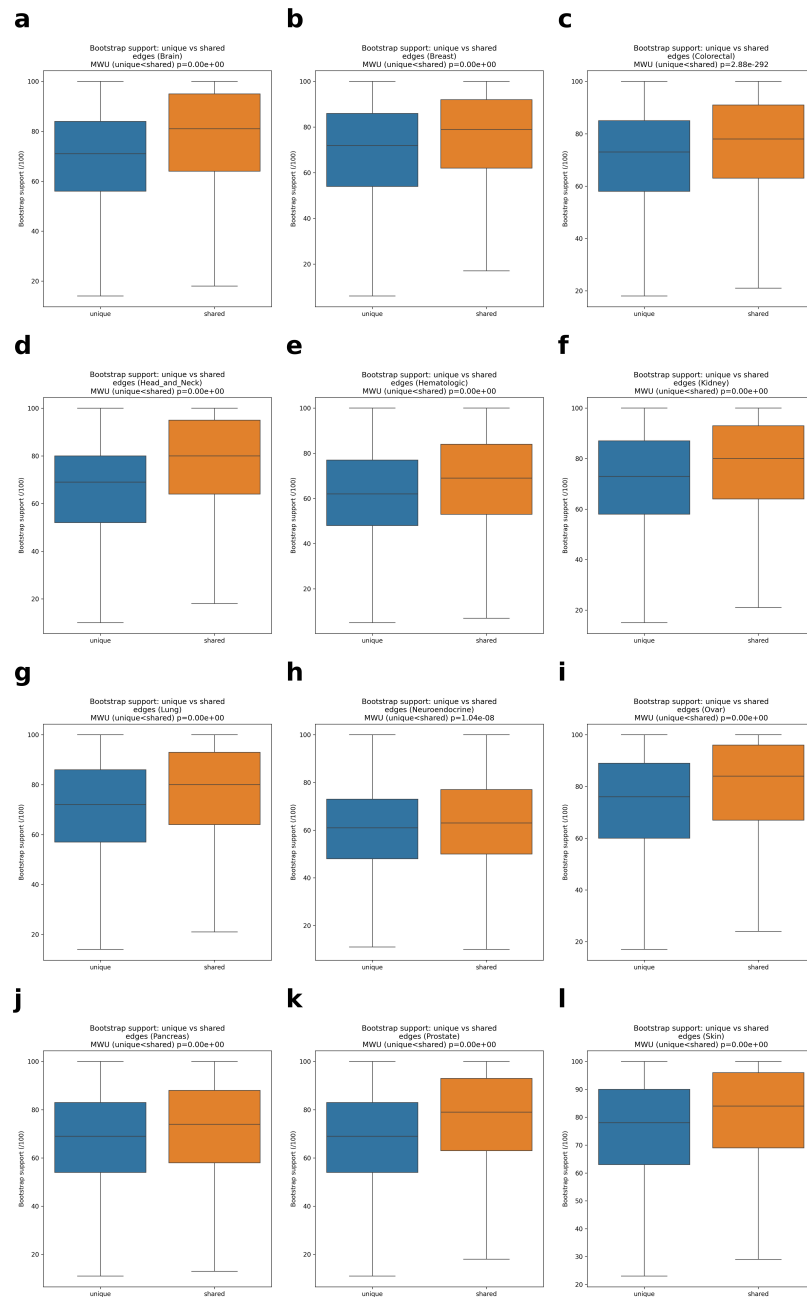


Supplementary Figure S14 Number of cells per patient in the normal, large-clone and small-clone populations, shown separately for each cancer type in the extended cohort: (a) Brain, (b) Breast, (c) Colorectal, (d) Head and neck; each bar is one patient. Part 1 of 3.

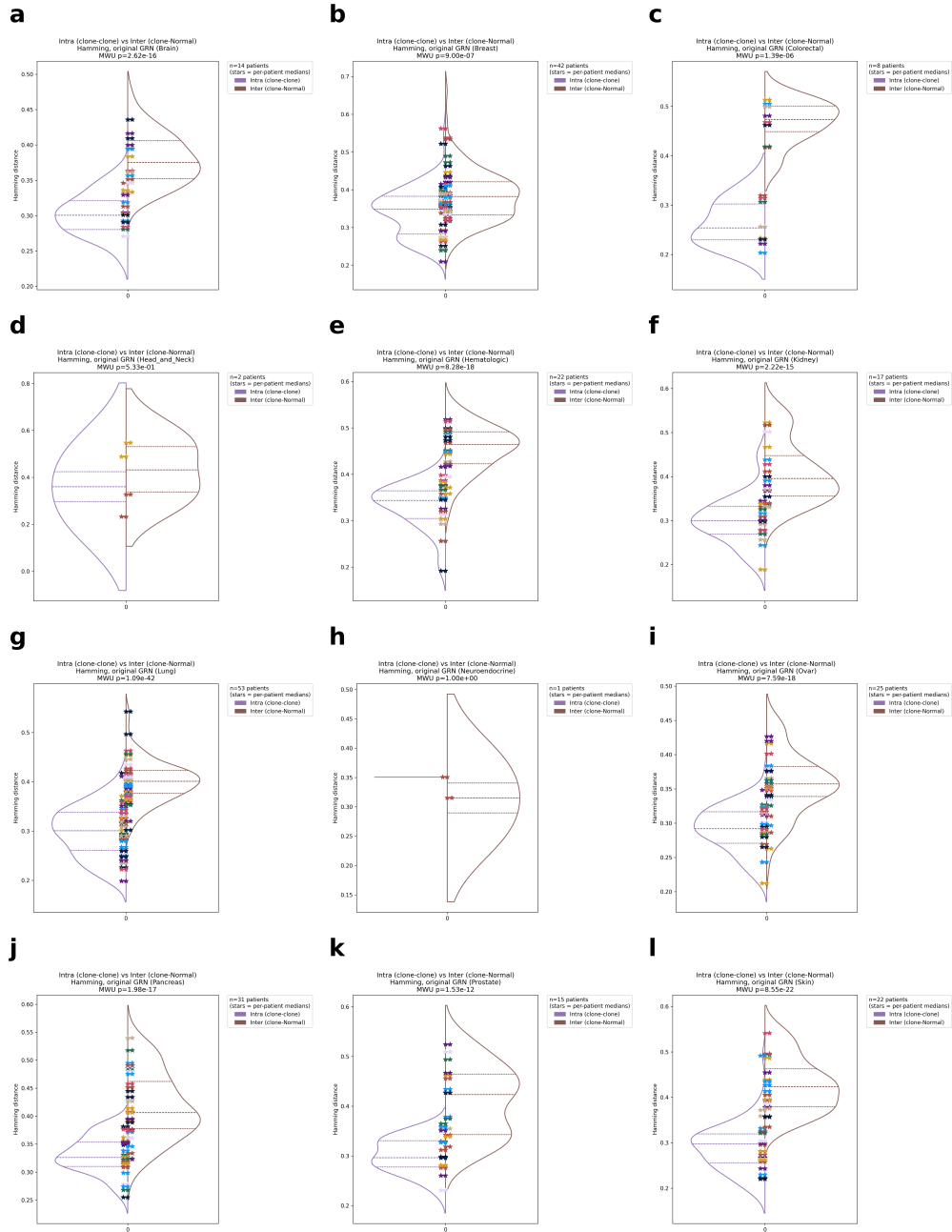


Supplementary Figure S15 Number of cells per patient in the normal, large-clone and small-clone populations, shown separately for each cancer type in the extended cohort: (e) Hematologic, (f) Kidney, (g) Lung, (h) Neuroendocrine; each bar is one patient. Part 2 of 3.





Supplementary Figure S17 Bootstrap support (out of 100) of edges unique to a clone versus edges shared with other populations, pooled across all patients within each cancer type of the extended cohort: (a) Brain, (b) Breast, (c) Colorectal, (d) Head and neck, (e) Hematologic, (f) Kidney, (g) Lung, (h) Neuroendocrine, (i) Ovarian, (j) Pancreatic, (k) Prostate, (l) Skin. Mann–Whitney p -values test unique < shared.



Supplementary Figure S18 Hamming distance between edge-presence vectors of the original GRNs, clone-clone (purple) versus clone-normal (brown), for each cancer type of the extended cohort: (a) Brain, (b) Breast, (c) Colorectal, (d) Head and neck, (e) Hematologic, (f) Kidney, (g) Lung, (h) Neuroendocrine, (i) Ovarian, (j) Pancreatic, (k) Prostate, (l) Skin. Stars mark per-patient medians; dashed and dotted lines mark the median and quartiles of each distribution; Mann-Whitney p -values are shown in each panel title.

Supplementary Tables

Supplementary Table S1 Datasets and patients (samples) of the extended cohort drawn from the Curated Cancer Cell Atlas (3CA); 253 patients in total. Identifiers are those used by 3CA.

Cancer type	Dataset	Patients
Brain	Choudhury et al. 2022 ($n = 7$)	MSC1, MSC3, MSC4, MSC5, MSC5-BTI, MSC6, MSC6-BTI
	Wang et al. 2019 ($n = 4$)	SF10022, SF12264, SF4297, SF6996
	Yuan et al. 2018 ($n = 3$)	PJ018, PJ025, PJ048
Breast	Bassez et al. 2021 ($n = 27$)	BIOKEY_1, BIOKEY_10, BIOKEY_11, BIOKEY_12, BIOKEY_14, BIOKEY_15, BIOKEY_16, BIOKEY_17, BIOKEY_19, BIOKEY_2, BIOKEY_21, BIOKEY_23, BIOKEY_24, BIOKEY_26, BIOKEY_27, BIOKEY_28, BIOKEY_29, BIOKEY_3, BIOKEY_30, BIOKEY_31, BIOKEY_35, BIOKEY_37, BIOKEY_4, BIOKEY_41, BIOKEY_6, BIOKEY_7, BIOKEY_9
	Qian et al. 2020 ($n = 3$)	47, 51, 54
	Wu et al. 2021 ($n = 12$)	CID3921, CID4066, CID4290A, CID4495, CID44971, CID44991, CID4513, CID4515, CID45171, CID4523, CID4530N, CID4535
Colorectal	Lee et al. 2020 ($n = 8$)	SMC01, SMC02, SMC04, SMC07, SMC08, SMC09, SMC16, SMC18
Head and neck	Chen et al. 2020 ($n = 1$)	P11
	Puram et al. 2017 ($n = 1$)	18

Supplementary Table S2 Datasets and patients (samples) of the extended cohort (continued).

Cancer type	Dataset	Patients
Hematologic	Caron et al. 2020 ($n = 2$)	PRE-T.1, PRE-T.2
	van Galen et al. 2019 ($n = 5$)	AML328-D0, AML419A-D0, AML556-D0, AML707B-D0, AML916-D0
	Liu et al. 2021 ($n = 7$)	27522.Primary, 27522.Relapse-2, 56203.Primary, 57075.Relapse-1, 77570.Primary, 81012.Primary, 83942.Primary
	Roider et al. 2020 ($n = 3$)	DLBCL3, FL4, tFL1
	Steen et al. 2021 ($n = 2$)	DLBCL002B, DLBCL007B
	Zhang et al. 2019 ($n = 3$)	FL1018T1, FL2001T1, FL2001T2
Kidney	Bi et al. 2021 ($n = 4$)	P76, P90, P906, P915
	Krishna et al. 2021 ($n = 1$)	UT2.Center
	Obradovic et al. 2021 ($n = 6$)	Patient2, Patient3, Patient4, Patient5, Patient7, Patient8
	Zhang et al. 2021 ($n = 6$)	SI.18854, SI.19703, SI.21561, SI.22604, SI.23459, SI.23843
Lung	Bischoff et al. 2021 ($n = 9$)	p018.T, p019.T, p023.T, p024.T, p027.T, p030.T, p031.T, p032.T, p033.T
	Chan et al. 2021 ($n = 16$)	PleuralEffusion, RU1066, RU1080C, RU1108a.Bambanker.Frozen, RU1124A.LN, RU1144.T, RU1145, RU1152, RU1181C, RU1195A, RU1215, RU1229A.Frozen, RU1262C, RU1293A, RU1322A.LN, RU426B
	Kim et al. 2020 ($n = 5$)	P0018, P0028, P0034, P1006, P1028
	Laughney et al. 2020 ($n = 4$)	RU255B, RU675Tum, RU679, RU681
	Maynard et al. 2020 ($n = 2$)	LT_S21, LT_S63
	Qian et al. 2020 ($n = 4$)	3, 4, 5, 8
	Xing et al. 2021 ($n = 8$)	NM4M, SSN05, SSN22, SSN24, SSN25, SSN27, SSN28, SSN29
	Zilionis et al. 2019 ($n = 5$)	p1, p3, p5, p6, p7
	Jansky et al. 2021 ($n = 1$)	NB04
Neuroendocrine		

Supplementary Table S3 Datasets and patients (samples) of the extended cohort (continued).

Cancer type	Dataset	Patients
Ovarian	Geistlinger et al. 2020 ($n = 4$)	T59, T76, T77, T90
	Olalekan et al. 2021 ($n = 2$)	omentum2834, omentum3401
	Olbrecht et al. 2021 ($n = 5$)	P1.omentum_tumor, P1.ovary_tumor, P2.peritoneum_tumor, P6.peritoneum_tumor, P7.peritoneum_tumor
	Qian et al. 2020 ($n = 4$)	11, 12, 13, 14
	Regner et al. 2021 ($n = 9$)	1, 10, 11, 2, 4, 5, 6, 8, 9
	Zhang et al. 2019 ($n = 2$)	VOA11543L, VOA11543R
Pancreatic	Lin et al. 2020 ($n = 7$)	MET05, MET06, P03, P06, P08, P09, P10
	Moncada et al. 2020 ($n = 1$)	PDAC_A
	Peng et al. 2019 ($n = 6$)	T14, T17, T22, T6, T8, T9
	Raghavan et al. 2021 ($n = 9$)	PANFR0383.Biopsy_None, PANFR0545.Biopsy_None, PANFR0562.Biopsy_None, PANFR0576.Biopsy_None, PANFR0578.Biopsy_None, PANFR0580.Biopsy_None, PANFR0598.Biopsy_None, PANFR0605.Biopsy_None, PANFR0635.Biopsy_None
	Steele et al. 2020 ($n = 8$)	PDAC_TISSUE_1, PDAC_TISSUE_11B, PDAC_TISSUE_15, PDAC_TISSUE_16, PDAC_TISSUE_2, PDAC_TISSUE_3, PDAC_TISSUE_6, PDAC_TISSUE_8
Prostate	Chen et al. 2021 ($n = 9$)	1, 12, 13, 2, 3, 6, 7, 8, 9
	Dong et al. 2020 ($n = 2$)	patient #1, patient #5
	Song et al. 2022 ($n = 4$)	MAY_PB1, PR5196, PR5249_T, PR5251_T
Skin	Biermann et al. 2022 ($n = 15$)	MBM02.sc, MBM03.sc, MBM04.sc, MBM05.sc, MBM05.sn, MBM11.sn, MBM12.sn, MBM14.sn, MBM18.sn, MBM19.sn, MBM20.sn, MPM05.sn, MPM06.sn, MPM07.sn, MPM09.sn
	Jerby-Arnon et al. 2018 ($n = 1$)	Mel79
	Ji et al. 2020 ($n = 6$)	P10, P2, P4, P7, P8, P9