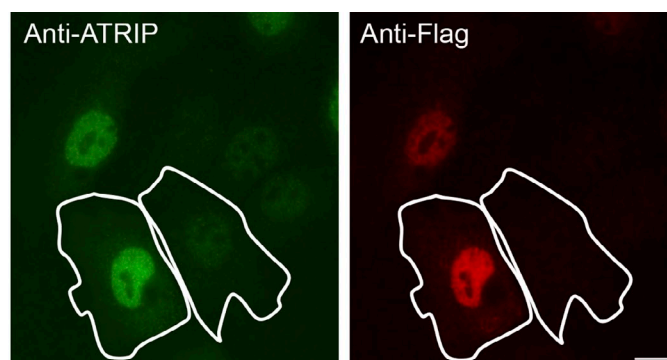


Fong et al., <http://www.jcb.org/cgi/content/full/jcb.201303145/DC1>

$$\text{Level of overexpression} = \frac{(\text{Fluorescence intensity}_{\text{transfected cell}} / \text{Area} - \text{Fluorescence intensity}_{\text{background}} / \text{Area})}{(\text{Fluorescence intensity}_{\text{untransfected cell}} / \text{Area} - \text{Fluorescence intensity}_{\text{background}} / \text{Area})}$$

ORFeome	Name	Level of overexpression	Localization of exogenous protein	Localization of endogenous protein
11006@a08	SNAI1	2.24±0.21	nucleoplasm	nucleoplasm
11006@e07	SMN2	2.93±0.37	nuclear bodies, cytoplasm	nuclear bodies, cytoplasm
11011@f11	TUBA1B	0.97±0.37	cytoskeleton	cytoskeleton
11013@g02	PRC1	2.49±0.32	cytoskeleton	cytoskeleton
11024@h06	MCM6	1.23±0.14	cytoplasm, nucleoplasm	nucleoplasm
11026@e03	SNRPD1	1.4±0.5	nucleoplasm, nucleolus	nucleoplasm, nuclear bodies
11027@a02	H3F3B	1.24±0.24	nucleoplasm	nucleoplasm
11030@g03	PCNA	3.38±1.36	cytoplasm	cytoplasm
11030@h05	RPA2	1.56±0.25	cytoplasm, nucleoplasm	nucleoplasm
11032@e04	BM11	2.15±0.43	nuclear bodies, nucleoplasm	nuclear bodies, nucleoplasm
11037@a09	NASP	1.35±0.13	nucleoplasm	nucleoplasm
11037@b03	TUBG1	4.22±0.69	cytoplasm	cytoplasm
11037@f04	AKT1	1.89±0.49	cytoplasm	cytoplasm
11052@a02	CDK2	1.08±0.08	cytoplasm, nucleoplasm	nucleoplasm
11056@h09	KDHRDS2	0.97±0.37	nuclear bodies, nucleoplasm	nuclear bodies, nucleoplasm
11065@a10	NONO	1.23±0.35	nuclear bodies, nucleoplasm	nuclear bodies, nucleoplasm
11067@e06	TOE1	1.08±0.17	nuclear bodies, nucleoplasm	nuclear bodies, nucleoplasm
11069@g07	COIL	3.03±0.14	nuclear bodies	nuclear bodies
11072@b03	DVL2	1.55±0.15	nucleoplasm	nucleoplasm
11072@f04	PML	1.2±0.03	nuclear bodies	nuclear bodies
11078@b04	WDR91	1.52±0.45	cytoplasm	cytoplasm
11085@g11	GAPDH	1.56±0.47	cytoplasm	cytoplasmic aggregates
11087@b08	PLK1	4.3±1.3	cytoplasm	cytoplasm
11089@f02	OCLN	4.08±1.01	cytoplasm, cell periphery	cytoplasm, cell periphery
11089@g03	ACD	2.06±0.16	cytoplasm, nucleoplasm	cytoplasm, nucleoplasm
31011@d04	SMARCA1	2.38±0.55	nucleoplasm	nucleoplasm
31037@f11	FANCD2	2.7±0.12	nucleoplasm	nucleoplasm
31038@a12	DKC1	3.25±0.25	nucleolus, nucleoplasm	cytoplasm, nucleoplasm
31042@g10	FBL	3.63±0.77	nucleolus	nucleolus
31043@g04	AURKB	5.49±1.2	cytoplasm, nucleoplasm	nucleoplasm
31044@c07	ATRIP	4.69±0.52	nucleoplasm	nucleoplasm
51008@e04	LAMC2	1.44±0.32	cytoplasm	cytoplasm
51016@b06	DYNLL1	3.67±0.87	cytoplasm	cytoplasm
51020@c05	RBBP7	2.71±0.71	cytoplasm, nucleoplasm	cytoplasm, nucleoplasm
51022@f02	TCF4	2.24±0.85	nucleoplasm	nucleoplasm
51024@d02	RASAL2	1.82±0.32	cytoplasm	cytoplasm

Figure S1. **The level of overexpression of the ORFeome library in our study.** Related to Fig. 1. A table shows the name of selected proteins from the ORFeome library with level of overexpression, localization of tagged protein, and localization of endogenous protein. Bar, 10 μ m.

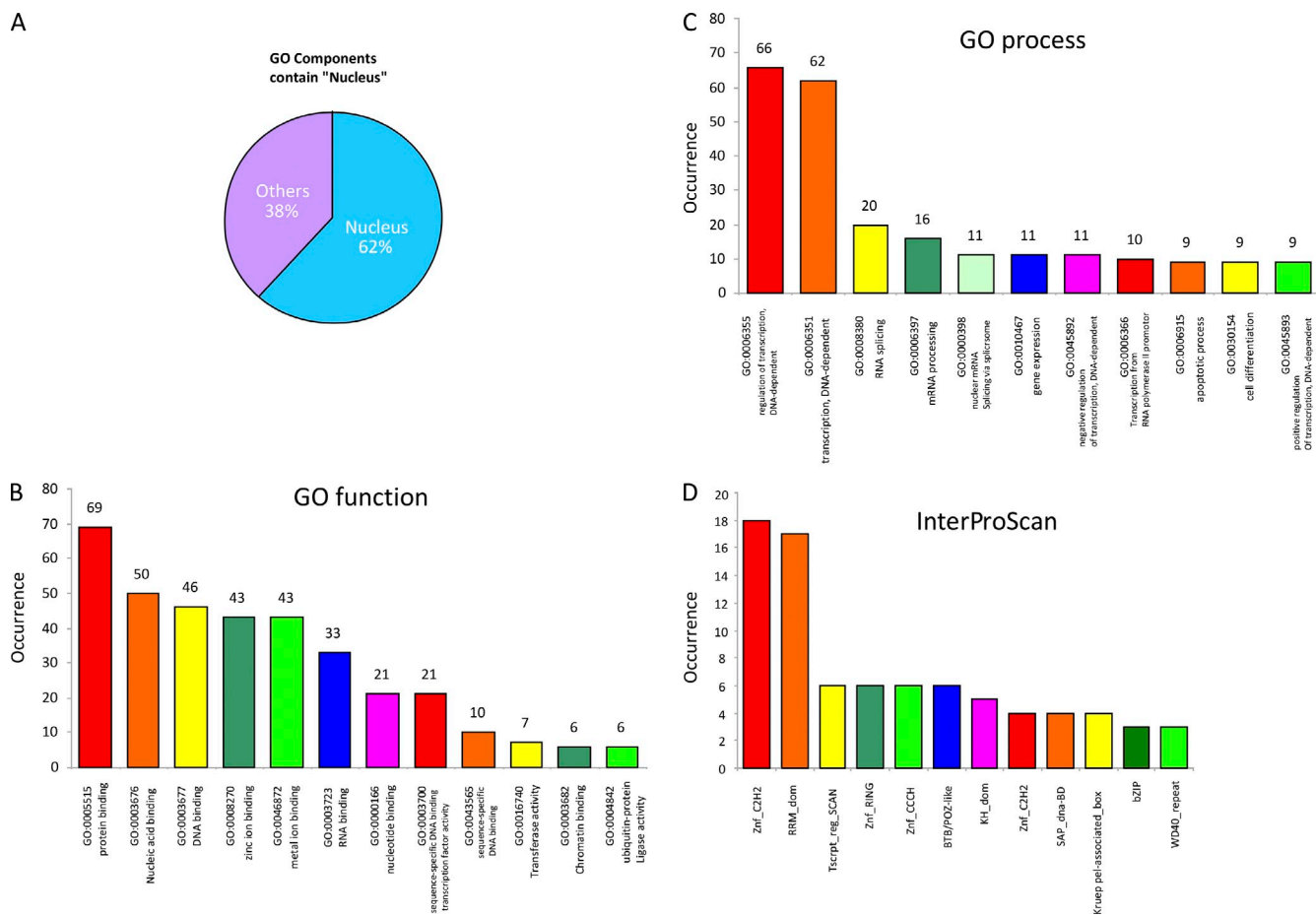


Figure S2. **Bioinformatics analysis of nuclear foci proteome.** (A) Percentage of proteins annotated with the GO component nucleus (please also see Table S1). (B) The bar graph shows the number of proteins in nuclear foci proteome that are annotated with each GO function (please also see Table S1). (C) The bar graph shows the number of proteins in nuclear foci proteome that are annotated with each GO process (please also see Table S1). (D) The bar graph shows the number of proteins that are predicted to contain the indicated motifs by InterProScan (please also see Table S5 and Table S6).

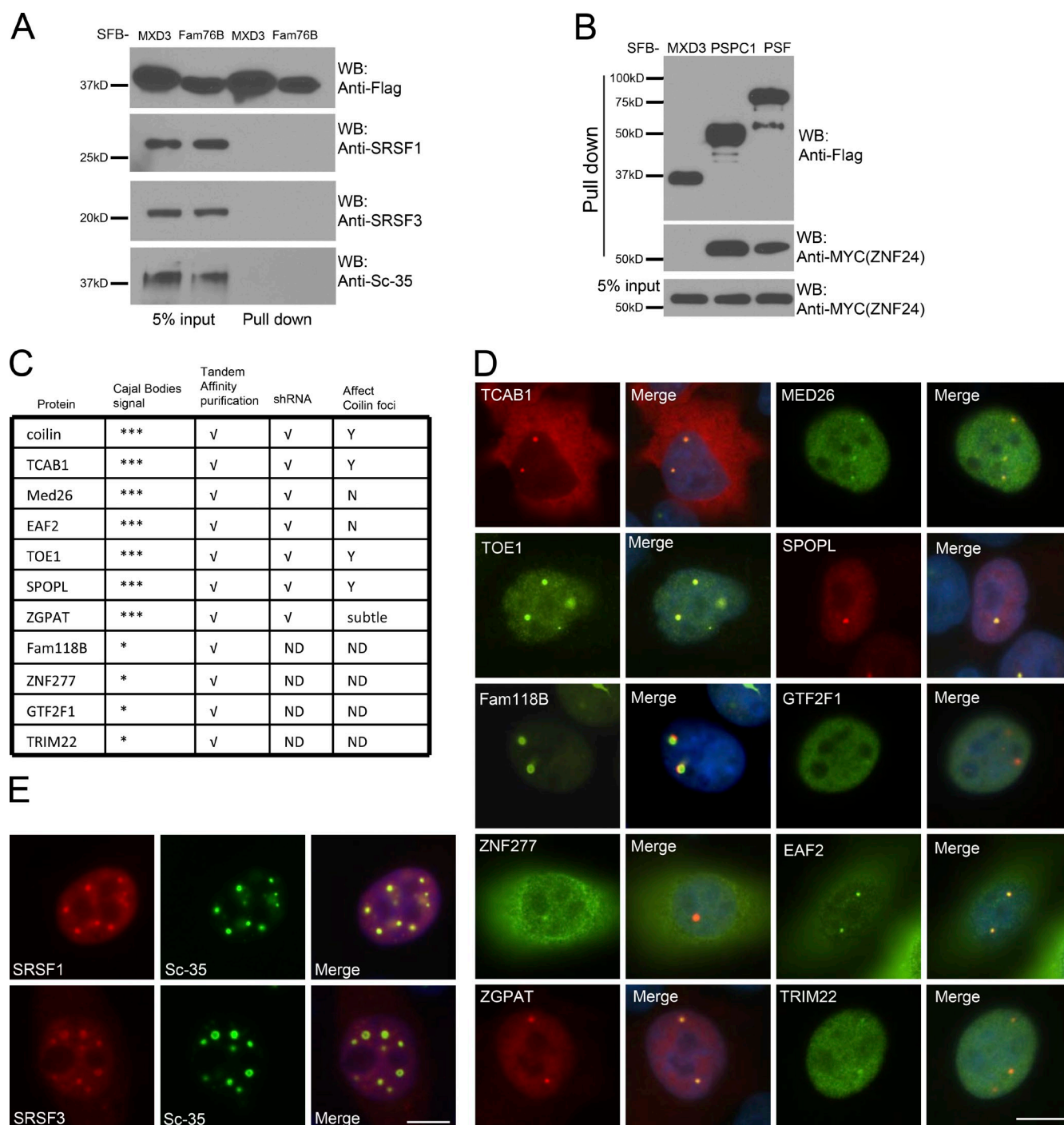
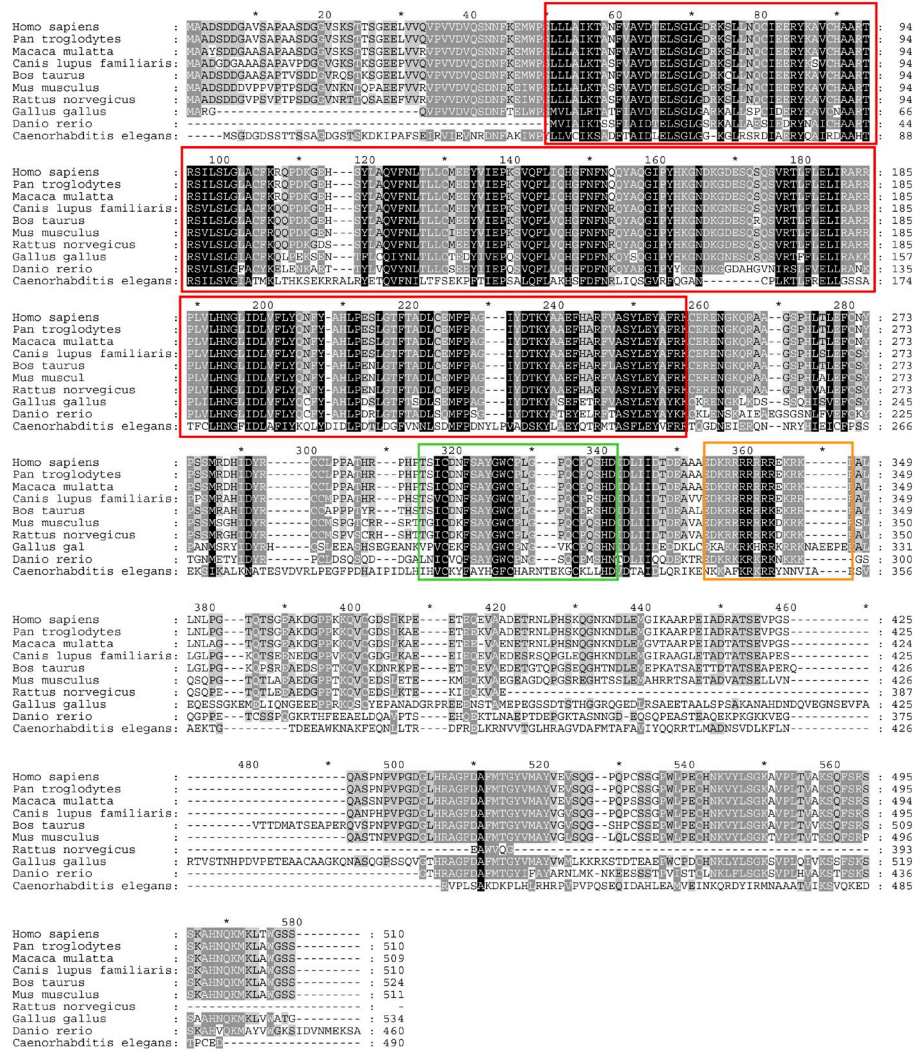
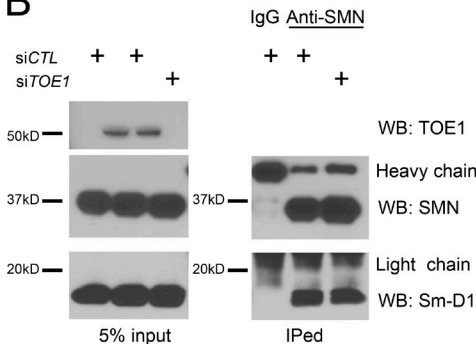


Figure S3. **Proteomic analysis of various nuclear subcompartments.** Related to Fig. 3 and Fig. 4. (A) Fam76B does not associate with the serine/arginine (SR)-rich family of pre-mRNA splicing factors. (B) ZNF24 interacts with core paraspeckle components. (C) A list shows the proteins that were subjected to proteomic analysis and shRNA-based phenotypic screen. Triple asterisks indicate stronger Cajal body signal, whereas the single asterisks indicate weaker Cajal body signal. (D) Localization of novel protein components of Cajal bodies. (E) TOE1 interactors (SRSF1 and SRSF3) localize to nuclear speckles. WB, Western blot. Bars, 10 μ m.

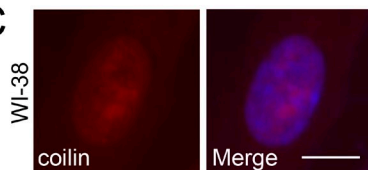
A



B



C



D

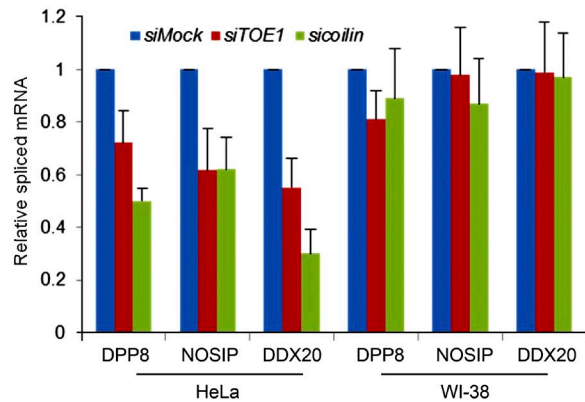


Figure S4. **TOE1 is required for endogenous mRNA splicing.** Related to Fig. 5 and Fig. 6. (A) Comparison of TOE1 amino acid sequences from various species. The N terminus region, the zinc finger domain, and NLS of TOE1 are indicated in red, green, and orange color, respectively. Letters in gray indicate the amino acids with same or similar properties shared and conserved in most of the species. Letters in black indicate the amino acids with same or similar properties conserved in all the species. Asterisks in mark every 20 amino acids. (B) TOE1 is not required for SMN binding to Sm-D1. WB, Western blot. (C) WI-38 cells do not contain Cajal bodies. Bars, 10 μ m. (D) Abundance of spliced mRNA for three endogenous genes was quantitated by qRT-PCR using primers spanning a particular splice junction: exons 4 and 5 for *DPP8*, exons 7 and 8 for *NOSIP*, and exons 2 and 3 for *DDX20*. Levels (\pm SD) of splicing efficiency are given relative to control (mock shRNA-treated cells) and normalized to 5S RNA ($n = 3$ independent experiments).

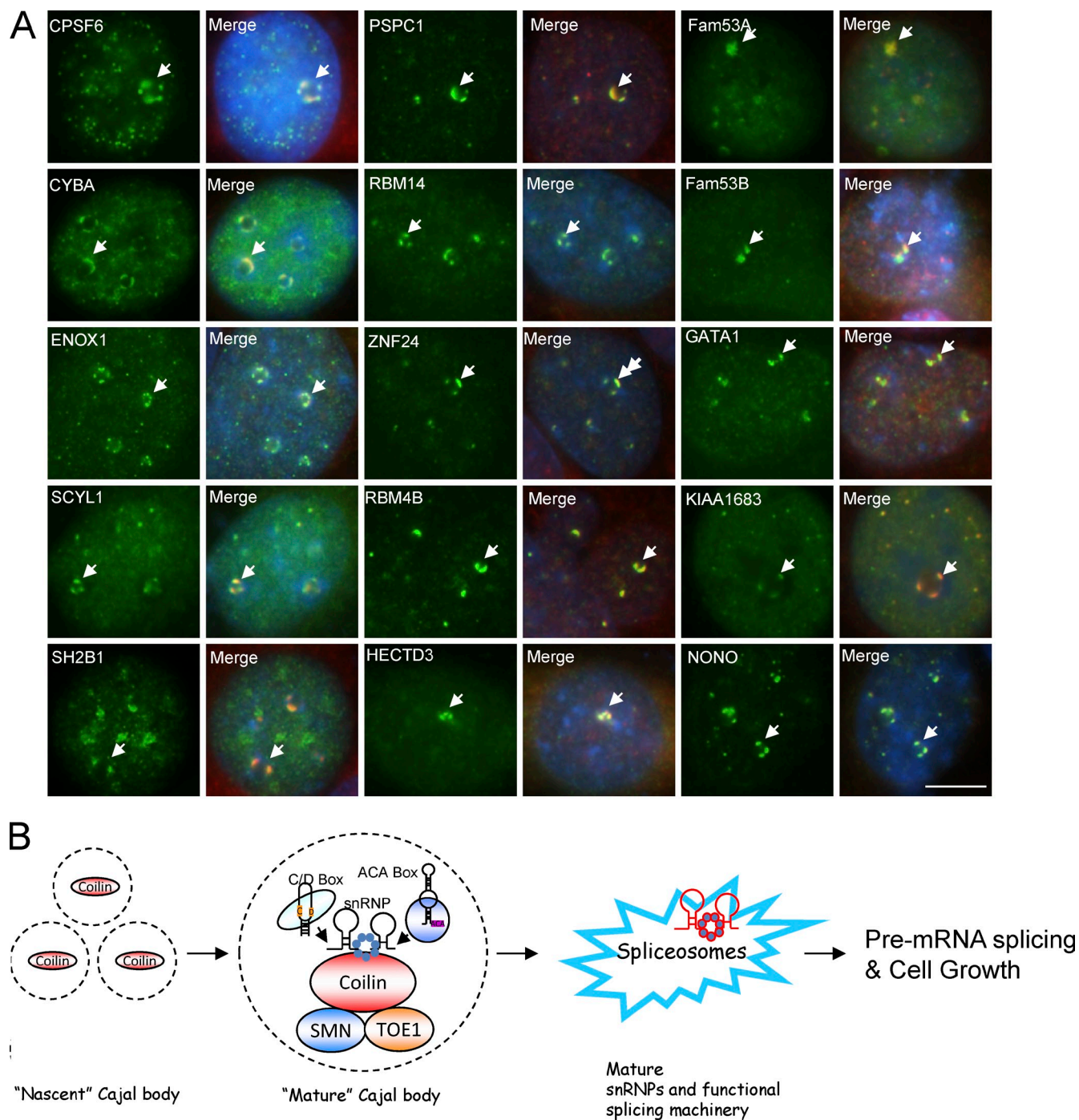


Figure S5. **Validation of identified paraspeckle proteins.** Related to Fig. 7 and Discussion. (A) Newly identified paraspeckle components showed relocation to perinucleolar caps upon actinomycin D treatment. Arrows indicate the position of paraspeckle-localized proteins after actinomycin D treatment. Bars, 10 μ m. (B) Proposed working model illustrates that TOE1 as an essential component for Cajal body assembly and function.

Table S1 is an inventory of nuclear foci proteome with GO analysis and is provided online as an Excel file.

Table S2 shows a comparison with NOPdb and is provided online as an Excel file.

Table S3 shows a comparison with different datasets and is provided online as an Excel file.

Table S4 shows the classification of unknown nuclear foci and is provided online as an Excel file.

Table S5 shows the InterProScan analysis and is provided online as an Excel file.

Table S6 shows the top hit protein motif among various nuclear bodies and is provided online as an Excel file.

Table S7 is a list of interacting proteins from mass spectrometry analysis and is provided online as an Excel file.