

SUPPLEMENTAL TABLES

Table S1. Alignments of sequencing reads. For RNA-seq (indicated by *), the format is: number of reads uniquely aligned to reference genome (percentage of total reads) + number of reads uniquely aligned to splicing junction library (percentage of reads fail to be aligned to reference genome). Double asterisk (**) indicates whether the reads were extended to 200 bp to generate tag density data. 200 bp is the average length of DNA segments sent to sequencing. Triple asterisk (***) indicated column contains GEO accession of the sequencing data downloaded from Gene Expression Omnibus with the prefix “GSM”, or ArrayExpress accession of the data submitted to ArrayExpress with the prefix “E-MTAB-”.

Cell Type	Sequencing	Number of Uniquely Aligned Reads (Percentage of Total Reads Aligned)*	Read Length	Extended to 200 bp **	Accession # ***
IMR90	GRO-seq	12,545,400 (48.70%)	35 bp	Yes	E-MTAB-742
		4,106,991 (37.95%)			
IMR90	H3K36me3 ChIP-seq	6,769,706 (43.45%)	36 bp	Yes	GSM521890
		6,148,696 (41.70%)			GSM521892
		10,720,642 (42.51%)			
		2,906,837 (49.25%)			
IMR90	RNAP ChIP-seq	22,998,721 (63.12%)	35 bp	Yes	E-MTAB-744
		2,142,977(41.99%) + 281,369 (15.04%)	43 bp		
IMR90	RNA-seq	3,166,967 (32.32%) + 515,432 (13.14%)	87 bp	No	GSM438363
		3,213,340 (32.32%)+523,428 (13.10%)			
		3,222,821 (32.55%)+523,856 (13.23%)			
IMR90	H4K20me1 ChIP-seq	6,336,288 (37.08%)	36 bp	Yes	GMS521915
IMR90	H3K9me3 ChIP-seq	3,128,614 (29.30%)	36 bp	Yes	GSM469974
MCF10A	GRO-seq	10,699,337 (44.72%)	35 bp	Yes	E-MTAB-742
		11,097,705 (42.65%)			
MCF10A	H3K36me3 ChIP-seq	21,220,856 (57.66%)	35 bp	Yes	E-MTAB-744
MCF10A	RNAP ChIP-seq	24,015,184 (63.83%)	35 bp	Yes	E-MTAB-744
MCF10A	RNA-seq	21,879,681 (55.90%)+7,648,486 (44.31%)	74 bp	No	E-MTAB-743
HIESC	GRO-seq	13,805,121 (59.39%)	34 bp	Yes	E-MTAB-742
		6,621,614 (53.14%)			
		4,424,186 (50.69%)			
HIESC	H3K36me3 ChIP-seq	6,871,013 (46.05%)	36 bp	Yes	GSM409312
		6,034,448 (52.62%)			GSM466737
		6,072,463 (50.21%)			
		6,353,061 (45.19%)			
		2,149,423 (43.50%)+342,142 (12.26%)	43 bp	No	
		4,164,030 (36.93%)+1,616,314 (22.73%)			
HIESC	RNA-seq	4,219,820 (36.44%)+1,645,731 (22.36%)	87 bp	No	GSM438361
		4,241,476 (36.60%)+1,649,529 (22.45%)	87 bp		
		4,265,070 (36.48%)+1,663,825 (22.40%)			

Table S2. mRNA-seq RPKM (CDS-aligned reads per kilobase per million unique mapped reads) results for splicing factor genes.* RPKM of hnRNPH/F was calculated as the average of RPKMs of hnRNPH1, hnRNPH2, hnRNPH3 and hnRNPF.

Splicing factor	mRNA-seq RPKM		
	IMR90	H1ESC	MCF10A
SF2/ASF	729.4	997.5	637.1
9G8	154.9	529.7	179.3
SC35	219.3	300.3	752.3
Tra2alpha	63.0	83.5	44.0
Tra2beta	57.4	182.4	102.3
SRp20	123.2	386.2	241.8
SRp40	310.3	1415.9	401.9
SRp55	291.8	344.0	355.3
hnRNPA1	111.6	4269.9	3121.1
hnRNPA2	251.7	695.6	462.1
hnRNPH/F*	150.5	222.4	281.8
MBNL	8.8	2.1	12.9
NOVA1	0.2	1.8	0
PTB	88.1	134.6	233.4
CUG-BP	33.1	158.6	51.9
YB1	37.6	360.6	195.9
FOX1	0.01	0.2	0.001

Table S3. De novo motifs related with cell type-specific GRO-seq patterns. We compared occurrences of all possible 8 bp, 7 bp and 6 bp motifs, i.e. exhaustive enumeration of 8-mers, 7-mers and 6-mers conformations, between samples of cell type-specific GRO-seq patterns Wilcoxon rank sum test. Significant motifs between IMR90 vs H1ESC were further split into clusters and their best matches in TRANSFAC (M prefix) or JASPAR (MA prefix) are shown in the tables together with p values of the matches. Motifs possibly act as accelerators of RNAP elongation in H1ESC and decelerators in IMR90. (Alternative hypothesis (H_a) of Wilcoxon rank sum test: motif occurrence in IMR90 > motif occurrence in H1ESC).

Region	Motif Length	Number of Significant Motifs	Familial Profile of Motif Cluster	Best Match in TRANSFAC/JASPAR (Homo sapiens)		Gene/Element	P value	
				Logo	ID			
exon	6 bp	46			M00671	TCF-4	5.115e-04	
intron	7 bp	54			M00941	MEF-2	4.151e-04	
	6 bp	284			MA0124.1	NKX3-1	3.986e-03	
					MA0083.1	SRF	1.763e-03	
					MA0037.1	GATA3	6.186e-05	
					M00671	TCF-4	1.228e-03	
					M00451	NKX3A	1.144e-04	
					M00471	TBP	2.001e-04	
					M00499	STAT5A	2.040e-03	
		M00302	NF-AT	2.469e-04				

Table S4. De novo motifs related with cell type-specific GRO-seq patterns. We compared occurrences of all possible 8 bp, 7 bp and 6 bp motifs, i.e. exhaustive enumeration of 8-mers, 7-mers and 6-mers conformations, between samples of cell type-specific GRO-seq patterns Wilcoxon rank sum test. Significant motifs between IMR90 vs H1ESC were further split into clusters and their best matches in TRANSFAC (M prefix) or JASPAR (MA prefix) are shown in the tables together with p values of the matches. Motifs possibly act as accelerators of RNAP elongation in IMR90 and decelerators in ESC H1. (Alternative hypothesis (H_a) of Wilcoxon rank sum test: motif occurrence in IMR90 < motif occurrence in H1ESC)

Region	Motif Length	Number of Significant Motifs	Familial Profile of Motif Cluster	Best Match in TRANSFAC/JASPAR (Homo sapiens)		Gene/Element	P value
				Logo	ID		
exon	6 bp	32			M00979	PAX6	1.503e-05
					MA0163.1	PLAG1	2.532e-04
					M00644	LBP-1	1.114e-05
intron	6 bp	292			M00304	Dde box	4.005e-05
					MA0095.1	YY1	3.047e-04
					M01037	GLI	4.613e-04
					M00649	MAZ	1.649e-04
					M00712	myogenin	2.953e-04
					M00008	Sp1	7.844e-05
					M00378	Pax-4	1.218e-04
					M00727	SF-1	2.863e-04
					M00720	CAC-binding	1.203e-03

Table S5. De novo motifs related with cell type-specific GRO-seq patterns. We compared occurrences of all possible 8 bp, 7 bp and 6 bp motifs, i.e. exhaustive enumeration of 8-mers, 7-mers and 6-mers conformations, between samples of cell type-specific GRO-seq patterns (same samples as in Figure 4) using Wilcoxon rank sum test. Significant motifs between MCF10A vs H1ESC were further split into clusters and their best matches in TRANSFAC (M prefix) or JASPAR (MA prefix) are shown in the tables together with p values of the matches. Motifs possibly act as accelerators of RNAP elongation in ESC H1 and decelerators in MCF10A. (Alternative hypothesis (H_a) of Wilcoxon rank sum test: motif occurrence in MCF10A > motif occurrence in H1ESC)

Region	Motif Length	Number of Significant Motifs	Familial Profile of Motif Cluster	Best Match in TRANSFAC/JASPAR (Homo sapiens)		Gene/Element	P value
				Logo	ID		
exon	6 bp	5			M00499	STAT5A	4.937e-03
intron	8 bp	3			M00006	MEF-2	1.416e-04
	7 bp	74			M00405	MEF-2	4.598e-05
	6 bp	327			M00624	DBP	3.924e-03
					M00292	Freac-4	9.155e-04
					M00495	Bach1	5.623e-03
					MA0033.1	FOXL1	1.622e-03
					M00921	GR	1.868e-03
					MA0084.1	SRY	2.874e-05
					M00941	MEF-2	4.512e-03
					M00149	SBF-1	1.313e-03
					M00462	GATA-6	2.078e-04
			M00135	Oct-1	1.657e-04		
		M01092	TCF	7.558e-04			

Table S6. De novo motifs related with cell type-specific GRO-seq patterns. We compared occurrences of all possible 8 bp, 7 bp and 6 bp motifs, i.e. exhaustive enumeration of 8-mers, 7-mers and 6-mers conformations, between samples of cell type-specific GRO-seq patterns (same samples as in Figure 4) using Wilcoxon rank sum test. Significant motifs between MCF10A vs H1ESC were further split into clusters and their best matches in TRANSFAC (M prefix) or JASPAR (MA prefix) are shown in the tables together with p values of the matches. Motifs possibly act as accelerators of RNAP elongation in MCF10A and decelerators in H1ESC. (Alternative hypothesis (H_a) of Wilcoxon rank sum test: motif occurrence in MCF10A < motif occurrence in H1ESC)

Region	Motif Length	Number of Significant Motifs	Familial Profile of Motif Cluster	Best Match in TRANSFAC/JASPAR (Homo sapiens)		Gene/Element	P value
				Logo	ID		
exon	6 bp	25			M00176	AP-4	2.578e-04
					M00005	AP-4	2.057e-04
intron	6 bp	330			M00468	AP-2rep	1.358e-3
					MA0138.1	REST	5.900e-04
					M00319	MEF-3	2.415e-03
					M00641	HSF	5.000e-03
					MA0080.1	SPI1	2.429e-04
					M00033	p300	1.196e-03
					M00993	TAL1	1.531e-06
					M00085	ZID	2.467e-03
					M00002	E47	6.228e-04
					M00518	RXR-alpha	9.954e-05
					MA0163.1	PLAG1	1.381e-04
					MA0079.2	SP1	7.333e-05
					M00055	N-Myc	3.420e-03