

Supplementary Information

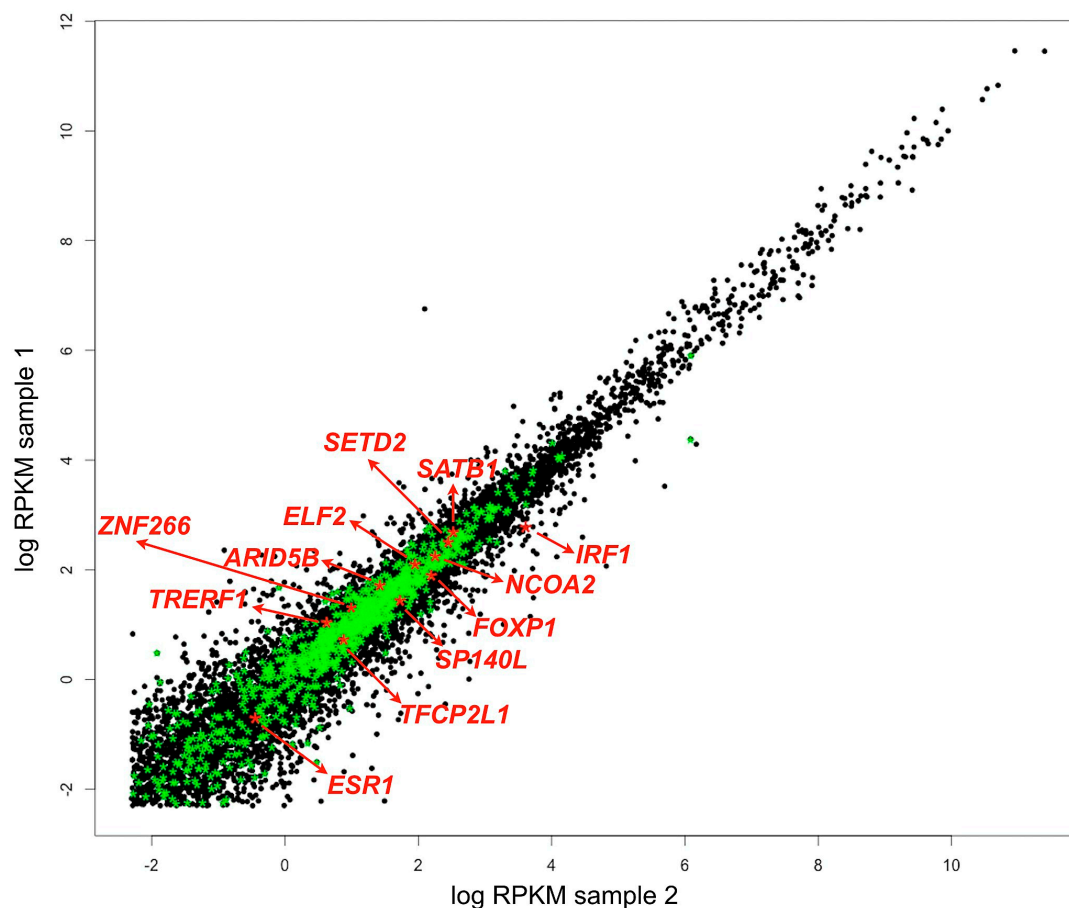


Figure S1. Scatter plot of log-transformed gene expression levels between two samples. Genes are indicated with black dots. Green stars represent transcription factors encoding genes and red stars indicate genes with evidence of alternative splicing events.

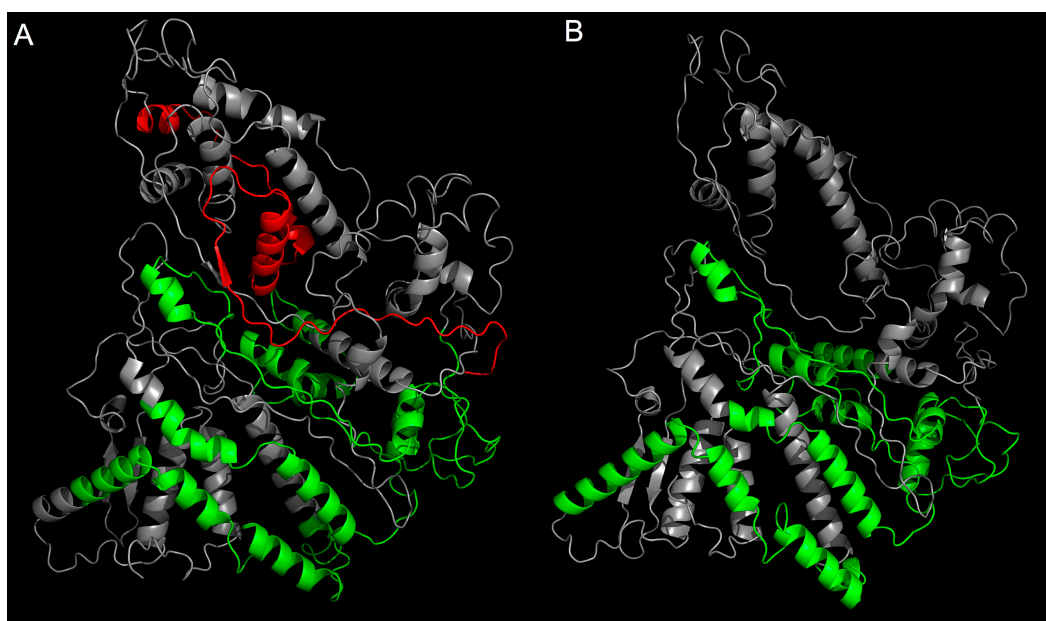


Figure S2. Tridimensional structures of annotated (panel A) and novel (panel B) SATB1 isoforms are shown as grey cartoons. Nuclear localization signal and DNA binding domain are highlighted in red and green, respectively.

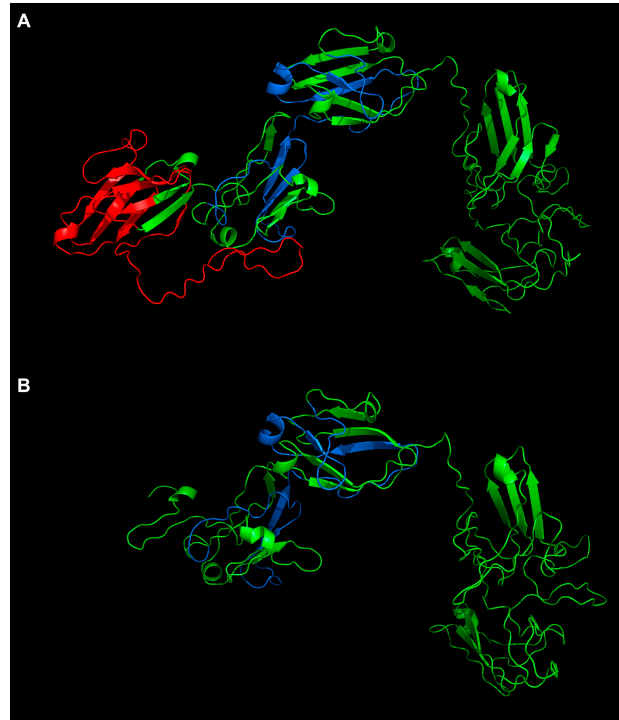


Figure S3. Tridimensional structures of annotated (panel **A**) and new (panel **B**) ELF2 isoforms are shown as green cartoons. *N*-terminal domain and the ETS-like domain are highlighted in red and blue, respectively.

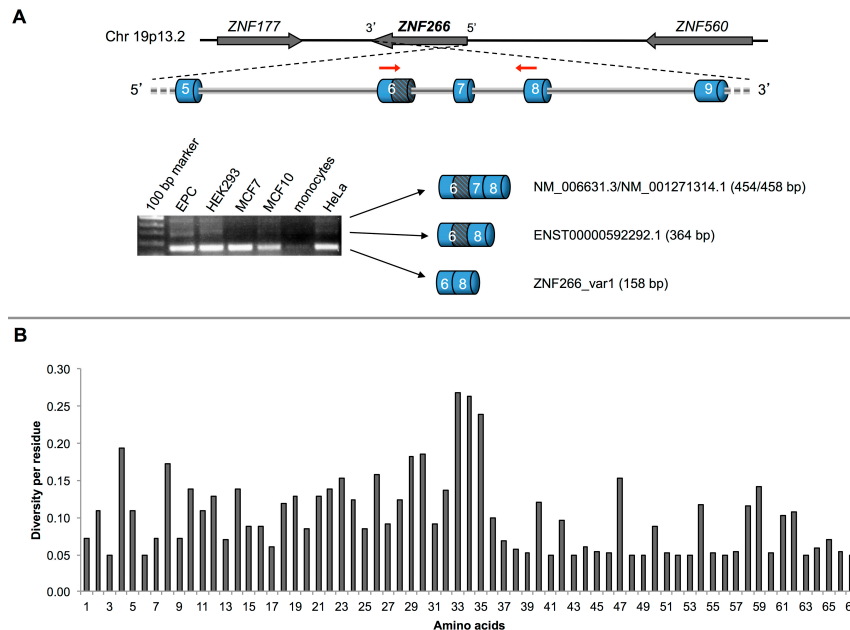


Figure S4. Panel **A**: Graphical representation of *ZNF266* transcripts. The genomic region encompassing *ZNF266* gene and the exon/intron structure are shown in the upper part. Red arrows indicate the primers annealing sites. In the lower part, analysis of PCR product by agarose gel electrophoresis in six cell lines/primary cells indicates the presence of three amplicons, schematized in the right part. The lowest PCR product correspond to the new identified *ZNF266* transcript; Panel **B**: Histogram of diversity per residue rate, relative to the additional 67 amino acids in the new *ZNF266* isoform.

Table S1. Expression pattern analysis of newly identified TFs transcripts. Relative expression of newly identified transcripts is schematized as ++ (present at high level), + (present), – (absent); “Not expressed” indicates that the corresponding gene is not expressed. EPC: Endothelial Progenitor Cells; HEK293: Human Embryonic Kidney-293 cell line; MCF7: Michigan Cancer Foundation-7 cell line; MCF10: Michigan Cancer Foundation-10 cell line; HeLa: Henrietta Lacks cell line.

Transcript	EPC	HEK293	MCF7	MCF10	HeLa	Monocytes
<i>ZNF266_var1</i>	++	++	++	++	++	+
<i>SATB1_var1</i>	+	++	–	Not expressed	+	Not expressed
<i>SP140L_var1</i>	++	+	+	+	+	+
<i>ARID5B_var1</i>	++	++	+	+	++	–
<i>NCOA2_var1</i>	+	+	+	+	+	Not expressed
<i>ELF2_var1</i>	+	+	+	–	–	–
<i>ELF2_var2</i>	++	+	+	–	–	–
<i>IRF1_var1</i>	+	+	+	+	+	+
<i>IRF1_var2</i>	+	+	+	+	+	+

Table S2. Evolutionary conservation of newly identified TFs transcripts. EST: Expressed Sequence Tag.

Transcript	Accession Number	Organism	Type
<i>ZNF266_var1</i>	JV705547	Macaca	mRNA from GenBank
	JV640626	Macaca	mRNA from GenBank
	AK305581	Pan	mRNA from GenBank
	HX641224	Callitrix	EST
	HX654921	Callitrix	EST
	EB529880	Macaca	EST
	BB625302	Mus	EST
	BY092133	Mus	EST
	BY133758	Mus	EST
	BY726782	Mus	EST
	CD546656	Mus	EST
	CN697858	Mus	EST
	CN702205	Mus	EST
	CX565991	Mus	EST
	CB605822	Rattus	EST
	FN800703	Rattus	EST
<i>SATB1_var1</i>	FS646931	Sus	EST
<i>SP140L_var1</i>	HX457795	Callitrix	EST
	HX540006	Callitrix	EST
	HX541471	Callitrix	EST
	HX547835	Callitrix	EST
	HX564060	Callitrix	EST
	HX564780	Callitrix	EST
	HX457795	Callitrix	EST
	BM423272	Macaca	EST
	DK580521	Macaca	EST
	DC853200	Macaca	EST
	CB553659	Macaca	EST

Table S2. Cont.

Transcript	Accession Number	Organism	Type
<i>ARID5B</i>_var1	/	/	/
<i>NCOA2</i>_var1	/	/	/
<i>ELF2</i>_var1	JR403486	Camelus	mRNA from GenBank
<i>ELF2</i>_var2	JU727093	Bubalus	mRNA from GenBank
	JO618916	Capra	mRNA from GenBank
	JT722928	Capra	mRNA from GenBank
	BC027739	Mus	mRNA from GenBank
	AK200215	Mus	mRNA from GenBank
	FE023941	Ovis	EST
<i>IRF1</i>_var1	KA619585	Bubalus	mRNA from GenBank
	JT619617	Capra	mRNA from GenBank
<i>IRF1</i>_var2	—	—	—