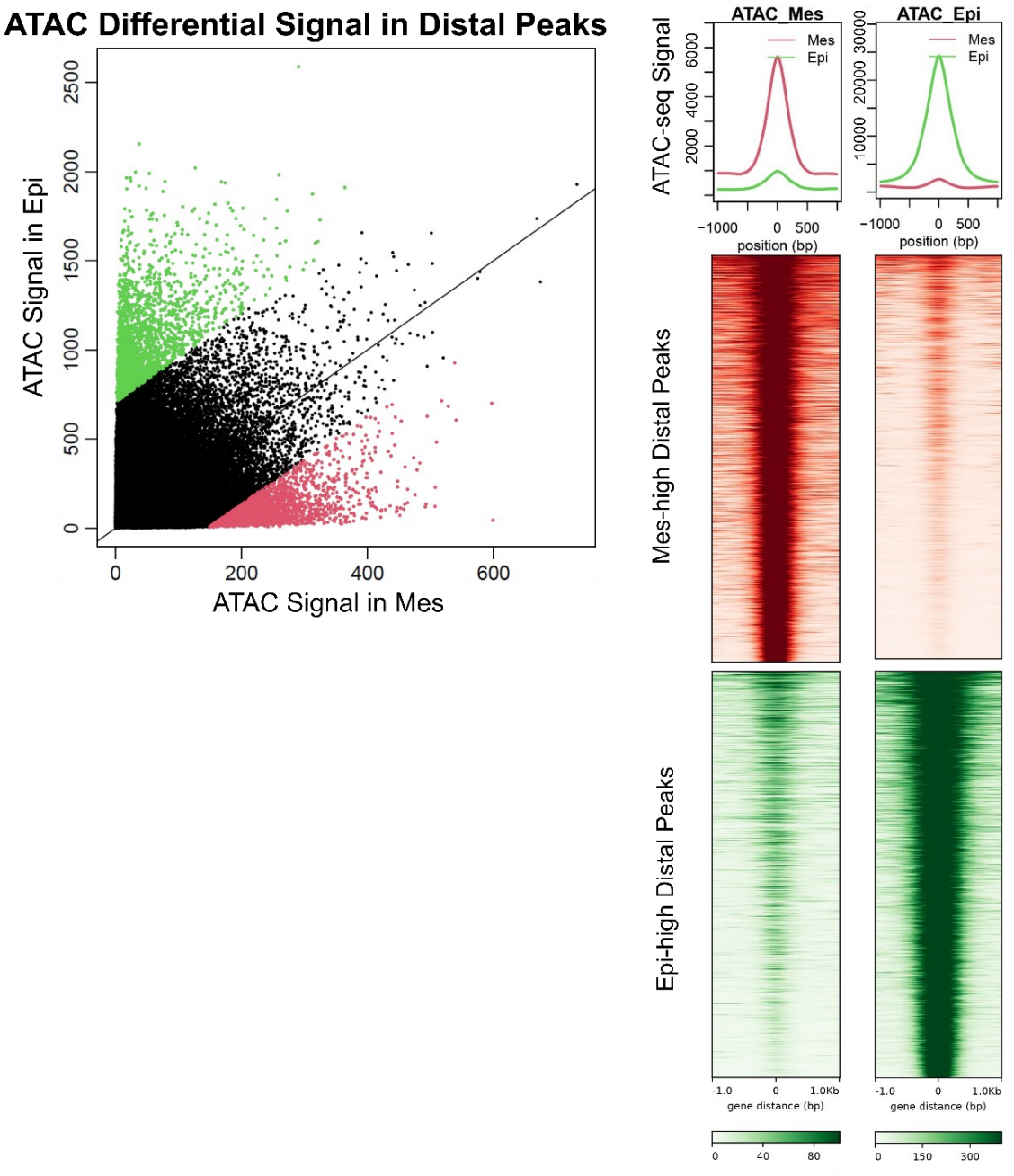
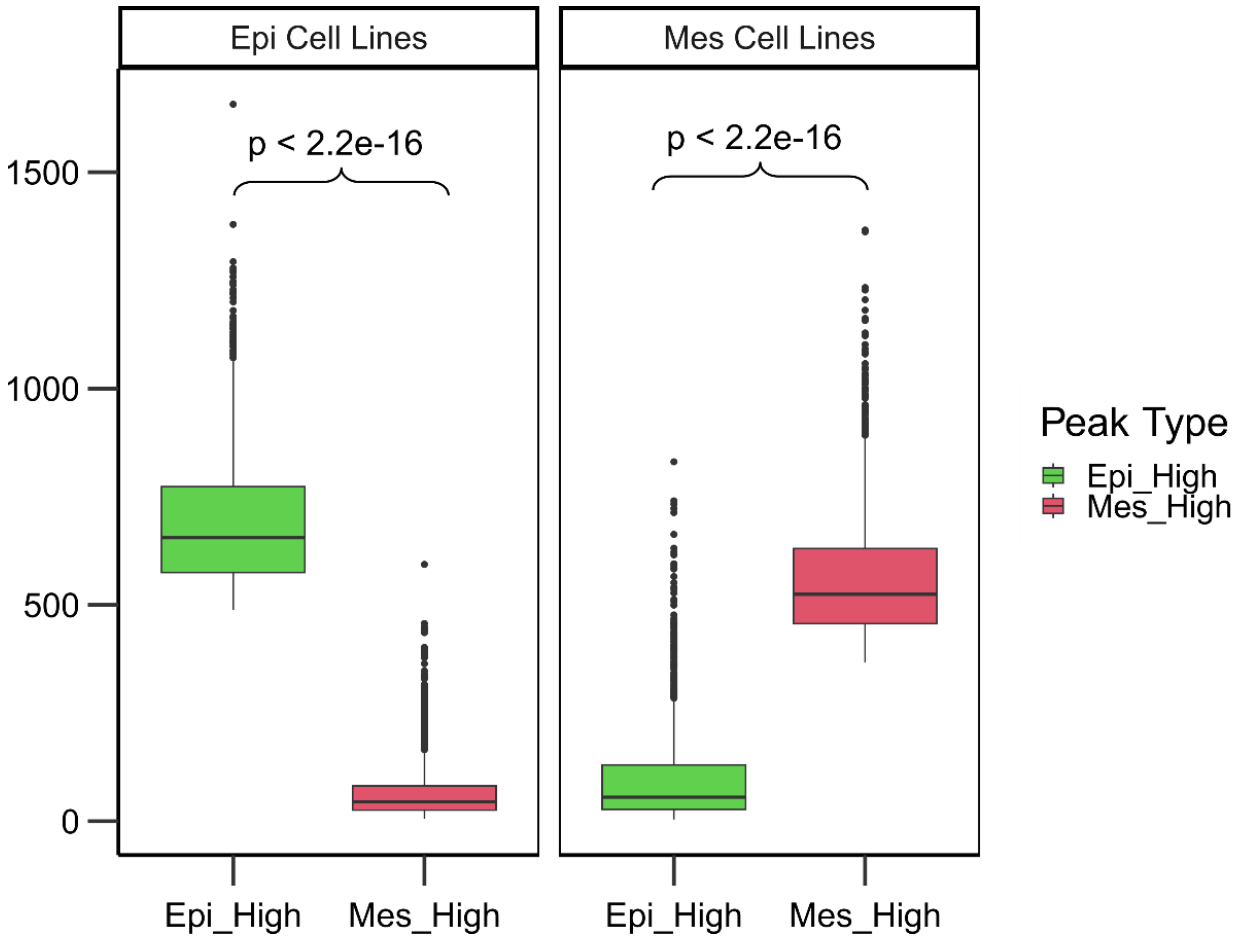
Supplemental Fig. S1) Average ATAC-seq signal in Mes-like and Epi-like differentially accessible peaks (n=2000 peaks in each set)



Distal ATAC peaks with the highest accessibility differences between Mes and Epi GC cell lines (2000 Mes-high & 2000 Epi-high peaks) are shown.

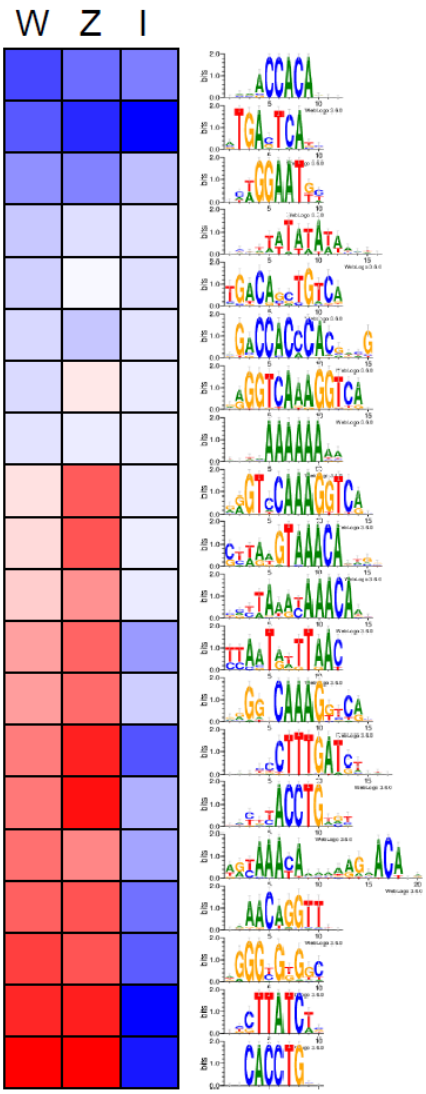
Supplemental Fig. S2) Statistical Validation of ATAC Signal Differences between Mes-high and Epi-high Peaks



We used linear regression to find top 2000 differentially accessible Mes-high and top 2000 Epi-high peaks. As expected, the *t*-test shows a significant difference between the ATAC signal of Mes-high and Epi-high peaks.

Supplemental Fig. S3) Active Motifs in Differentially Accessible Peaks in Mes vs. Epi Cell Line ATAC-seq (2000 Mes-high vs. 2000 Epi-high peaks)

N	ID	Motif	W	Z	I
1	1008	RUNX2	1.00	2.77	0.51
2	895	AP-1	0.90	4.04	1.00
3	1150	TEAD	0.60	2.37	0.25
4	629	Tbp	0.30	0.61	0.13
5	1891	TGIF2LX	0.25	0.14	0.12
6	1428	GLI2	0.21	1.07	0.11
7	1661	NR2F1	0.20	-0.43	0.08
8	1956	ZNF384	0.15	0.35	0.08
9	1464	HNF4A	-0.16	-3.10	0.08
10	46	FREAC4	-0.20	-3.23	0.07
11	1393	FOXE1	-0.27	-2.06	0.08
12	89	HNF1B	-0.51	-2.92	0.40
13	80	HNF4-01-	-0.53	-2.78	0.19
14	222	Lef1	-0.71	-4.20	0.67
15	339	AREB6	-0.73	-4.52	0.31
16	1059	FOXA1:AR	-0.77	-2.29	0.36
17	1436	GRHL2	-0.86	-3.24	0.56
18	1146	KLF5	-1.04	-3.22	0.64
19	1094	Gata2	-1.18	-4.24	0.99
20	1072	E2A	-1.38	-4.81	0.91

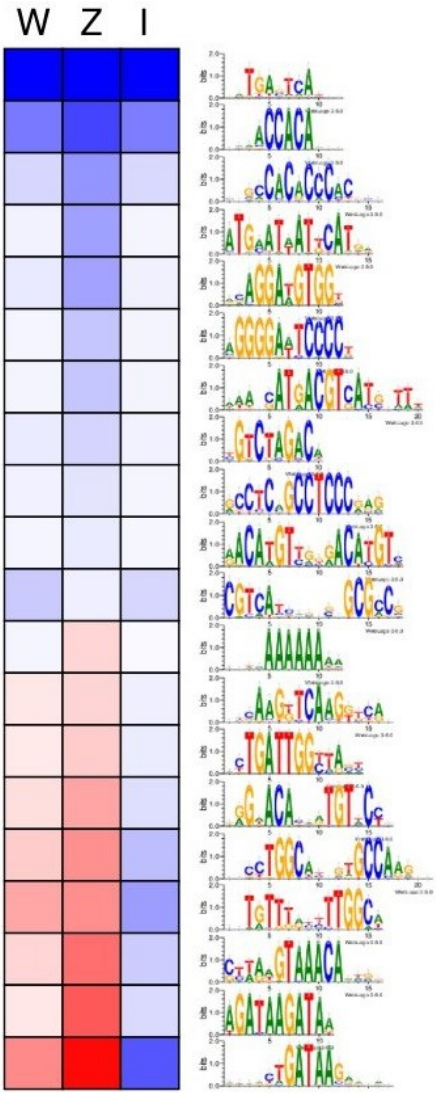


Top 2000 differentially accessible distal peaks between Mes-like and Epi-like GC cell lines (Supplemental Fig. S1) were compared to identify known TFBS DNA motifs, using gkmPWM.

W is the normalized weight for a particular motif found using lasso regression to explain the gapped *k*-mer (gkmSVM) weight space. *Z* is the approximate location of the motif on the gkmSVM weight distribution in terms of *z*-score, and *I* is the relative increase in error when removing the TF/motif from the list.

Supplemental Fig. S4) Active Motifs in Differentially Accessible Peaks in TCGA-STAD Primary Tumor vs. Normal Adult Stomach ATAC-seq

N	ID	Motif	W	Z	I
1	1029	Fosl2	1.00	5.30	1.00
2	1008	RUNX2	0.52	3.90	0.51
3	1541	KLF9	0.16	2.18	0.15
4	1195	OCT:OCT	0.05	1.98	0.03
5	942	ETS:RUNX	0.08	1.89	0.06
6	1627	NFKB1	0.04	1.19	0.04
7	1252	BACH2	0.04	1.18	0.04
8	1794	SMAD5	0.08	0.90	0.05
9	1960	ZNF460	0.05	0.54	0.04
10	1900	TP63	0.04	0.43	0.03
11	820	CREB-ext	0.21	0.34	0.16
12	1956	ZNF384	0.04	-0.85	0.03
13	1674	NR6A1	-0.09	-0.89	0.06
14	47	NFY-C	-0.08	-1.09	0.07
15	1668	NR3C2	-0.12	-1.87	0.12
16	1620	NFIB	-0.20	-2.29	0.26
17	1015	NF1:FOXA	-0.34	-2.32	0.38
18	46	FREAC4	-0.16	-2.98	0.20
19	43	EVI1	-0.09	-3.46	0.14
20	375	Gata5	-0.46	-5.15	0.66

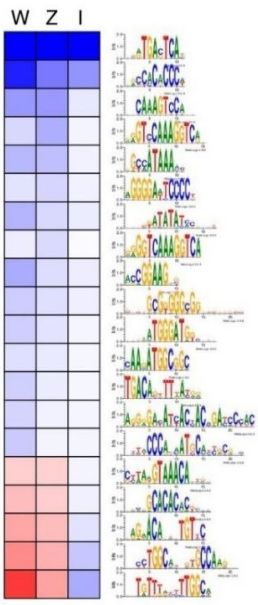


Top 2000 differentially accessible distal peaks between TCGA-STAD (TCGA-BR-A4J6) and healthy adult stomach (ENCODE ENCBS441WEO) were compared to identify TFBS DNA motifs, using gkmPWM.

Supplemental Fig. S5) Active Motifs in Differentially Accessible Peaks in All TCGA-STAD Primary Tumors vs. Normal Adult Stomach ATAC-seq

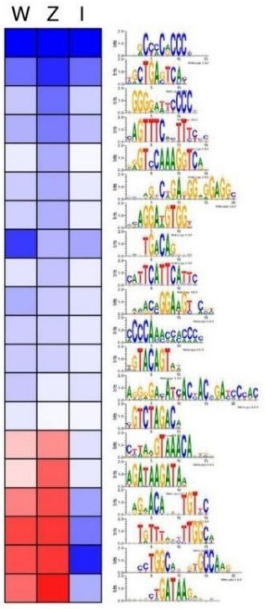
TCGA-STAD ATAC (blue) vs. Normal Stomach (red, ENCODE:ENCBS441WEO)

N	ID	Motif	W	Z	I
1	1377	FOSL2	1.00	7.62	1.00
2	1527	Klf1	0.88	3.99	0.42
3	1463	HNF4A	0.40	3.02	0.09
4	1464	HNF4A	0.15	2.42	0.04
5	1270	CDX1	0.28	1.87	0.05
6	1627	NFKB1	0.11	1.26	0.06
7	600	Six6	0.30	1.13	0.07
8	1777	Rxra	0.07	1.08	0.01
9	261	NRF2	0.32	1.01	0.07
10	826	EGR1	0.19	0.92	0.06
11	176	Zfp410	0.19	0.81	0.03
12	1036	YY1	0.12	0.62	0.03
13	59	MEIS1BHO	0.18	0.53	0.03
14	1955	ZNF382	0.17	0.41	0.04
15	35	STAF	0.21	0.33	0.04
16	46	FREAC4	-0.19	-1.33	0.04
17	239	Gm397	-0.26	-1.44	0.06
18	1181	PGR	-0.35	-2.01	0.10
19	1620	NFIB	-0.45	-2.36	0.21
20	1015	NF1:FOXA	-0.77	-2.76	0.32



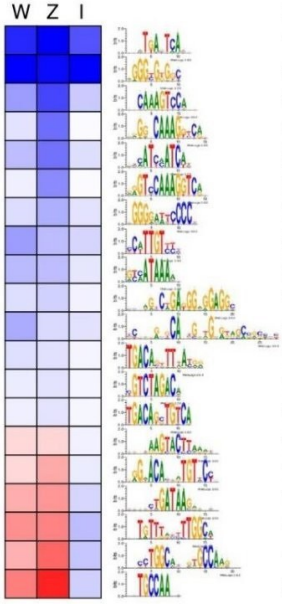
TCGA-CD-A48C (STAD1)

N	ID	Motif	W	Z	I
1	1164	KLF3	1.00	3.87	1.00
2	137	NFE2	0.57	3.25	0.56
3	1628	NFKB2	0.21	2.20	0.18
4	39	ISRE	0.31	1.98	0.26
5	1464	HNF4A	0.07	1.28	0.03
6	1221	ZSCAN22	0.20	1.27	0.09
7	942	ETS:RUNX	0.18	1.19	0.07
8	577	MEIS1	0.77	1.14	0.35
9	1947	ZNF24	0.11	0.90	0.07
10	86	ELK1	0.32	0.80	0.12
11	54	RREB1	0.19	0.72	0.09
12	1923	ZBTB32	0.21	0.69	0.09
13	1955	ZNF382	0.21	0.29	0.12
14	1792	SMAD3	0.09	0.11	0.04
15	46	FREAC4	-0.23	-1.69	0.12
16	43	EVI1	-0.14	-2.43	0.09
17	1181	PGR	-0.49	-2.70	0.36
18	1015	NF1:FOXA	-0.73	-3.02	0.53
19	1620	NFIB	-0.69	-3.06	0.88
20	375	Gata5	-0.58	-3.47	0.32



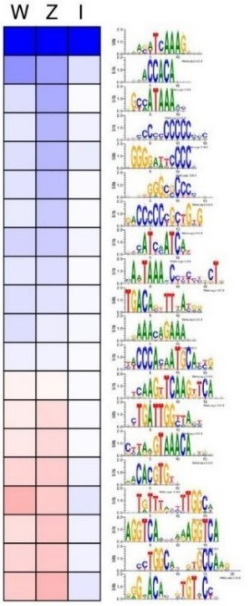
TCGA-VQ-A94O (STAD2)

N	ID	Motif	W	Z	I
1	1106	Fra1	0.85	5.55	0.67
2	1146	KLF5	1.00	5.31	1.00
3	1463	HNF4A	0.39	4.00	0.21
4	80	HNF4-01-	0.11	3.18	0.02
5	1122	HOXA2	0.26	3.08	0.12
6	1464	HNF4A	0.06	2.52	0.02
7	1628	NFKB2	0.13	1.74	0.11
8	1813	Sox6	0.38	1.47	0.16
9	1086	Cdx2	0.27	1.44	0.10
10	1221	ZSCAN22	0.12	0.99	0.05
11	40	PAX5	0.32	0.78	0.11
12	59	MEIS1BHO	0.12	0.66	0.03
13	1792	SMAD3	0.07	0.62	0.03
14	1890	TGIF2	0.06	0.57	0.03
15	362	Nkx3-1	-0.15	-0.90	0.05
16	1668	NR3C2	-0.14	-1.85	0.07
17	375	Gata5	-0.32	-2.27	0.16
18	1015	NF1:FOXA	-0.43	-2.68	0.25
19	1620	NFIB	-0.30	-3.28	0.21
20	1619	NFIA	-0.50	-4.79	0.18



TCGA-VQ-A8PJ (STAD5)

N	ID	Motif	W	Z	I
1	1865	TCF7L2	1.00	7.78	1.00
2	1008	RUNX2	0.43	2.89	0.12
3	1270	CDX1	0.13	2.53	0.02
4	316	Zfp281	0.09	2.20	0.05
5	1628	NFKB2	0.13	2.12	0.07
6	535	Plagl1	0.10	1.93	0.03
7	1933	ZIC1	0.08	1.59	0.02
8	1122	HOXA2	0.17	1.49	0.04
9	38	POLY-C	0.09	1.03	0.01
10	59	MEIS1BHO	0.12	0.85	0.03
11	1835	Stat2	0.13	0.51	0.03
12	1944	ZNF143	0.06	0.38	0.02
13	22	GCNF	-0.05	-0.25	0.01
14	47	NFY-C	-0.08	-1.07	0.02
15	46	FREAC4	-0.09	-1.15	0.02
16	1280	CLOCK	-0.17	-1.51	0.05
17	1015	NF1:FOXA	-0.29	-1.55	0.10
18	1748	Rarb	-0.13	-1.66	0.06
19	1620	NFIB	-0.16	-1.66	0.06
20	1227	Ar	-0.20	-2.21	0.09



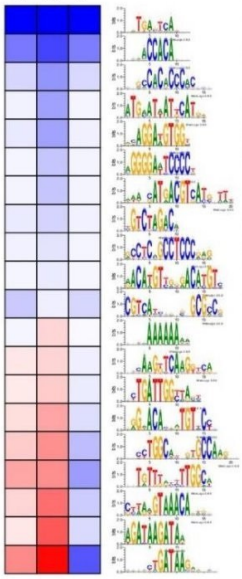
TCGA-VQ-A91W (STAD6)

Top 2000 differentially accessible distal peaks between TCGA-STAD ATAC-seq samples and healthy adult stomach (ENCODE ENCBS441WEO) were compared to identify TFBS DNA motifs, using gkmPWM.

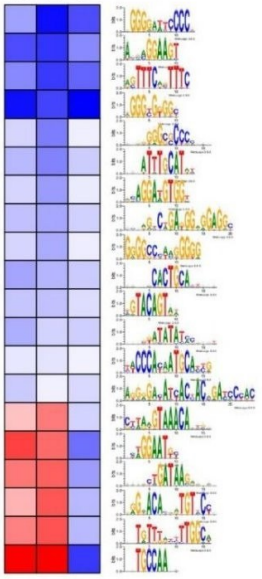
Supplemental Fig. S5 Continued

TCGA-STAD ATAC (blue) vs. Normal Stomach (red, ENCODE:ENCBS441WEO)

N	ID	Motif	W	Z	I
1	1029	Fosl2	1.00	5.30	1.00
2	1008	RUNX2	0.52	3.90	0.51
3	1541	KLF9	0.16	2.18	0.15
4	1195	OCT:OCT	0.05	1.98	0.03
5	942	ETS:RUNX	0.08	1.89	0.06
6	1627	NFKB1	0.04	1.19	0.04
7	1252	BACH2	0.04	1.18	0.04
8	1794	SMAD5	0.08	0.90	0.05
9	1960	ZNF460	0.05	0.54	0.04
10	1900	TP63	0.04	0.43	0.03
11	820	CREB-ext	0.21	0.34	0.16
12	1956	ZNF384	0.04	-0.85	0.03
13	1674	NR6A1	-0.09	-0.89	0.06
14	47	NFY-C	-0.08	-1.09	0.07
15	1668	NR3C2	-0.12	-1.87	0.12
16	1620	NFIB	-0.20	-2.29	0.26
17	1015	NF1:FOXA	-0.34	-2.32	0.38
18	46	FREAC4	-0.16	-2.98	0.20
19	43	EVI1	-0.09	-3.46	0.14
20	375	Gata5	-0.46	-5.15	0.66

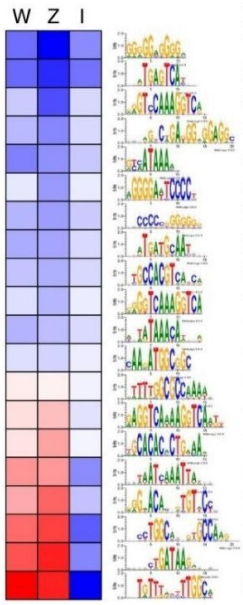


N	ID	Motif	W	Z	I
1	1628	NFKB2	0.39	3.47	0.70
2	967	ELF5	0.74	2.91	0.46
3	1091	ISRE	0.49	2.84	0.61
4	404	Klf4	1.00	2.70	1.00
5	535	Plagl1	0.18	1.87	0.06
6	1714	POU2F2	0.29	1.78	0.20
7	942	ETS:RUNX	0.17	1.43	0.08
8	1221	ZSCAN22	0.26	1.25	0.18
9	41	PLAG1	0.17	1.19	0.07
10	417	Zbtb3	0.38	1.09	0.20
11	1923	ZBTB32	0.24	0.89	0.13
12	600	Six6	0.34	0.64	0.19
13	1944	ZNF143	0.10	0.38	0.05
14	1955	ZNF382	0.22	0.22	0.15
15	46	FREAC4	-0.25	-1.90	0.17
16	1150	TEAD	-0.79	-2.15	0.56
17	375	Gata5	-0.56	-2.29	0.38
18	1668	NR3C2	-0.32	-2.43	0.27
19	1015	NF1:FOXA	-0.52	-2.48	0.27
20	1619	NFIA	-1.07	-3.67	0.78



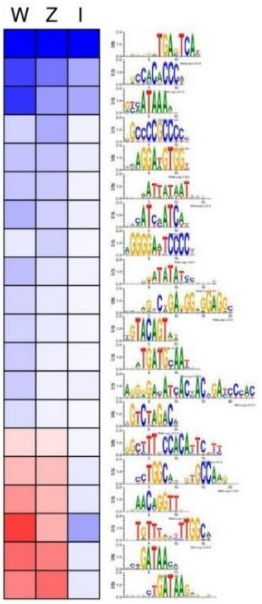
TCGA-BR-A4J6 (STAD12)

N	ID	Motif	W	Z	I
1	890	ZNF281	0.96	3.22	0.45
2	1367	FOS:JUND	1.00	2.67	0.55
3	1464	HNF4A	0.35	2.22	0.15
4	1221	ZSCAN22	0.46	1.50	0.15
5	1086	Cdx2	0.83	1.39	0.21
6	1627	NFKB1	0.16	1.28	0.08
7	596	Zic1	0.49	1.22	0.21
8	1278	CEBPG	0.56	1.13	0.15
9	1283	CREB3L1	0.39	1.02	0.15
10	1781	RXRG	0.22	0.90	0.07
11	126	XFD2	0.50	0.81	0.14
12	1036	YY1	0.22	0.55	0.07
13	1314	E2F3	-0.13	-0.16	0.05
14	1743	RARA	-0.22	-0.81	0.11
15	1968	ZSCAN4	-0.19	-1.16	0.04
16	1703	PHOX2B	-0.70	-1.27	0.45
17	1227	Ar	-0.57	-1.96	0.24
18	1620	NFIB	-0.98	-2.39	0.64
19	375	Gata5	-1.14	-2.79	0.47
20	1015	NF1:FOXA	-1.70	-2.81	1.00



TCGA-BR-A4IY (STAD13)

N	ID	Motif	W	Z	I
1	813	AP1	1.00	6.89	1.00
2	1527	Klf1	0.74	3.70	0.34
3	1086	Cdx2	0.80	2.70	0.34
4	969	Sp1	0.17	2.28	0.06
5	942	ETS:RUNX	0.23	1.61	0.08
6	565	Sox14	0.22	1.43	0.05
7	1122	HOXA2	0.30	1.22	0.07
8	1627	NFKB1	0.08	1.22	0.04
9	600	Six6	0.25	0.80	0.05
10	1221	ZSCAN22	0.17	0.59	0.04
11	1923	ZBTB32	0.17	0.57	0.04
12	1243	ATF4	0.19	0.44	0.04
13	1955	ZNF382	0.20	0.41	0.06
14	1794	SMAD5	0.14	0.32	0.03
15	871	UA11	-0.14	-0.74	0.04
16	1620	NFIB	-0.27	-1.73	0.09
17	1436	GRHL2	-0.42	-1.88	0.09
18	1015	NF1:FOXA	-0.76	-2.14	0.37
19	451	GATA1	-0.59	-3.42	0.09
20	375	Gata5	-0.48	-3.97	0.09



TCGA-BR-A4CS (STAD14)

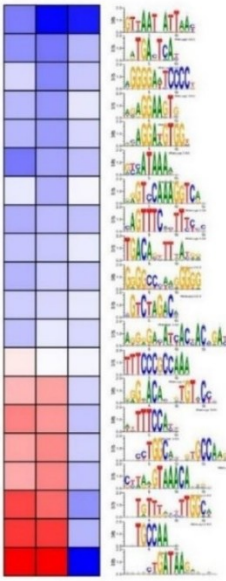
TCGA-HF-A5NB (STAD16)

While there is significant heterogeneity among the tumor samples, and only a fraction of the cells in a tumor are likely in a mesenchymal state, we find significant evidence that the Mes-like regulatory program identified in the cell lines is activated in the STAD samples relative to normal stomach. All of the 10 STAD ATAC samples with >10k distal peaks detect some activation of RUNX or AP-1 when trained against normal stomach DHS (DHS_882): 7 detect AP-1 (STAD1,2,5,12,14,16,19), and 7 detect RUNX (STAD2,6,12,13,16,19,21).

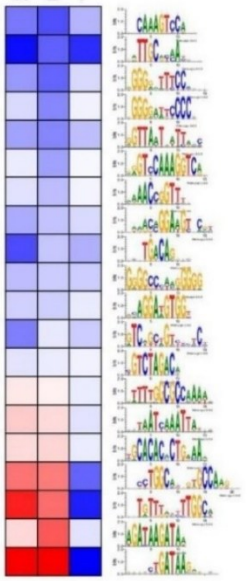
Supplemental Fig. S5 Continued

TCGA-STAD ATAC (blue) vs. Normal Stomach (red, ENCODE:ENCBS441WEO)

N	ID	Motif	W	Z	I
1	1462	HNF1B	0.96	4.41	0.89
2	1380	FOSL2:JU	0.81	2.38	0.28
3	1627	NFKB1	0.30	2.10	0.21
4	978	PU.1	0.52	1.84	0.10
5	942	ETS:RUNX	0.54	1.72	0.16
6	1086	Cdx2	1.00	1.57	0.24
7	1464	HNF4A	0.16	1.37	0.03
8	39	ISRE	0.59	1.18	0.24
9	59	MEIS1BHO	0.43	0.95	0.08
10	1709	PLAG1	0.54	0.87	0.20
11	1794	SMAD5	0.36	0.68	0.09
12	1955	ZNF382	0.46	0.38	0.15
13	1318	E2F8	-0.16	0.08	0.04
14	1227	Ar	-0.54	-1.66	0.16
15	954	NFAT	-0.92	-1.81	0.22
16	1620	NFIB	-0.67	-2.07	0.18
17	46	FREAC4	-0.64	-2.11	0.24
18	1015	NF1:FOXA	-1.39	-2.63	0.42
19	1619	NFIA	-1.50	-3.50	0.28
20	375	Gata5	-1.92	-4.54	1.00



N	ID	Motif	W	Z	I
1	1463	HNF4A	0.55	3.42	0.29
2	818	CEBPB	1.00	3.08	0.82
3	844	NFKB1	0.42	3.02	0.15
4	1628	NFKB2	0.15	2.34	0.05
5	131	HNF1A	0.38	2.33	0.25
6	1464	HNF4A	0.08	1.84	0.03
7	1435	GRHL1	0.12	1.29	0.04
8	86	ELK1	0.35	1.16	0.14
9	577	MEIS1	0.74	1.09	0.31
10	41	PLAG1	0.28	0.99	0.19
11	942	ETS:RUNX	0.18	0.93	0.08
12	1158	ZNF317	0.53	0.43	0.19
13	1792	SMAD3	0.11	0.38	0.05
14	1314	E2F3	-0.09	-0.28	0.07
15	1703	PHOX2B	-0.21	-0.70	0.11
16	1968	ZSCAN4	-0.14	-0.79	0.07
17	1620	NFIB	-0.57	-2.56	0.61
18	1015	NF1:FOXA	-0.92	-2.84	0.87
19	43	EVI1	-0.15	-3.23	0.11
20	375	Gata5	-1.06	-5.01	1.00

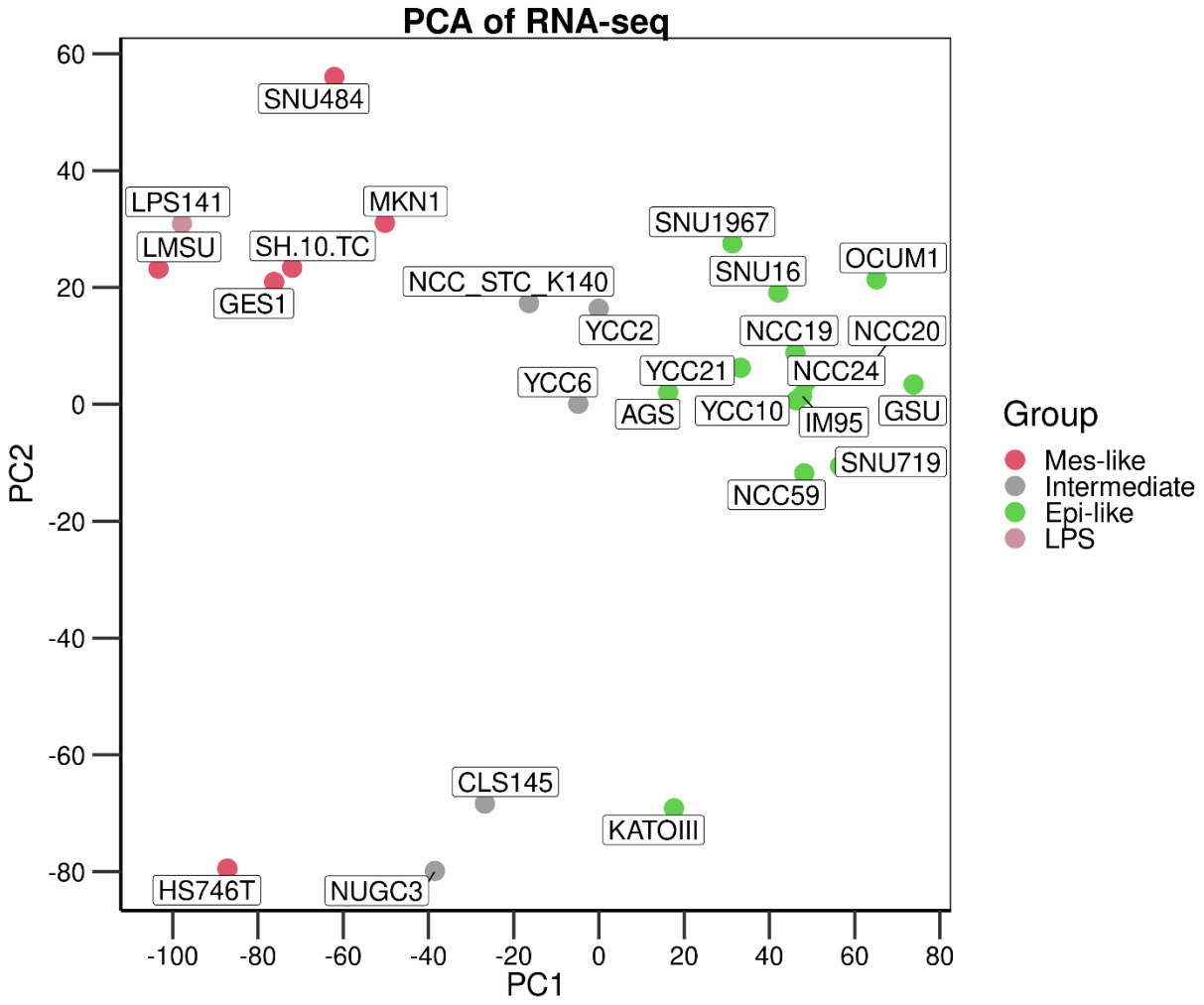


TCGA-BR-A4J4 (STAD19)

TCGA-CD-A486 (STAD21)

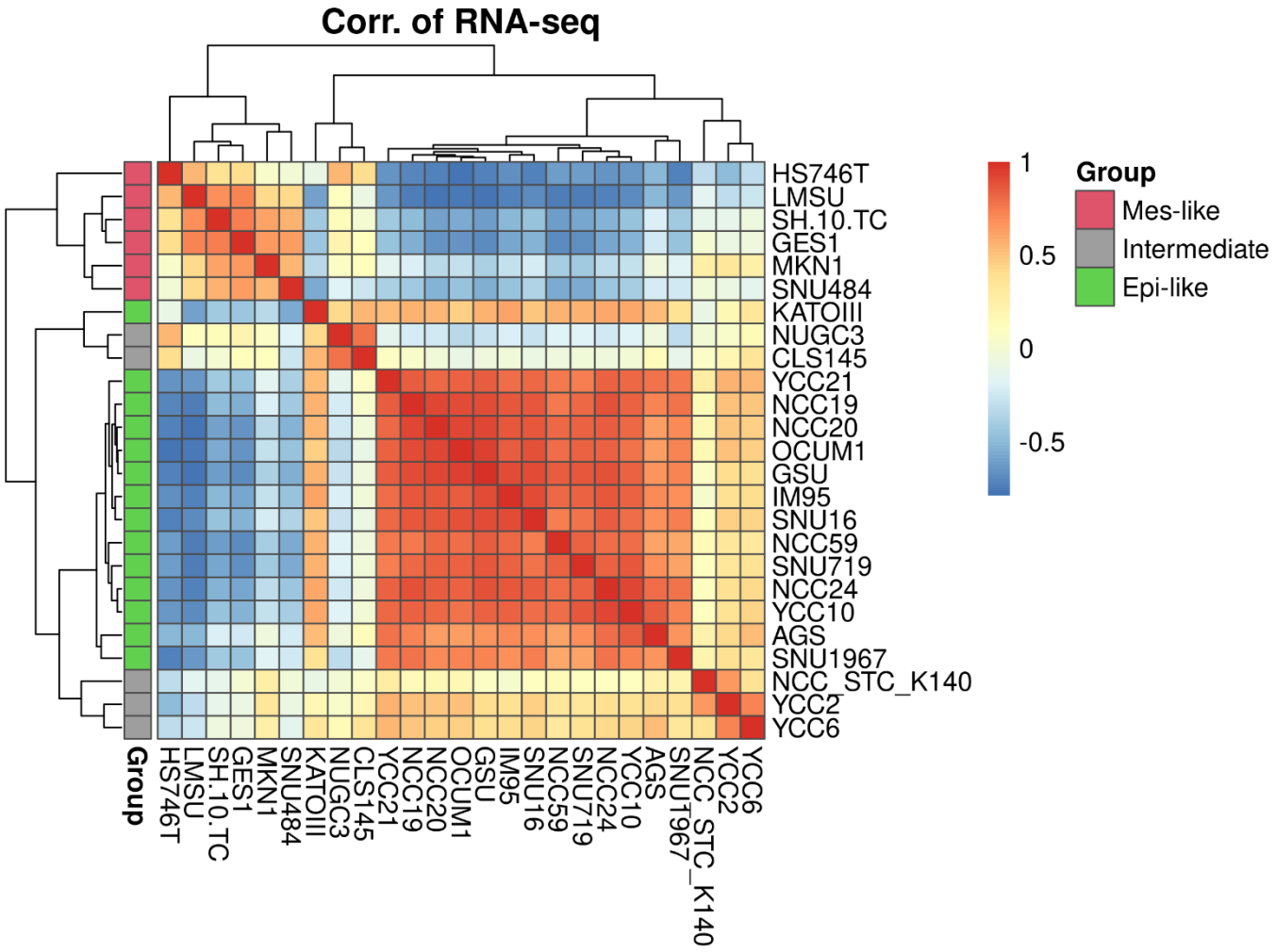
W is the normalized weight for a particular motif found using lasso regression to explain the gapped *k*-mer (gkmSVM) weight space. *Z* is the approximate location of the motif on the gkmSVM weight distribution in terms of *z*-score, and I is the relative increase in error when removing the TF/motif from the list.

Supplemental Fig. S6) LPS141 Transcriptional Profile Compared to GC Cell Lines



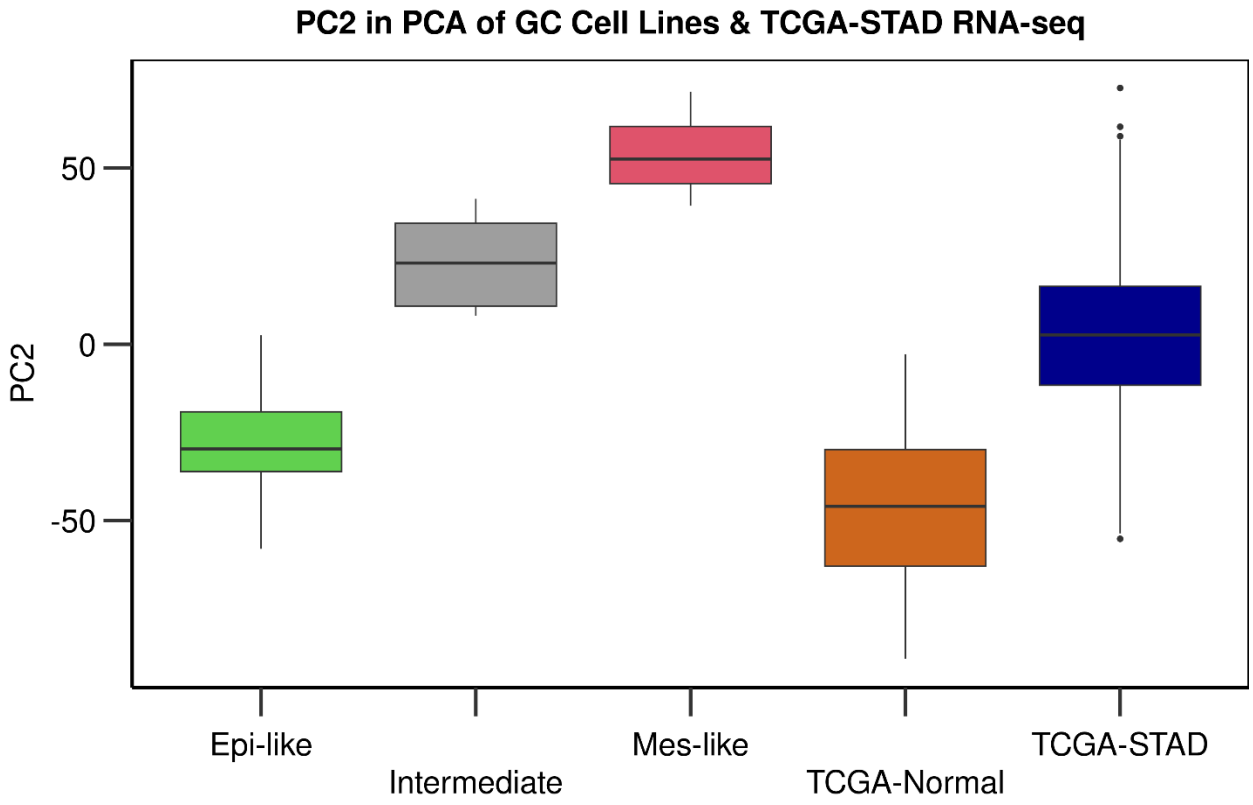
PCA of RNA-seq over ~11,300 tissue-specific (see Methods) protein-coding genes. LPS141 mesenchymal liposarcoma cell line has a very similar transcriptional profile to Mes-like GC cell lines.

Supplemental Fig. S7) Correlation Heatmap of RNA-seq in GC Cell Lines



Correlation heatmap of RNA-seq profiles is consistent with the ATAC-seq clustering.

Supplemental Fig. S8) PC2 of RNA-seq PCA



t-test p-values are:

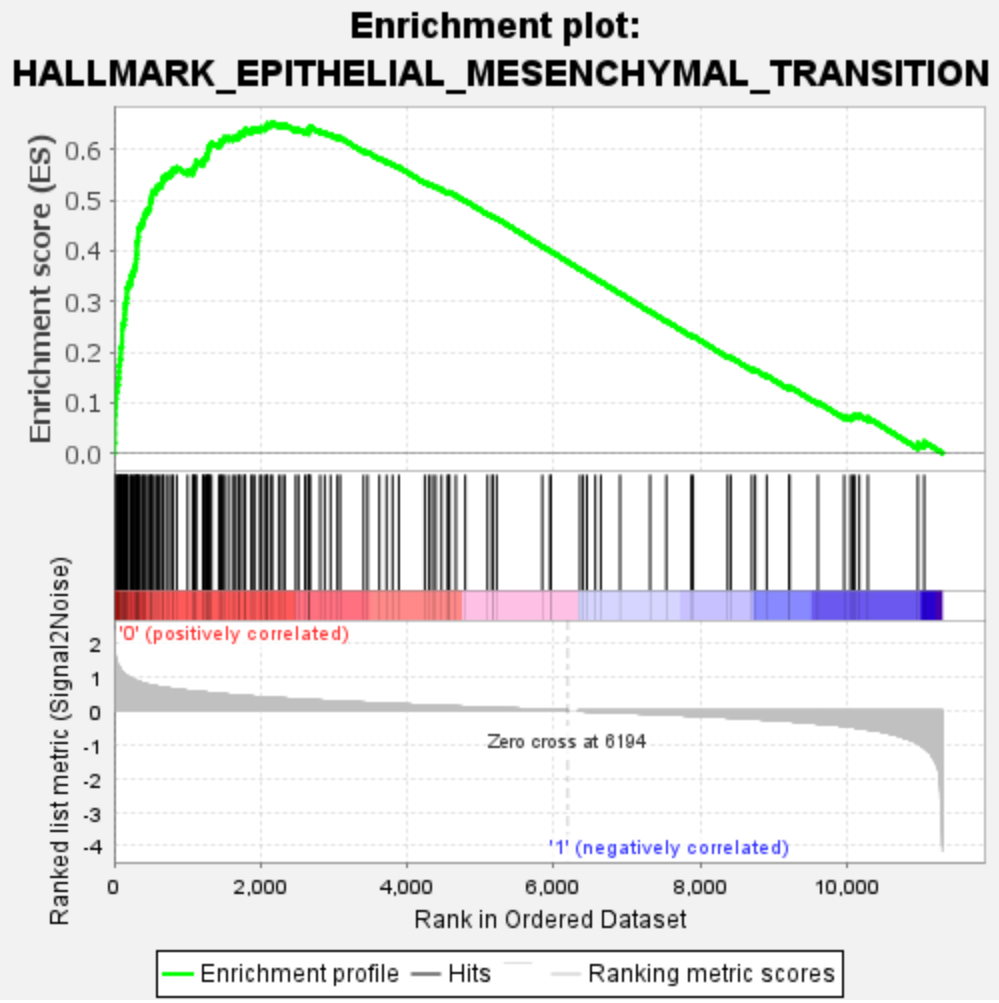
- (Mes vs. Intermediate) = 6×10^{-3}
- (Intermediate vs. Epi) = 2×10^{-4}
- (TCGA-STAD vs. TCGA-Normal) = 4×10^{-7}

Supplemental Fig. S9) Gene Set Enrichment Analysis (GSEA) of Hallmark Gene Sets for Mes vs. Epi GC Cell Lines

NAME	SIZE	ES	NES	NOM p-val	FDR q-val	FWER p-val	RANK AT MAX
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	167	0.6537	1.7740	0.006787	0.127605	0.098	2157
HALLMARK_UV_RESPONSE_DN	109	0.5611	1.7649	0.002203	0.070798	0.11	1824
HALLMARK_HEDGEHOG_SIGNALING	30	0.5001	1.4848	0.061674	0.459550	0.616	2714
HALLMARK_ANGIOGENESIS	32	0.4318	1.4159	0.064593	0.506993	0.759	2974
HALLMARK_HYPOXIA	137	0.3726	1.3525	0.110619	0.556421	0.839	2148
HALLMARK_MYOGENESIS	150	0.3575	1.1952	0.208333	0.898935	0.96	2591
HALLMARK_SPERMATOGENESIS	106	0.3149	1.1467	0.261851	0.920096	0.977	2599
HALLMARK_COMPLEMENT	151	0.2772	1.1057	0.291469	0.925865	0.986	2475
HALLMARK_TGF_BETA_SIGNALING	31	0.3515	1.1021	0.375926	0.834174	0.988	2096
HALLMARK_WNT_BETA_CATENIN_SIGNALING	27	0.3609	1.0843	0.350427	0.797242	0.991	1578
HALLMARK_APICAL_JUNCTION	137	0.2898	1.0344	0.402715	0.844931	0.997	1294
HALLMARK_ALLOGRAFT_REJECTION	145	0.2656	1.0307	0.384804	0.782085	0.998	1433
HALLMARK_G2M_CHECKPOINT	125	0.4053	1.0194	0.49434	0.745905	0.999	5109
HALLMARK_APOPTOSIS	95	0.2420	0.9759	0.496659	0.784445	1	744
HALLMARK_GLYCOLYSIS	125	0.2200	0.9412	0.529175	0.800791	1	1030
HALLMARK_APICAL_SURFACE	37	0.3030	0.9257	0.551282	0.782200	1	2169
HALLMARK_ADIPOGENESIS	84	0.2313	0.9101	0.587054	0.764469	1	2506
HALLMARK_COAGULATION	112	0.2474	0.8992	0.633929	0.742667	1	2457
HALLMARK_INFLAMMATORY_RESPONSE	163	0.2507	0.8852	0.602871	0.729837	1	1865
HALLMARK_E2F_TARGETS	113	0.3851	0.8851	0.617761	0.693487	1	5021
HALLMARK_MITOTIC_SPINDLE	108	0.2739	0.8659	0.575368	0.689749	1	4053
HALLMARK_DNA_REPAIR	46	0.2116	0.7971	0.636015	0.762965	1	4915
HALLMARK_TNFA_SIGNALING_VIA_NFKB	146	0.2047	0.6804	0.826962	0.875369	1	2150

GSEA between Mes and Epi GC cell lines was performed using “Hallmark” gene sets.

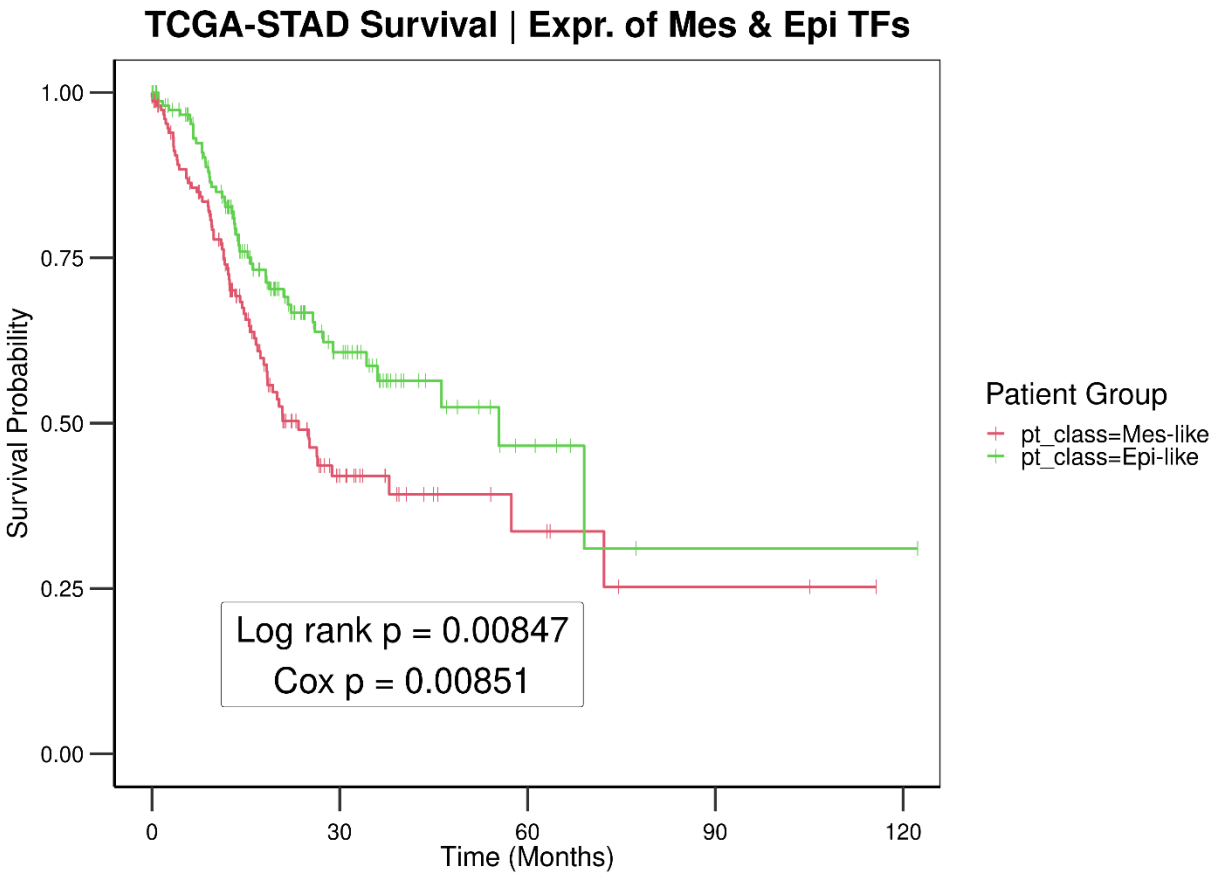
Supplemental Fig. S10) GSEA Enrichment Plot for Hallmark of EMT Gene Set



In the enrichment plot:

- Cluster 1: Epi-like GC Cell Lines
- Cluster 3: Mes-like GC Cell Lines

Supplemental Fig. S11) TCGA-STAD KM-plot Based on Mes vs Epi TF Expression

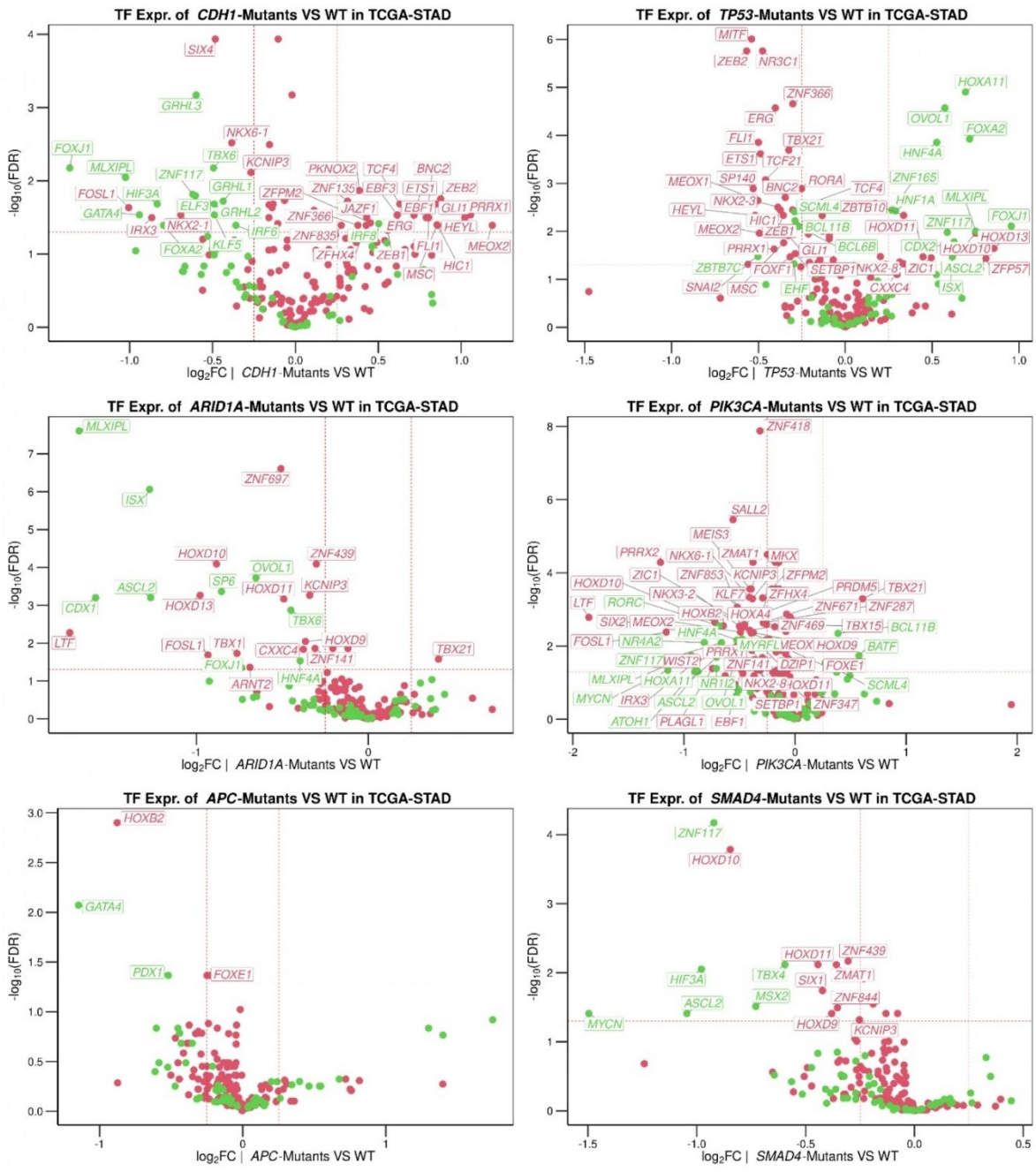


We calculated the difference between the expression of Mes and Epi differentially expressed TFs on TCGA-STAD RNA-seq. Mes and Epi TFs were defined from the differential gene expression analysis of the GC cell lines, and those differentially expressed genes that are TFs are used here for stratifying TCGA-STAD patients.

Difference = [Average(Mes TF expression) – Average(Epi TF expression)].

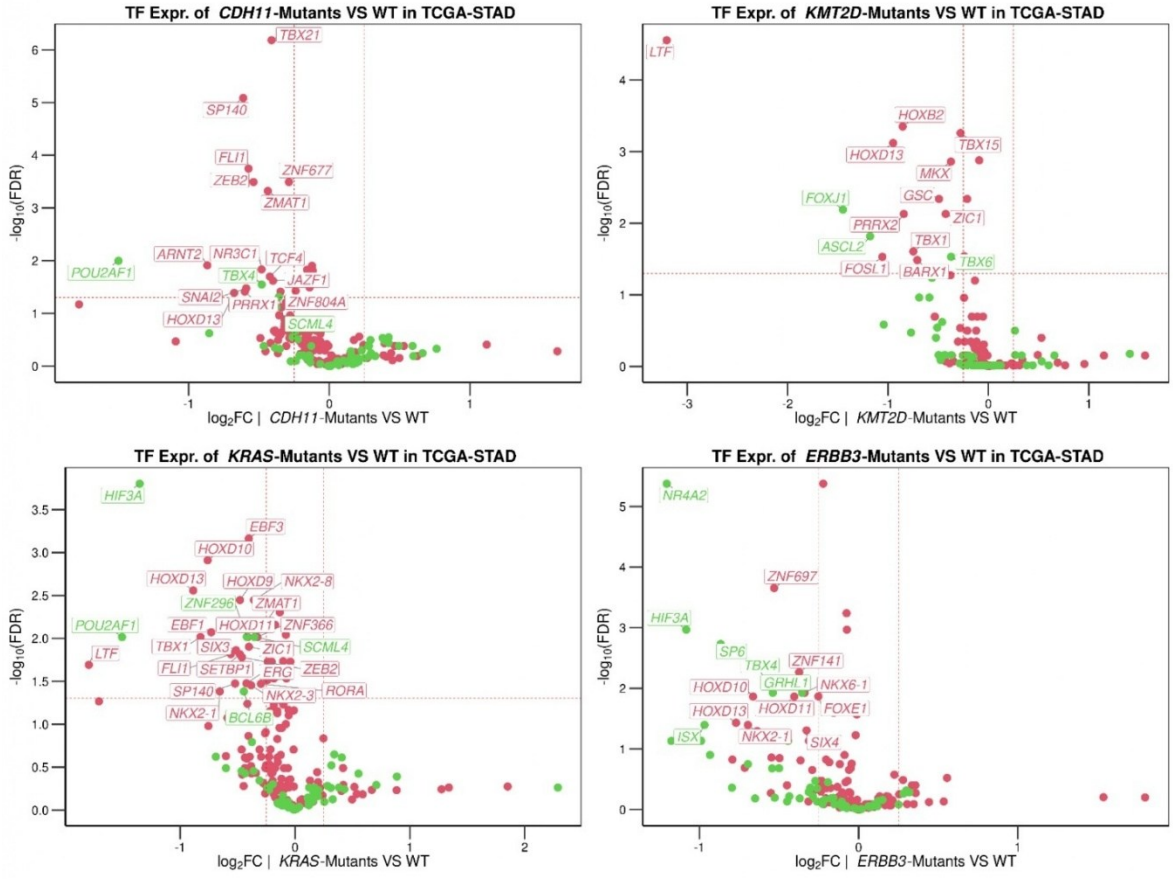
TCGA-STAD samples were sorted by the difference in the average expression of Mes vs. Epi TFs. Those with a differential value above the median were assigned to the STAD Mes-like patient group, and the rest of the STAD patients were assigned to the STAD Epi-like group for this survival plot (and similarly for ACRG survival plot Supplemental Fig. 2G).

Supplemental Fig. S12) Changes in TF Expression in TCGA-STAD Samples with or without Driver Gene Mutation

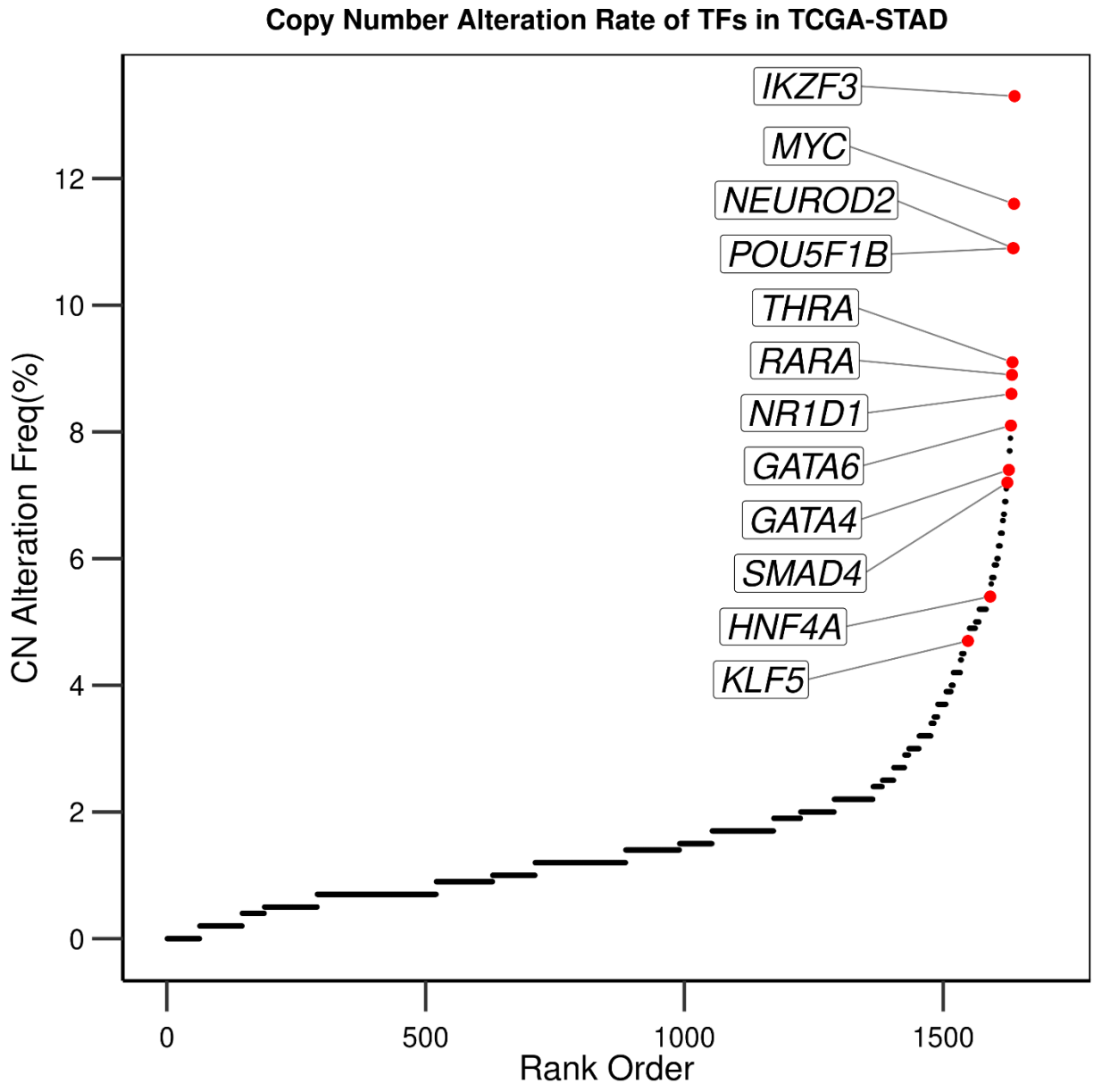


X-axis: $\log_2$(TF RNA in STAD Mutant Sample / TF RNA in STAD WT)

Supplemental Fig. S12 Continued

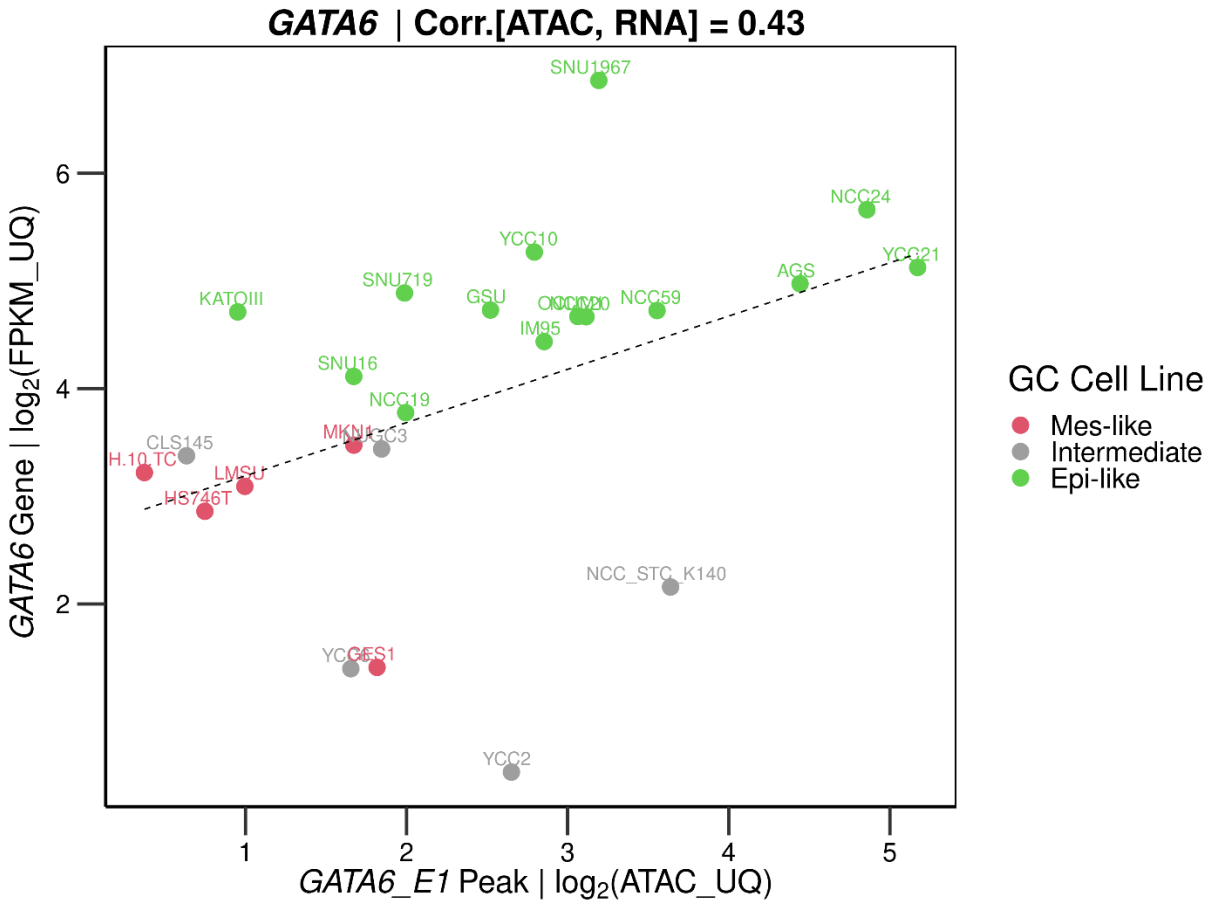

TF Expr. of CDH11-Mutants VS WT in TCGA-STAD
-log10(FDR)
log2FC | CDH11-Mutants VS WT
TBX21
SP140
FLI1
ZEB2
ZNF677
ZMAT1
ARNT2
NR3C1
TCF4
POU2AF1
TBX4
JAZF1
SNAI2
PRRX1
ZNF804A
HOXD13
SCML4
TF Expr. of KMT2D-Mutants VS WT in TCGA-STAD
-log10(FDR)
log2FC | KMT2D-Mutants VS WT
LTF
HOXB2
HOXD13
TBX15
MKX
FOXJ1
GSC
ZIC1
PRRX2
TBX1
ASCL2
FOSL1
BARX1
TBX6
TF Expr. of KRAS-Mutants VS WT in TCGA-STAD
-log10(FDR)
log2FC | KRAS-Mutants VS WT
HIF3A
EBF3
HOXD10
HOXD13
HOXD9
NKX2-8
ZNF296
ZMAT1
POU2AF1
EBF1
HOXD11
ZNF366
LTF
TBX1
SIX3
ZIC1
SCML4
FLI1
SETBP1
ERG
ZEB2
SP140
NKX2-3
RORA
NKX2-1
BCL6B
TF Expr. of ERBB3-Mutants VS WT in TCGA-STAD
-log10(FDR)
log2FC | ERBB3-Mutants VS WT
NR4A2
ZNF697
HIF3A
SP6
TBX4
ZNF141
HOXD10
GRHL1
NKX6-1
HOXD13
HOXD11
FOXE1
ISX
NKX2-1
SIX4

Supplemental Fig. S13) Copy Number Alteration Rates in TCGA-STAD



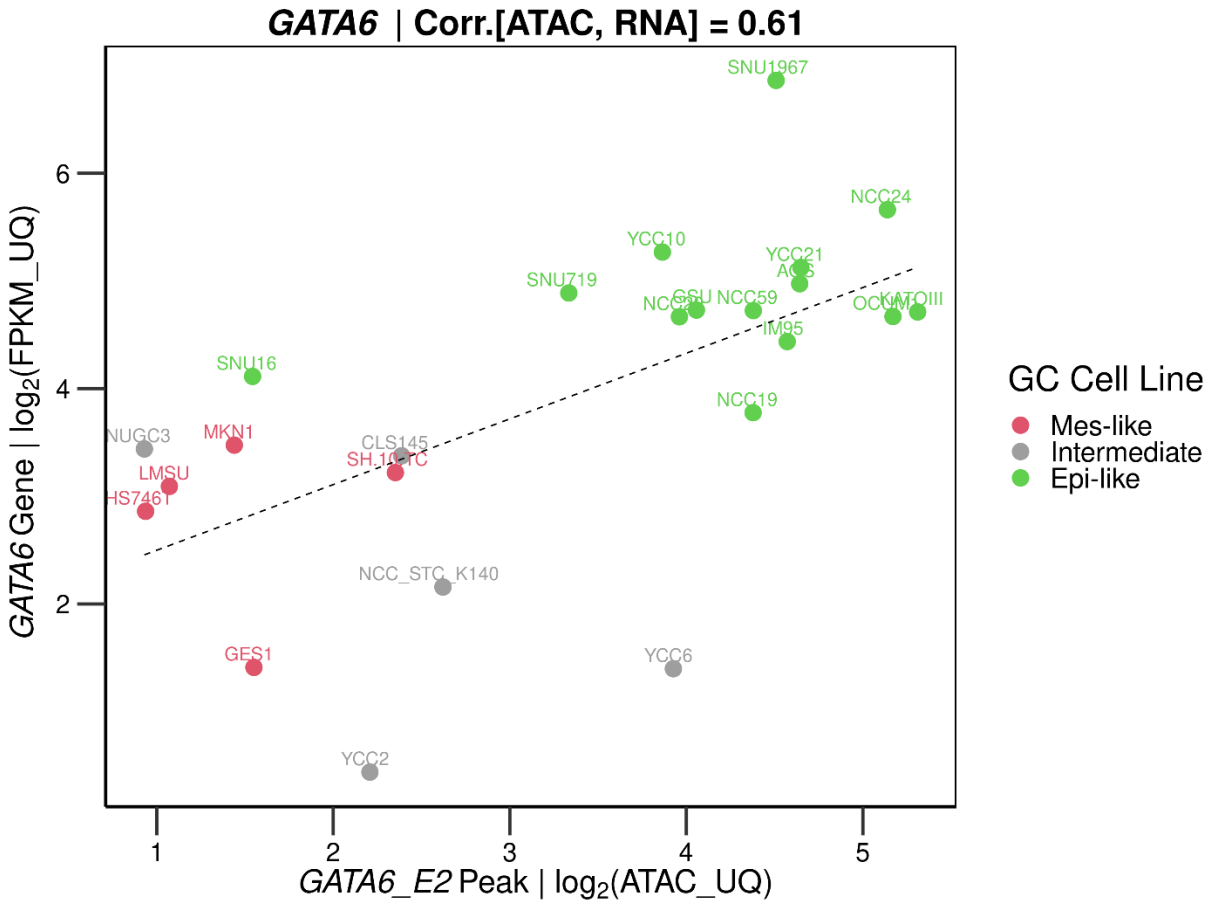
Copy number alteration = DNA copy number amplification rate + DNA copy number deletion rate

Supplemental Fig. S14) Correlation of GATA6 Enhancer (*E1*) Accessibility and *GATA6* Gene Expression



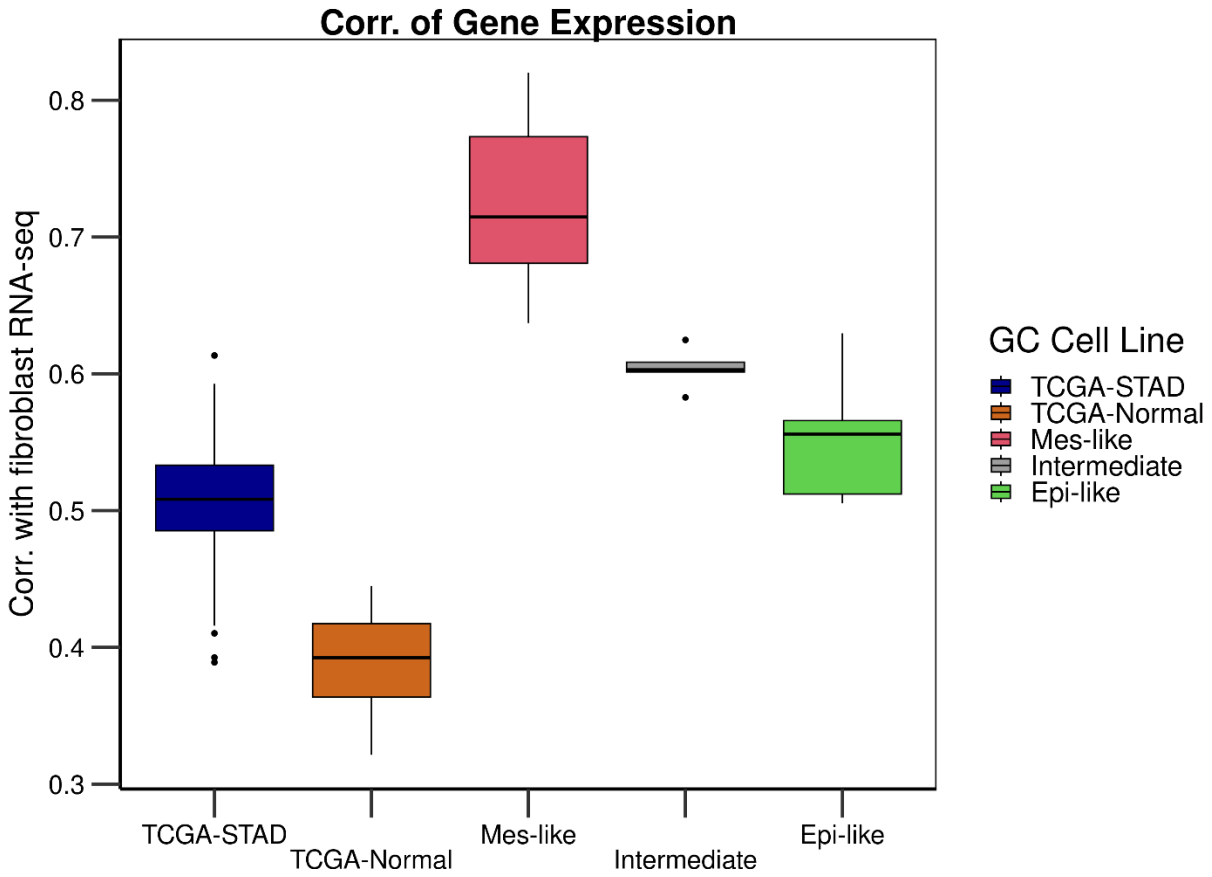
Pearson's correlation of [*GATA6* RNA, *GATA6_E1* ATAC] signals. ATAC-seq and RNA-seq values are upper-quartile normalized.

Supplemental Fig. S15) Correlation of GATA6 Enhancer (*E2*) Accessibility and *GATA6* Gene Expression



Pearson's correlation of [*GATA6* RNA, *GATA6_E2* ATAC] signals. ATAC-seq and RNA-seq values are upper-quartile normalized.

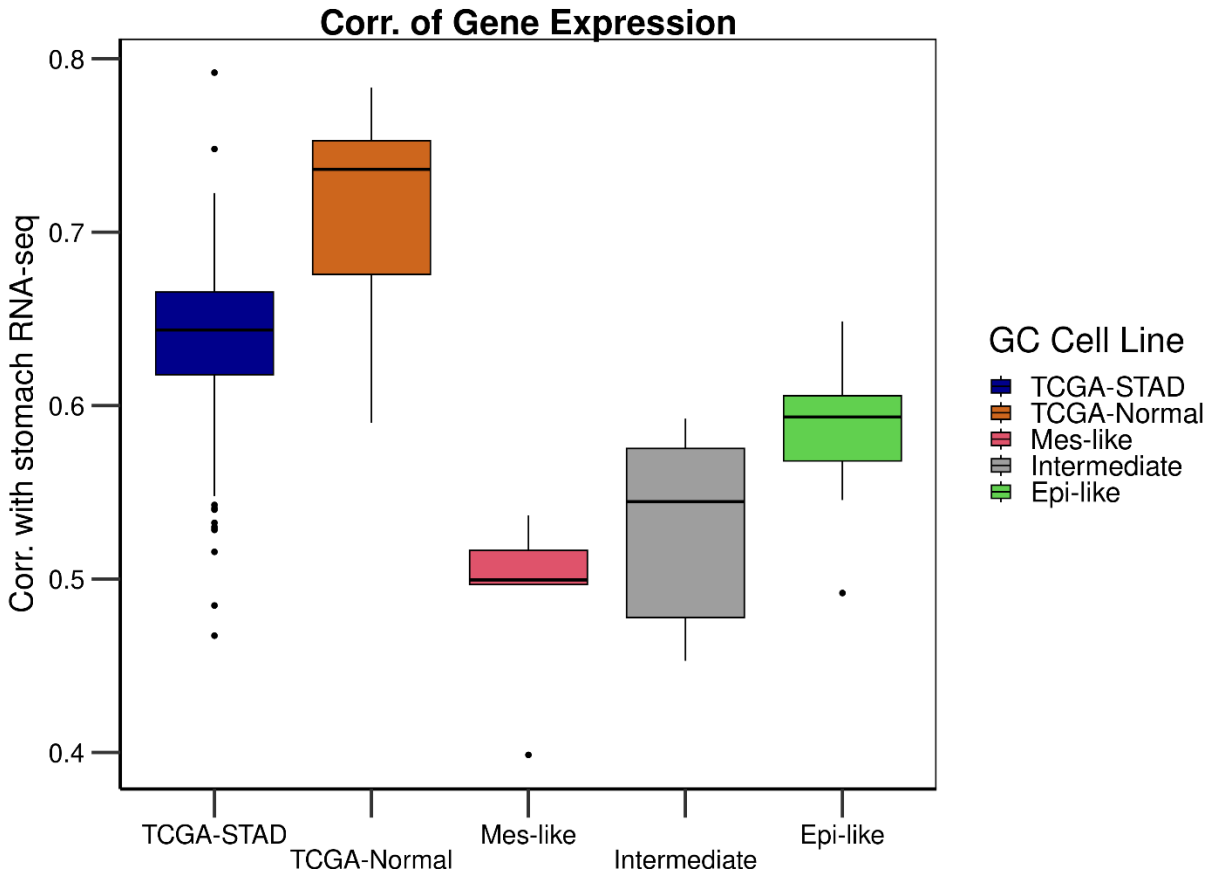
Supplemental Fig. S16) Correlation of RNA-seq for GC Cell Lines and TCGA-STAD with ENCODE Primary Lung Fibroblast



t-test p-values:

- (Mes vs. Epi) = 9.9×10^{-4}
- (TCGA-STAD vs. TCGA-Normal) = 1.2×10^{-10}

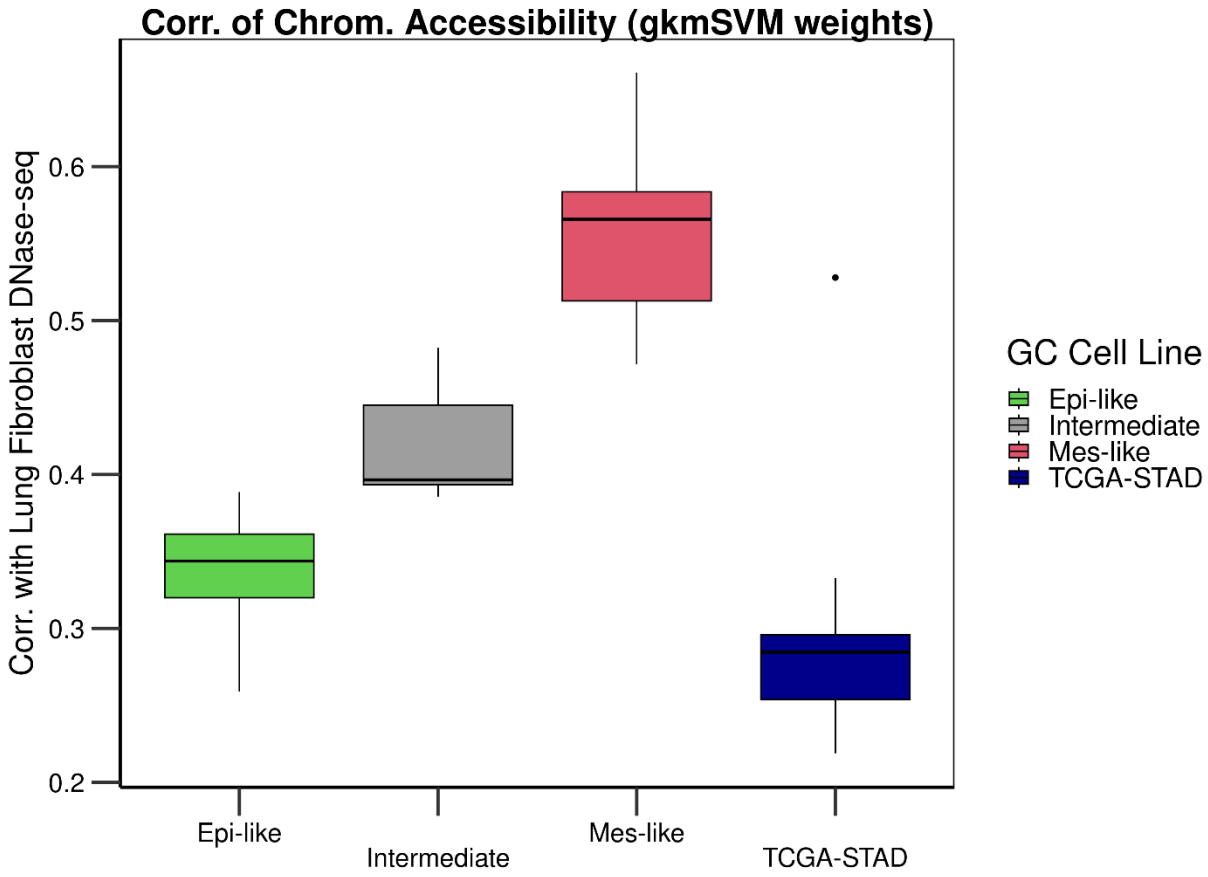
Supplemental Fig. S17) Correlation of RNA-seq for GC Cell Lines and TCGA-STAD with ENCODE Primary Stomach



t-test p-values:

- (Mes vs. Epi) = 0.0032
- (TCGA-STAD vs. TCGA-Normal) = 9.3×10^{-5}

Supplemental Fig. S18) Correlation of Chromatin Accessibility for GC Cell Lines and TCGA-STAD with ENCODE Primary Lung Fibroblast

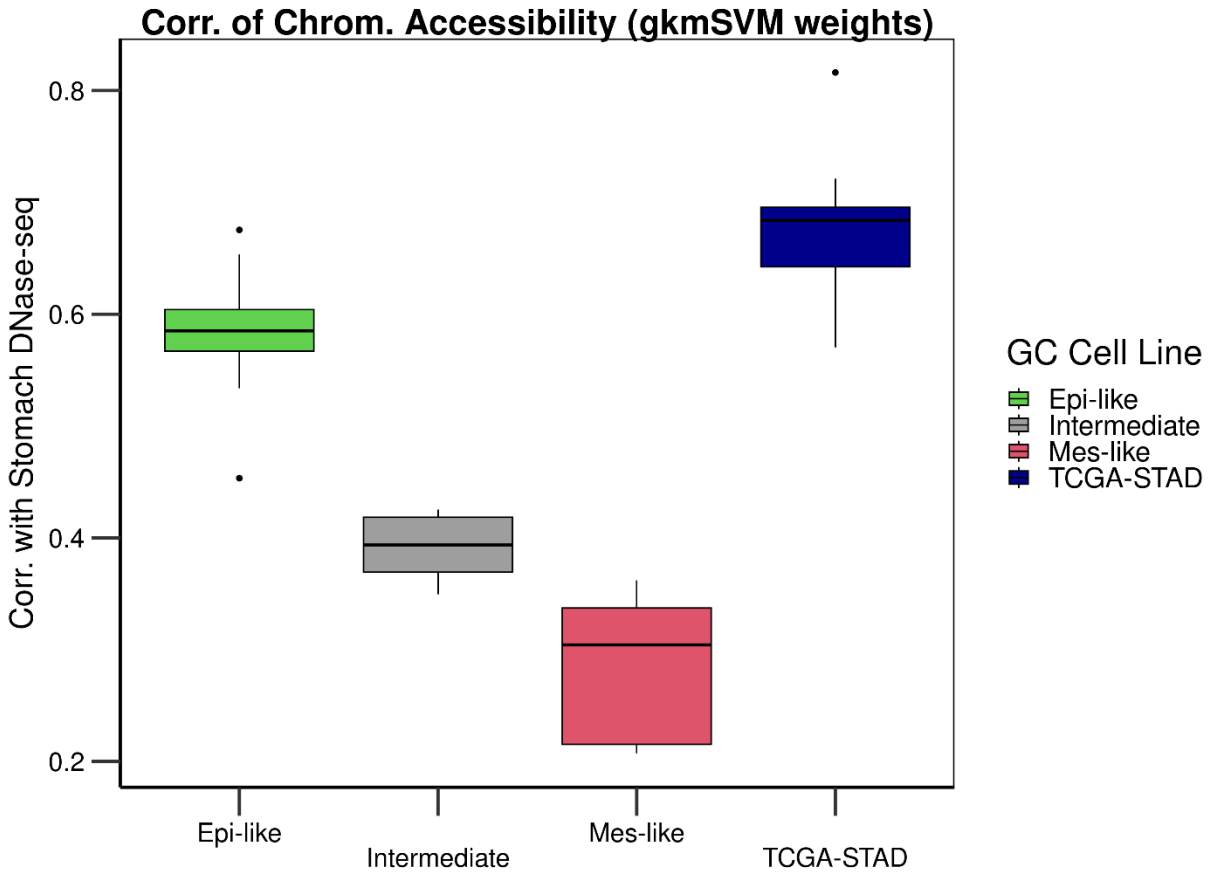


Chromatin accessibility is measured by ATAC-seq in the GC cell lines and with DNase-seq in the ENCODE healthy primary tissue. For each sample, gkm-SVM models were trained on distal enhancer DNA regions (peak length = 300). The gkm-SVM output is a weight vector for (4^{11} / 2) *k*-mers (k = 11) which shows the overrepresentation and underrepresentation of that particular *k*-mer (transcription factor binding site) in the ATAC-seq or DNase-seq. Correlation of these weight vectors were used to measure similarity in the regulatory landscape and chromatin accessibility of different samples.

t-test p-values:

- (Mes vs. Epi) = 0.0014
- (TCGA-STAD vs. Epi) = 0.2396

Supplemental Fig. S19) Correlation of Chromatin Accessibility for GC Cell Lines and TCGA-STAD with ENCODE Primary Stomach

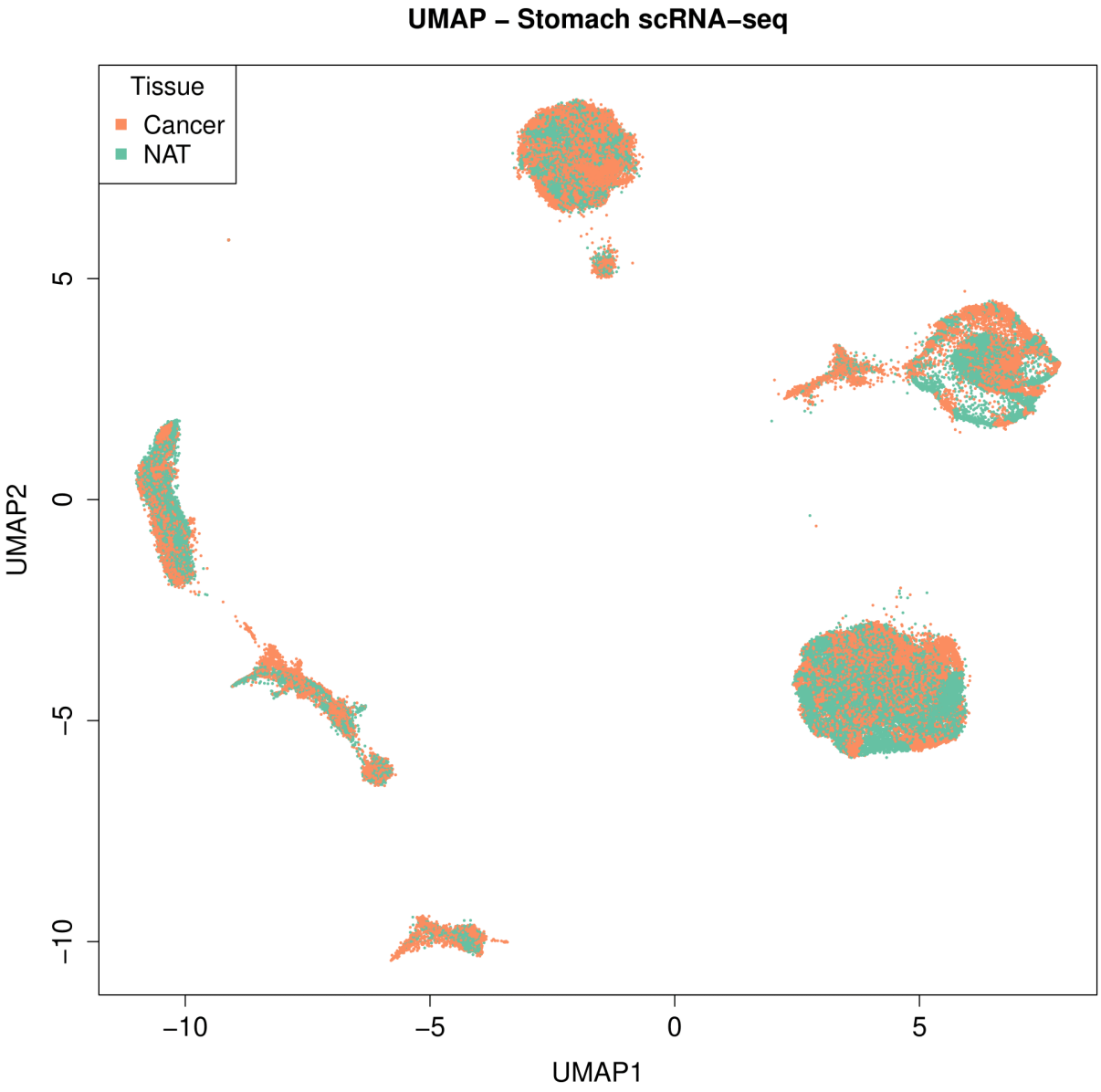


Correlation of gkm-SVM *k*-mer weight vectors was used to measure similarity in the regulatory landscape and chromatin accessibility of different samples.

t-test p-values:

- (Mes vs. Epi) = 0.00018
- (TCGA-STAD vs. Epi) = 0.0039

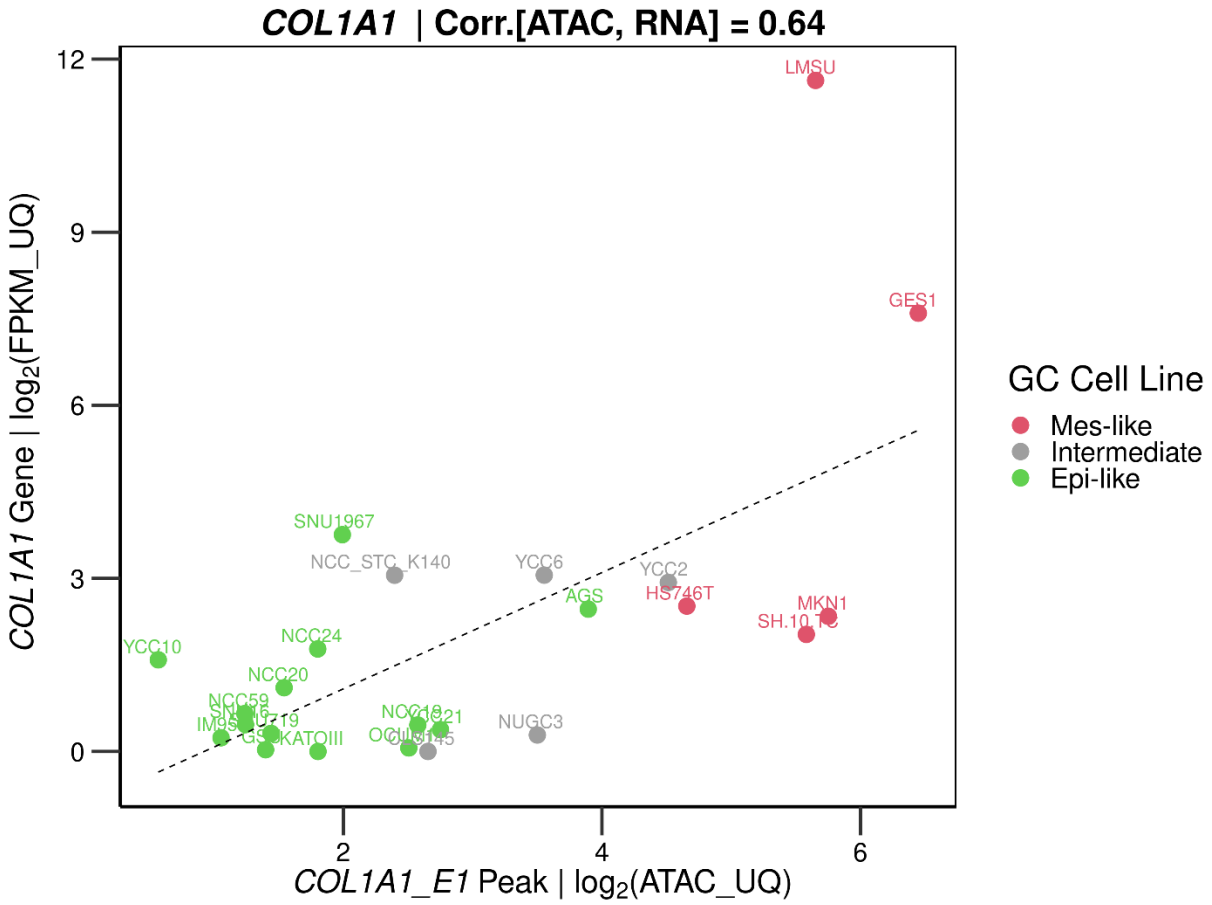
Supplemental Fig. S20) UMAP of scRNA-seq of GC Tumors and Normal Adjacent Tissue



NAT: Normal Adjacent Tissue

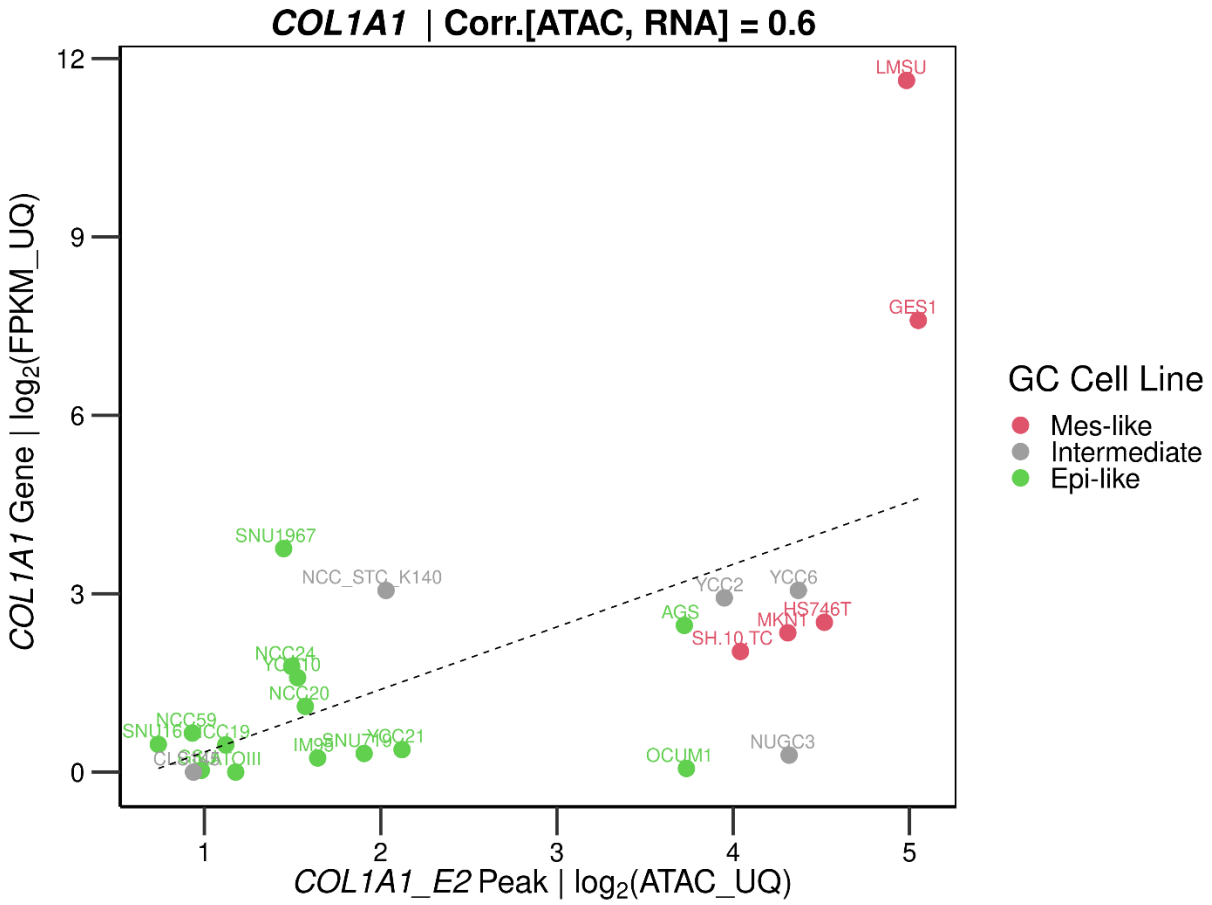
Cancer and NAT cells spread uniformly across the 6 single-cell groups detected by clustering over UMAP.

Supplemental Fig. S21) Correlation of COL1A1 Enhancer (*E1*) Accessibility and *COL1A1* Gene Expression



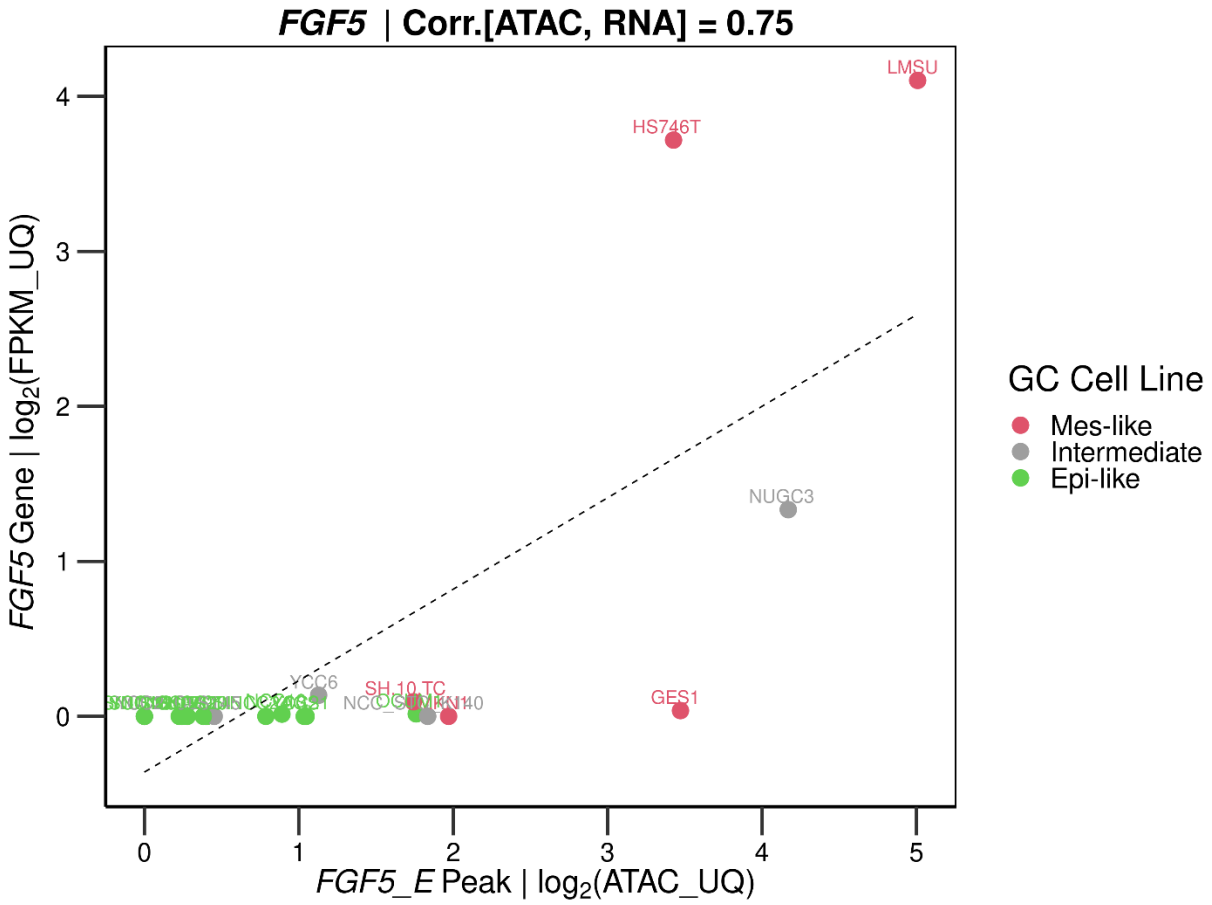
Pearson's correlation of [*COL1A1* RNA, *COL1A1_E1* ATAC] signals. ATAC-seq and RNA-seq values are upper-quartile normalized.

Supplemental Fig. S22) Correlation of COL1A1 Enhancer (*E2*) Accessibility and *COL1A1* Gene Expression



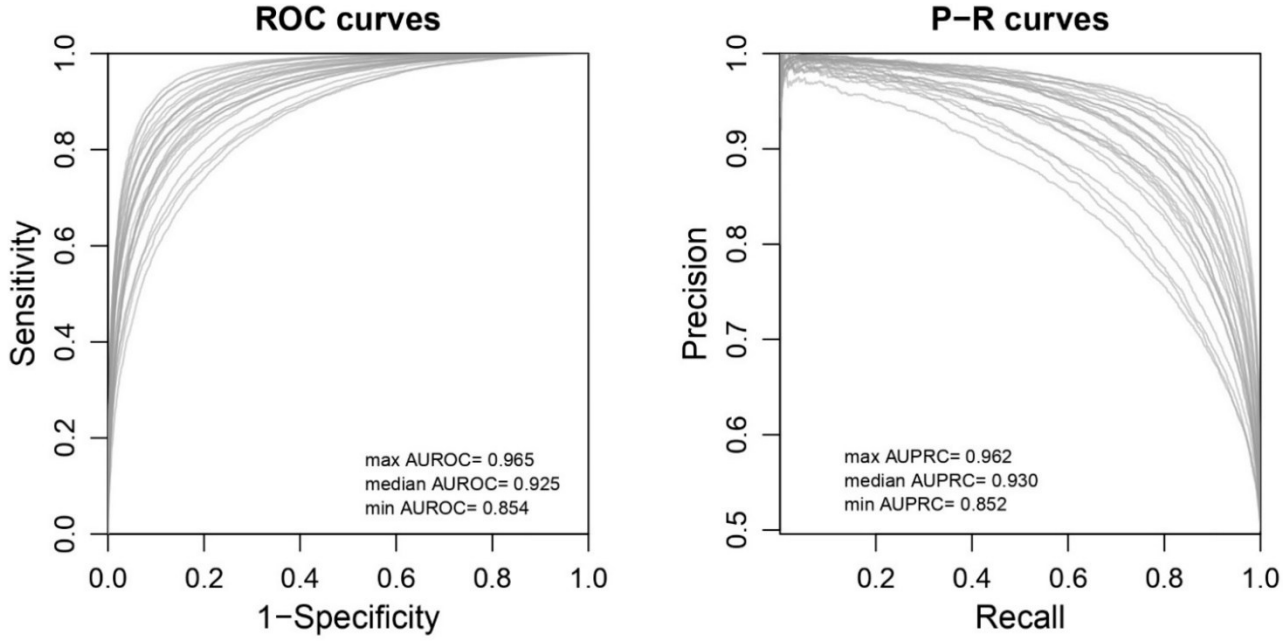
Pearson's correlation of [*COL1A1* RNA, *COL1A1_E2* ATAC] signals. ATAC-seq and RNA-seq values are upper-quartile normalized.

Supplemental Fig. S23) Correlation of FGF5 Enhancer (*E*) Accessibility and *FGF5* Gene Expression



Pearson's correlation of [*FGF5* RNA, *FGF5_E* ATAC] signals. ATAC-seq and RNA-seq values are upper-quartile normalized.

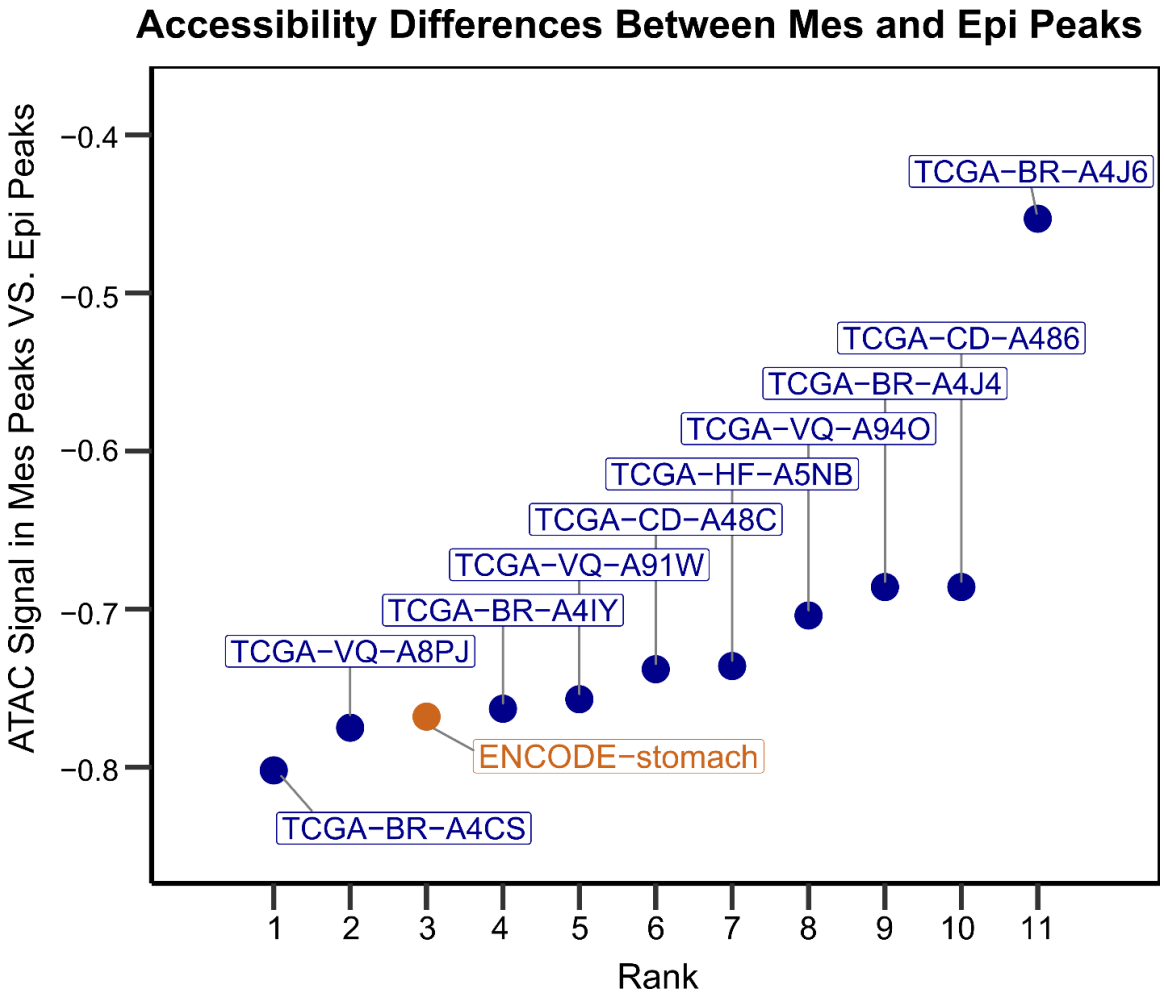
Supplemental Fig. S24) ROC Curves – GC cell lines VS. random GC-matched Genomic Regions



gkm-SVM models were trained on the top10k distal ATAC peaks of each cell line (n=25) vs. 10k random GC-matched genomic regions

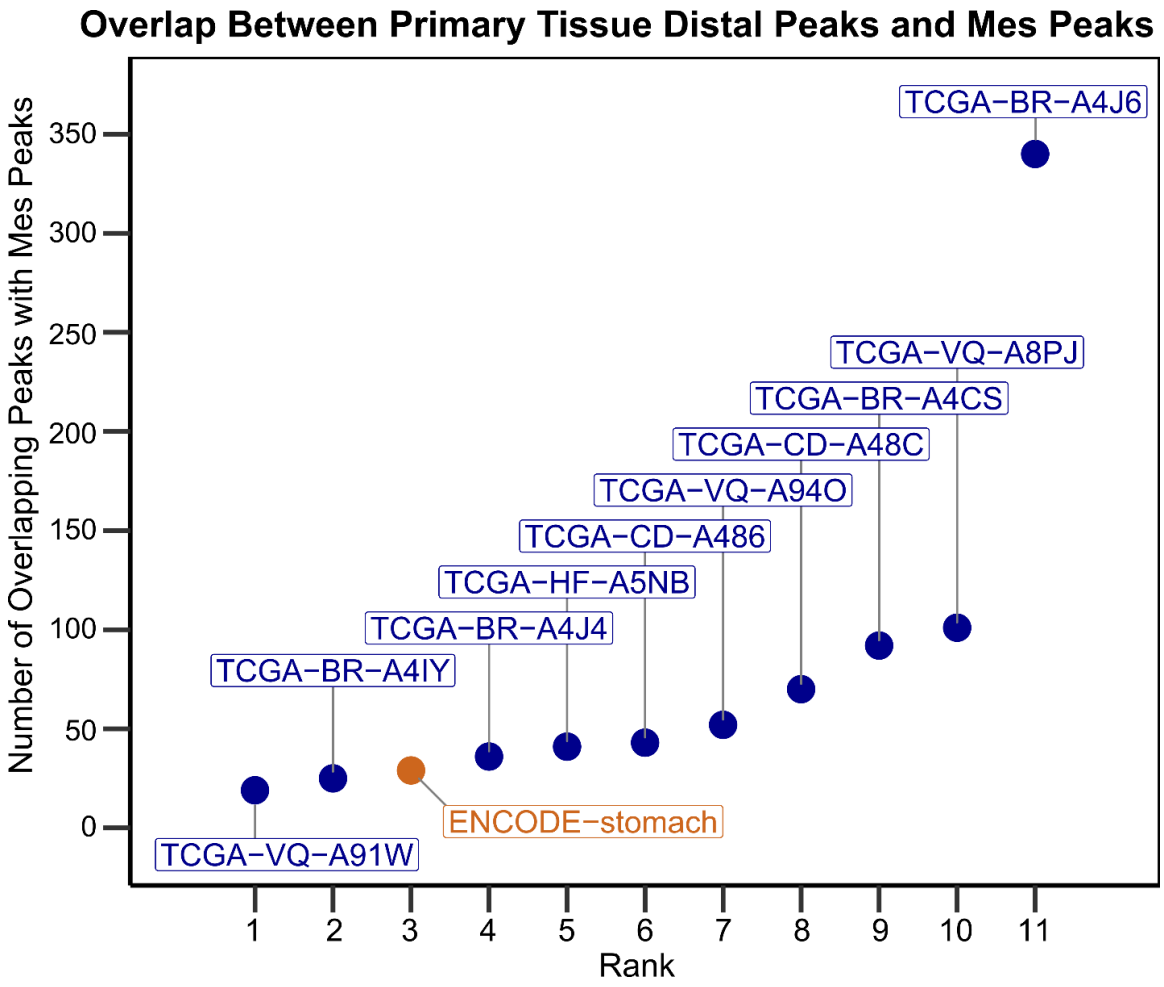
Peak length = 300bp

Supplemental Fig. S25) ATAC-seq Signal of Mes vs. Epi Peaks in TCGA-STAD



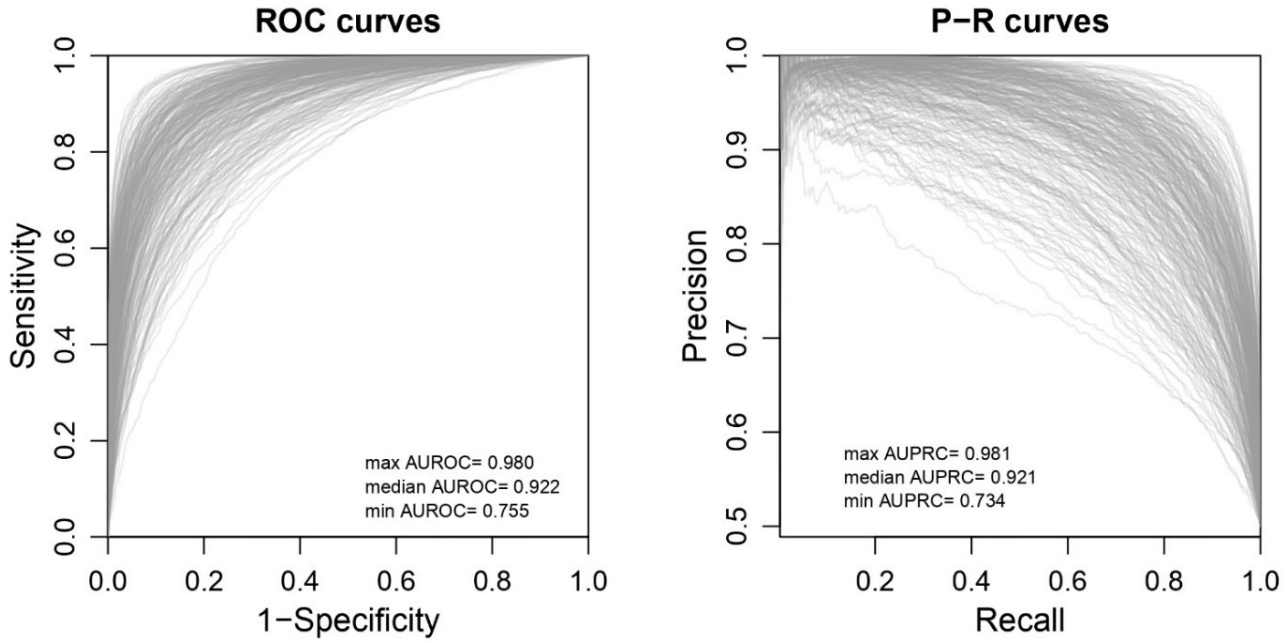
The Y-axis shows the ATAC signal values calculated over the 2000 Mes-high and 2000 Epi-high peaks as follows: (Mes - Epi) / (Mes + Epi), which is between [-1, +1].

Supplemental Fig. S26) Number of Overlaps between Top TCGA-STAD ATAC Peaks and Mes-high Peaks



Y-axis shows the number of overlapping peaks between the top ATAC peaks in TCGA-STAD samples and the 2000 Mes-high peaks.

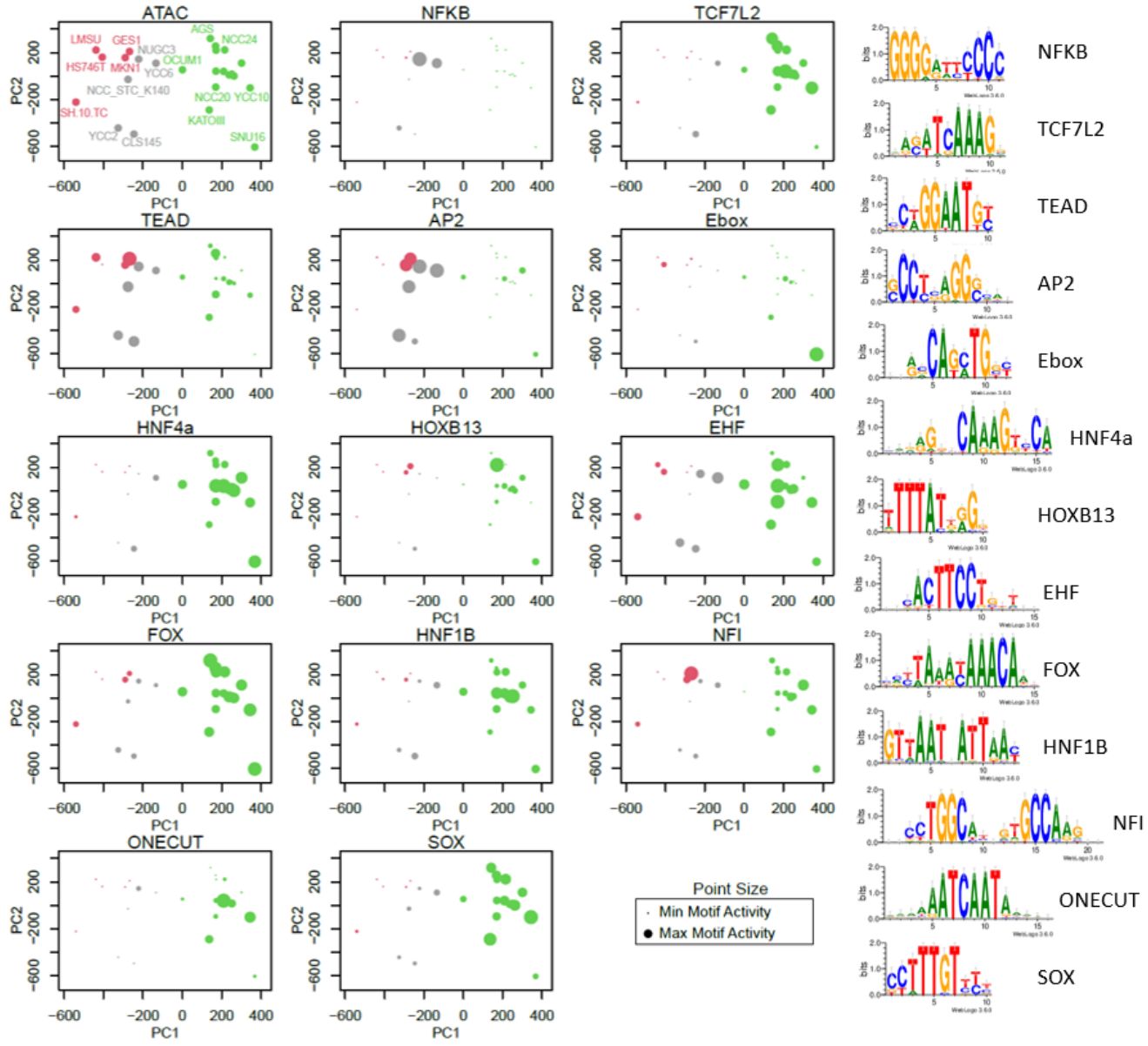
Supplemental Fig. S27) ROC Curves – Pairs of GC Cell Lines Trained Against Each Other



gkm-SVM models were trained on the most differentially accessible distal ATAC peaks (n=2000 positive peaks and n=2000 negative peaks) of each pair of GC cell lines (n=25) vs. all other GC cell lines (300 pairs = (25 * 24) / 2)

Peak length = 300bp

Supplemental Fig. S28) PCA of ATAC-seq | TF Motif Activity



gkm-SVM inferred activity (dot size) of TFs not shown in Fig 1E, across all samples.