

Supplemental Figures for

A systematic analysis of *Trypanosoma brucei* chromatin factors identifies novel protein interaction networks associated with sites of transcription initiation and termination

Desislava P. Staneva*, Roberta Carloni*, Tatsiana Auchynnikava Pin Tong, Juri Rappsilber, A. Arockia Jeyaprakash, Keith R. Matthews[†] and Robin C. Allshire[†]

* These authors contributed equally to this work.

[†] Co-corresponding authors:

Robin Allshire: robin.allshire@ed.ac.uk

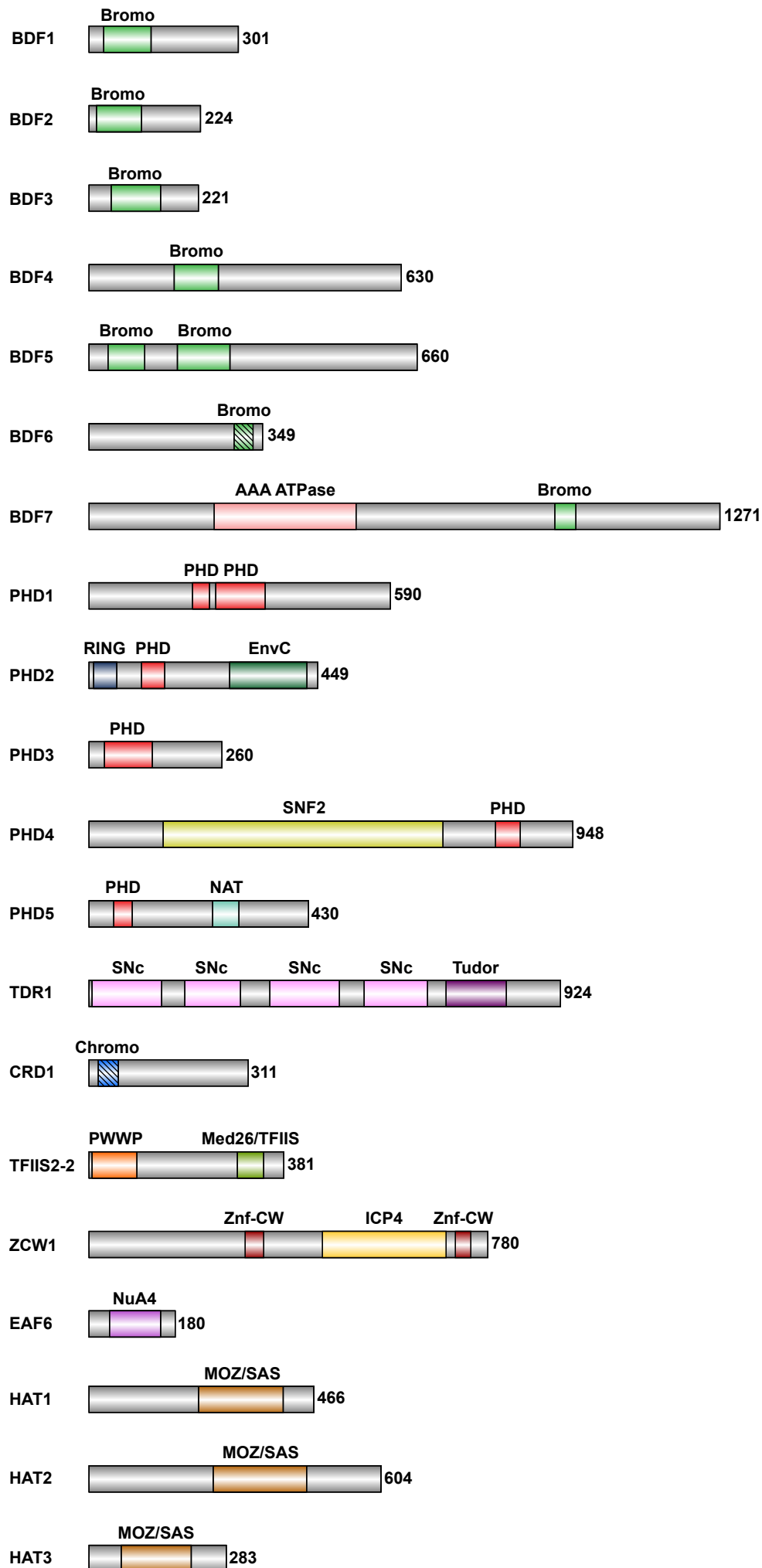
Keith Matthews: keith.matthews@ed.ac.uk

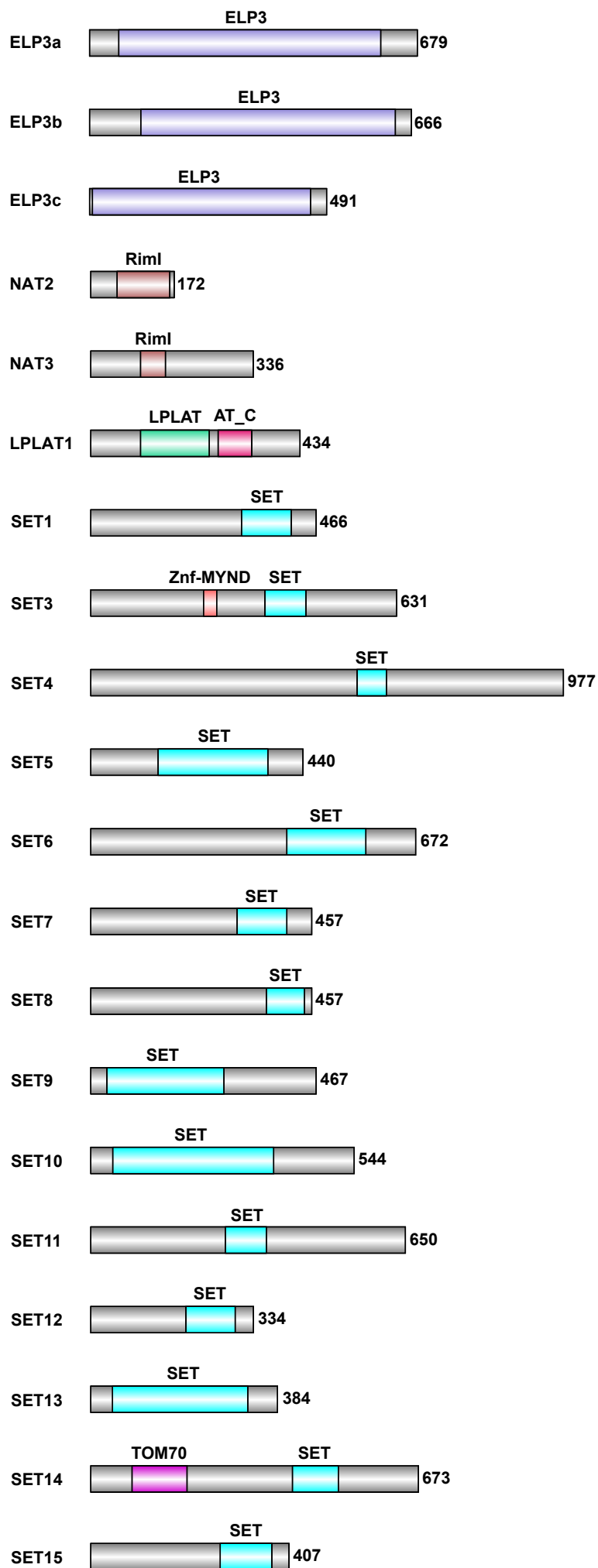
Contents

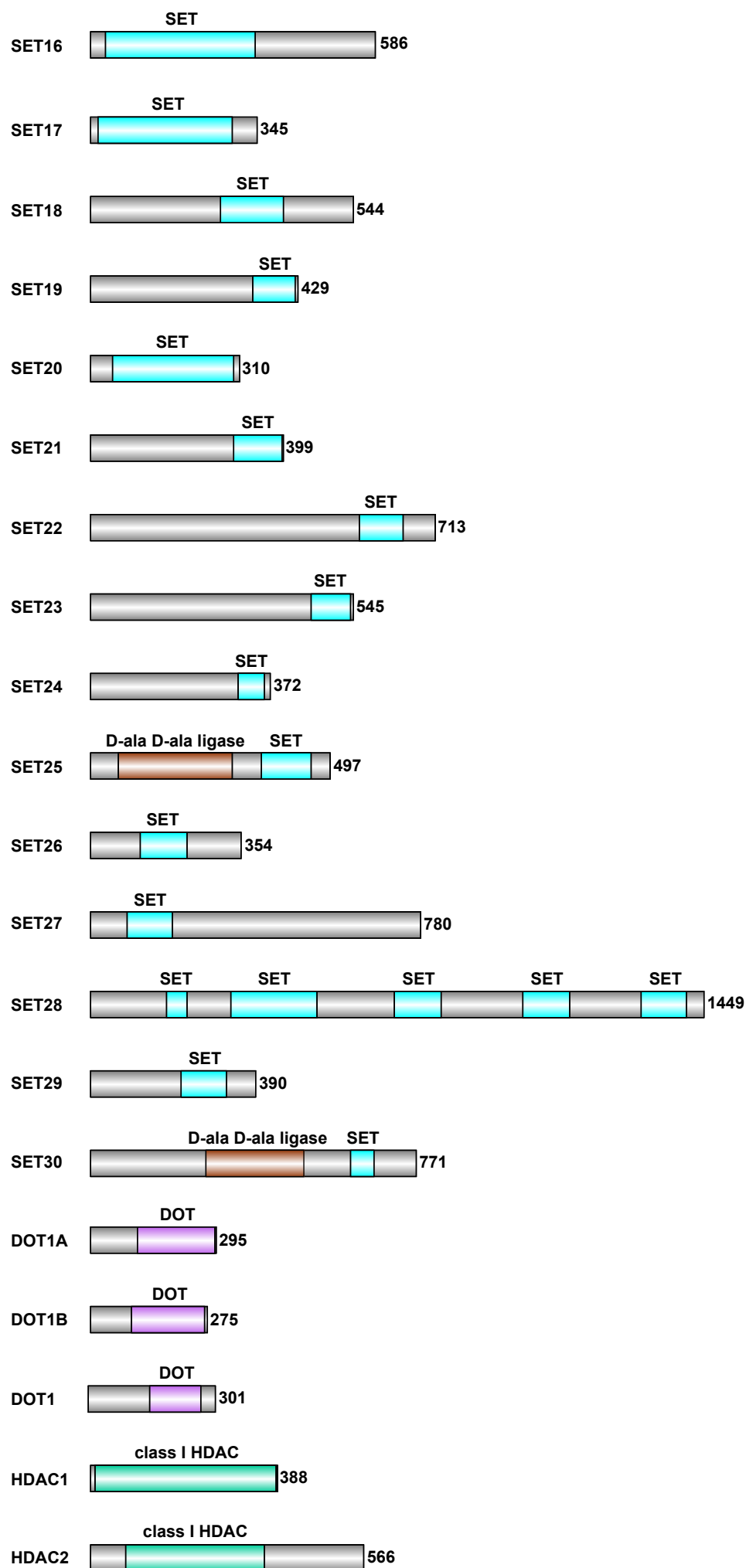
Page

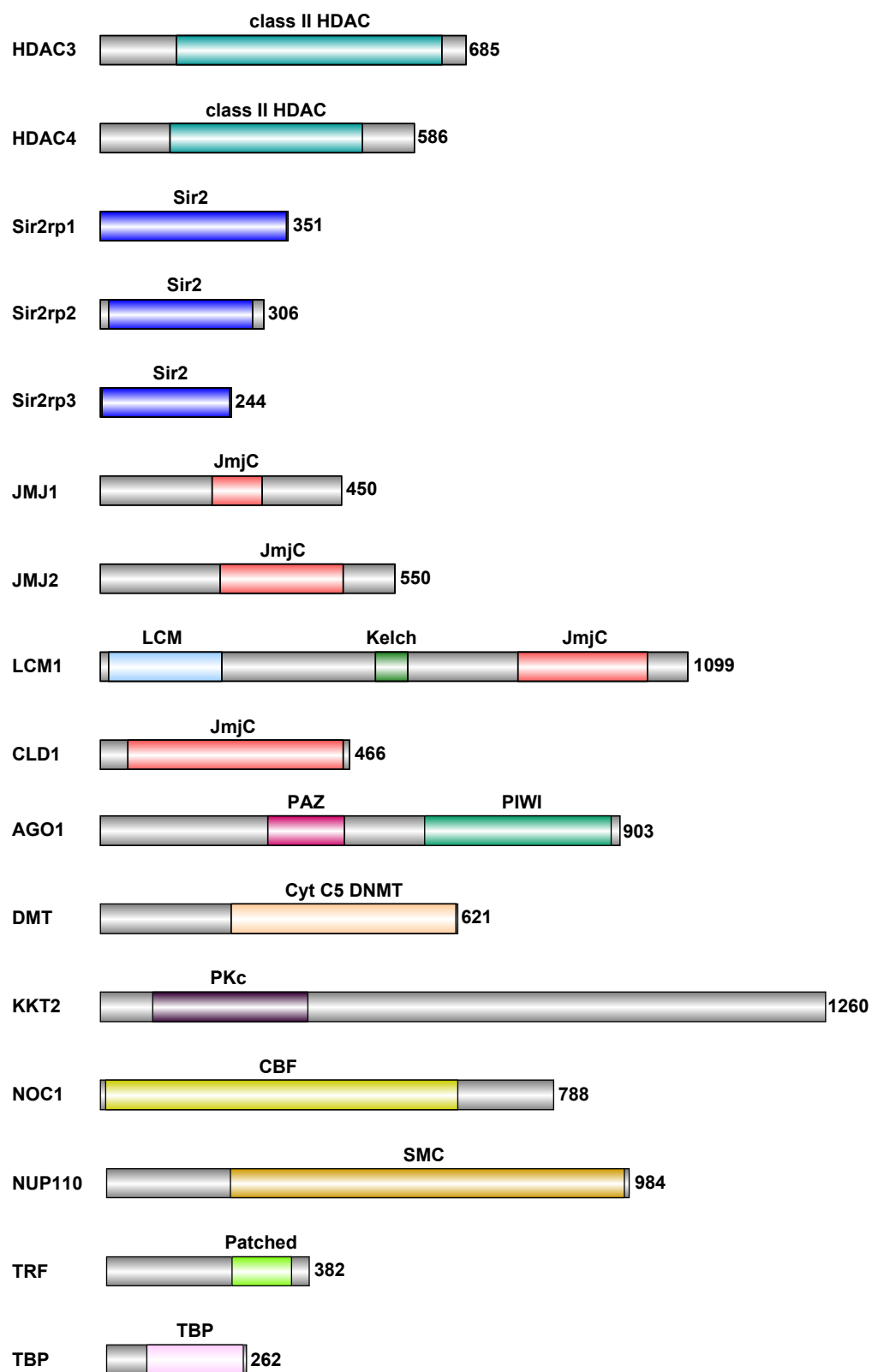
1	Supplemental Fig. S1. Domain architecture of the candidate proteins.
8	Supplemental Fig. S2. Cellular localization of <i>T. brucei</i> proteins with no specific genomic enrichment by ChIP-seq.
11	Supplemental Fig. S3. Comparison of different imaging procedures.
13	Supplemental Fig. S4. Localization of YFP-tagged proteins at different stages of the cell cycle.
29	Supplemental Fig. S5. YFP-tagged proteins with no specific genomic enrichment by ChIP-seq.
31	Supplemental Fig. S6. Examples of ChIP-seq enrichment of TSR- and RNAPIII/TTR-associated factors.
33	Supplemental Fig. S7. Average metagene plots.
35	Supplemental Fig. S8. Proteomic analysis of affinity selections for several non-chromatin associated proteins.

A









B

	10	20	30	40	50	60	
<i>Tb_CRD1</i>	...VNEPLSNPDD	SALEYCV	EAQSPV	NRITSTW	EFRSK	LF...PHAFAAV	LDTDLSE 60
<i>Sp_Chp1</i>	VEDILADRVNKN	GINEY YIKWAGY	DWYDNTW	EPEQNLFGA	EKV LKK	WKKRKKL	76
<i>Sp_Swi6</i>	...KVLKHRMARK	GGYEYLLKWE	GYDDPS	DNTWS	SEADCSGC	KQLIEAYWNEH	GG R 138
<i>Hs_Cbx5</i>	VEKYLDRRVVK	...GQVEYLLKWK	GF...SEE	HNTWEP	EKNLD	CPELISEFMK	KYKKM 66
<i>Mm_Cbx7</i>	VESIRKKRVRK	...GKVEYLVKWK	GW...PPKY	STWEP	EH...L	DPRLVMAYE	EKEERD 63

C

	158	160	170	180	190	200							
<i>BDF6 Tb_XP_001219038</i>	...TS	EQKARWH	NLRRE	LISPD	LIDLVQ	QFTLV	DGAAVFLAPVV...PSTVMEFN 207						
<i>BDF5 Tc_A0A2V2W6E9</i>	...	MEERRQ	KFY P	...EELV	ALVRS	LDRPQ	DEGLFSMDVLVVYPHLEQEYT 46						
<i>XP_003858810.1 Ld</i>	...	SAGT	SAAHKTAAA	PPS...TREM	VLVDS	LNRR	REDGGAFSVDVAEAYPDLRDSYR 202						
<i>BRD7 Hs_Q9NP11</i>	...	PLTSS	SLAQEE	VEQTPL...QEAL	NQLMR	QLQK	DPSPAFSFPVT...DFIAPGYS 50						
<i>TIF1A Hs_O15164.3</i>	CRDL	SKPEVEY	DCDAPS	HNSEKK	KTEGL	VKLTP	IDKRKCERLLLLFLYCHEMSLAFQDPV...PLTVPDYV 936						
<i>BPTF Hs_Q12830</i>	...LQ	SEAE	LI DEY	VCPQCQ	STED	AMT VLT	PLTEKDYEG LKRYLRSLQAHKMAWPFLEPYD...PNDAPDYV 2965						
	210	220	230	240	250	260	270						
<i>BDF6 Tb_XP_001219038</i>	GV	RS	SGPY	VTV	IHQPLS	LMCV	KRRVLAARRD	YELHKKHSGAYLP	TGQNNVGVGSRKRE	RGNG	GVT	SR	ASAVS 278
<i>BDF5 Tc_A0A2V2W6E9</i>	RV	...	CP	KRCD	LATA	AEKA...ANE	AYS	YDVN	...	...	...	...	72
<i>XP_003858810.1 Ld</i>	KI	...	CP	RP	MNL	LIL	MRORA...KEGY	YTS	SGS	ATV	YGD	VAA	S 238
<i>BRD7 Hs_Q9NP11</i>	MI	...	KHP	MD	FST	MKE	I...KN	NDYQS	...	...	...	...	73
<i>TIF1A Hs_O15164.3</i>	KI	...	KNP	MD	LST	IKK	RL...QEDY	SMYSK	...	...	...	...	961
<i>BPTF Hs_Q12830</i>	GV	...	KEP	MD	LAT	ME	ERV...QR	RYEK	...	...	...	...	2988
	280	290	300	310	320	330	340						
<i>BDF6 Tb_XP_001219038</i>	QPPH	F	S	I	A	T	NSGE	KG	N	V	I	R	T
<i>BDF5 Tc_A0A2V2W6E9</i>	...	...	...	...	...	...	...	...	...	...	...	...	...
<i>XP_003858810.1 Ld</i>	...	...	...	...	...	...	...	...	...	...	...	...	...
<i>BRD7 Hs_Q9NP11</i>	...	...	...	...	...	...	...	...	...	...	...	...	...
<i>TIF1A Hs_O15164.3</i>	...	...	...	...	...	...	...	...	...	...	...	...	...
<i>BPTF Hs_Q12830</i>	...	...	...	...	...	...	...	...	...	...	...	...	...

Supplemental Fig. S1. Domain architecture of the candidate proteins.

A. Conserved sequences and domains in the candidate proteins (described briefly below) were identified using NCBI's Conserved Domains tool (<https://www.ncbi.nlm.nih.gov/cdd>). The shaded domains of BDF6 and CRD1 are weakly predicted.

Bromo – domain which binds acetylated lysines; found in chromatin-associated proteins and in histone acetyltransferases;

AAA ATPase – has chaperone-like functions that aid assembly, function or disassembly of protein complexes;

PHD (plant homeodomain) - a Cys₄-His-Cys₃ zinc finger motif found in nuclear proteins involved in chromatin-mediated transcriptional regulation; some PHD fingers bind methylated histones;

RING – a specialised type of zinc finger often found in ubiquitin protein ligases;

EnvC – bacterial protein which activates cell wall hydrolases and is required for daughter cell separation following cell division;

SNF2 – domain found in proteins involved in a variety of processes including transcription regulation, DNA repair, DNA recombination and chromatin unwinding;

NAT – N-acetyltransferase;

SNc - Staphylococcal nuclease fold;

Tudor – domain which can recognise methylated histone lysines and arginines; present in several RNA-binding proteins;

Chromo (chromatin organisation modifier) – domain which binds methylated histones; involved in chromatin organisation, specifically heterochromatin formation;

PWWP – domain which contains a Pro-Trp-Trp-Pro motif; binds methylated histone lysines; found in DNA-binding proteins that function as transcription factors;

Med26/TFIIS - TFIIS helical bundle-like domain; component of the mediator complex involved in the regulation of RNAPII-transcribed genes;

Znf-CW – a zinc finger domain containing conserved Cys and Trp residues; implicated in DNA binding and protein-protein interactions, particularly recognition of methylated histones;

ICP4 – Herpesvirus protein required for transcription of viral genes; binds DNA in a sequence-specific manner;

NuA4 – histone acetyltransferase subunit;

MOZ/SAS - suggested to be homologous to acetyltransferases;

ELP3 - radical SAM enzyme/protein acetyltransferase; this family includes elongator complex protein 3 (ELP3) which is a component of the RNAPII holoenzyme;

RimI – ribosomal protein acetyltransferase found in bacteria; mediates acetylation of N-terminal residues;

LPLAT - lysophospholipid acyltransferase;

AT_C – domain found at the C-terminus of several acyltransferases;

SET - Su(var)3-9, Enhancer-of-zeste, Trithorax; catalytic domain of lysine methyltransferases;

Znf-MYND - MYND-type zinc finger; protein-protein interaction domain;

TOM70 – component of the translocase of outer membrane (TOM) complex involved in mitochondrial import;

D-ala D-ala ligase – bacterial enzyme involved in peptidoglycan synthesis and cell wall biogenesis;

DOT - Disruptor of telomeric silencing; domain which regulates gene expression via histone methylation;

Class I HDACs - Zn-dependent histone deacetylases;

Class II HDACs- Zn-dependent histone deacetylases;

Sir2 HDACs – sirtuins; NAD-dependent histone deacetylases;

JmjC – found in metalloenzymes that adopt the cupin fold; function in histone demethylation

LCM - leucine carboxyl methyltransferase;

Kelch – sequence motif present in proteins with diverse functions including cytoskeletal support and oxidation;

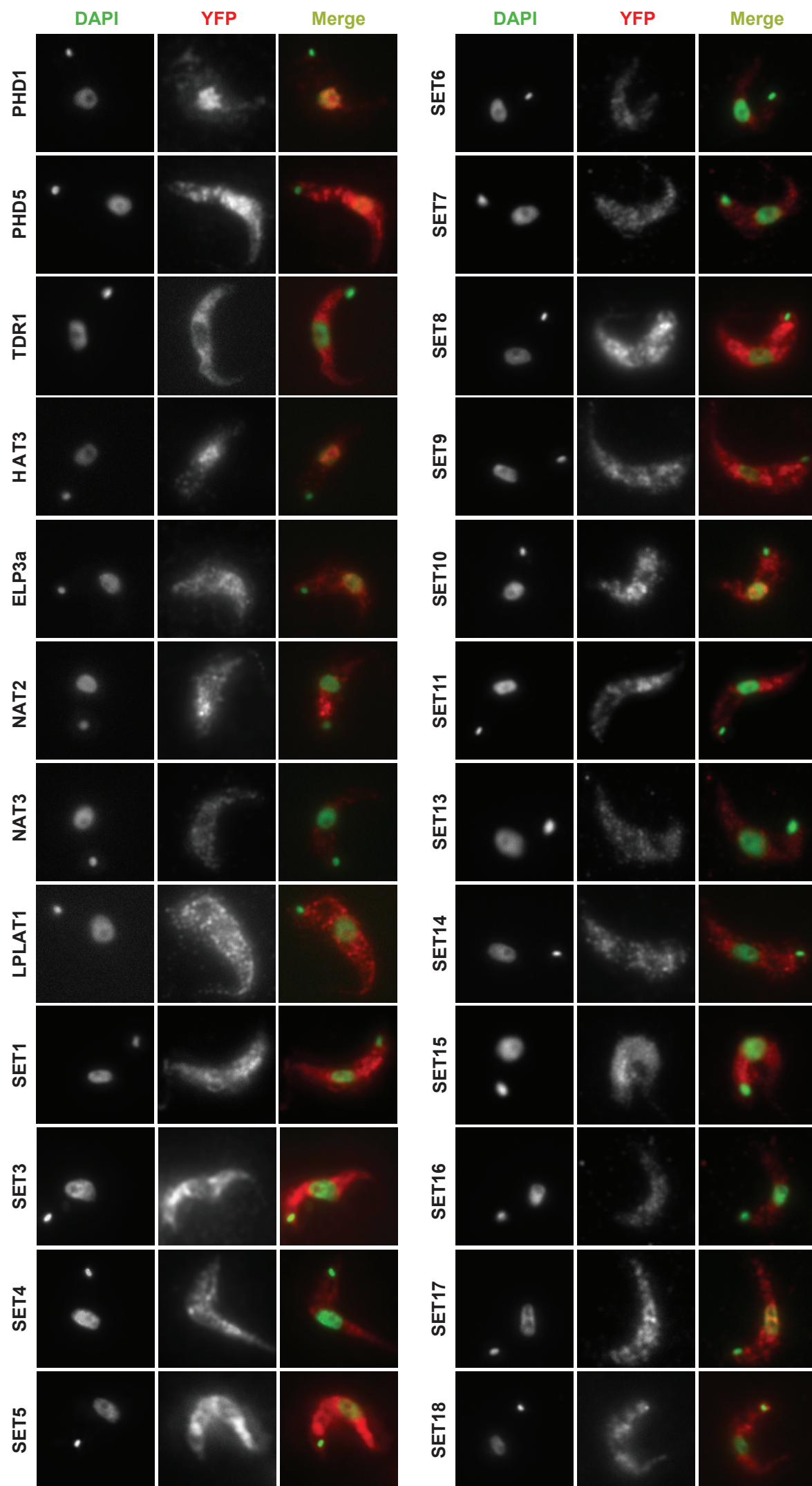
SMC (structural maintenance of chromosomes) – chromosome segregation protein;

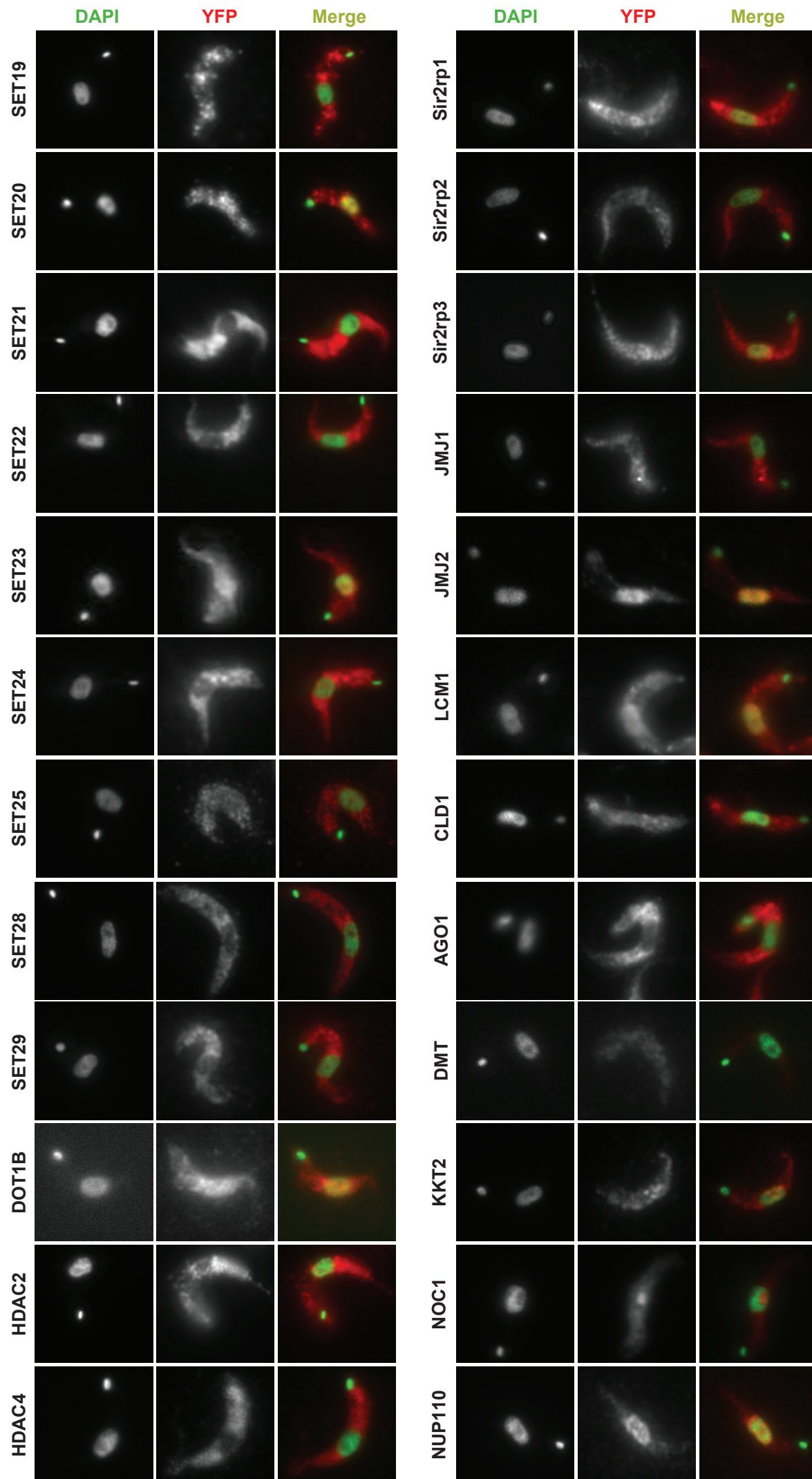
Patched – transmembrane receptor for Sonic Hedgehog;

TBP (TATA-binding protein) – part of the DNA-binding transcription factor complex TFIID;

B. Multiple sequence alignment of the putative TbCRD1 chromodomain with *S. pombe* Chp2 and Swi6; Human Cbx5 and Mouse Cbx7. Level of sequence similarity is indicated as a color gradient from cyan (low similarity) to red (high similarity).

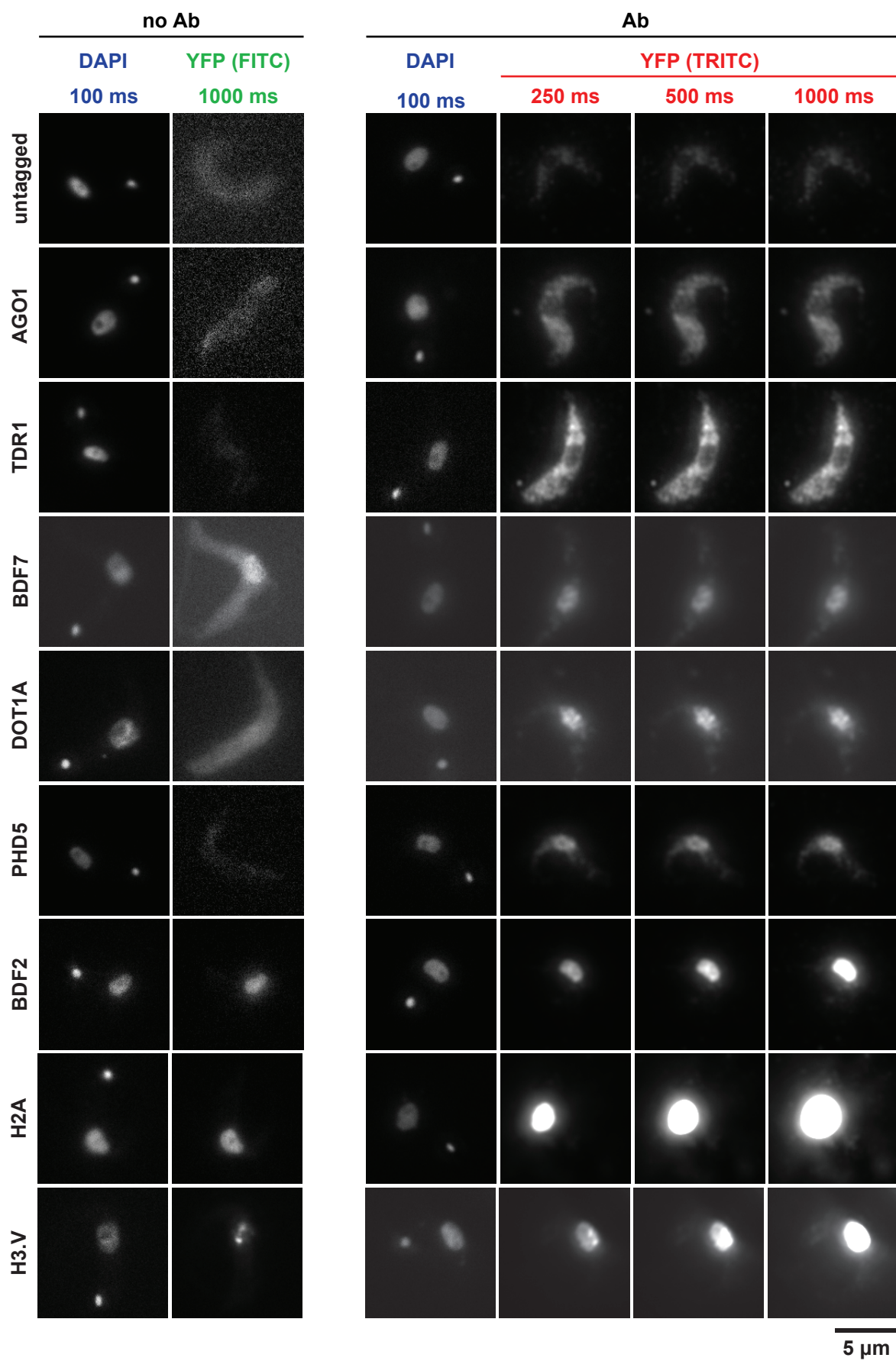
C. Multiple sequence alignments of the TbBDF6 Bromo domain with indicated protein accession numbers including Human BRD7, TIF1A and BPTF. Level of sequence similarity is indicated as a color gradient from cyan (low similarity) to red (high similarity).





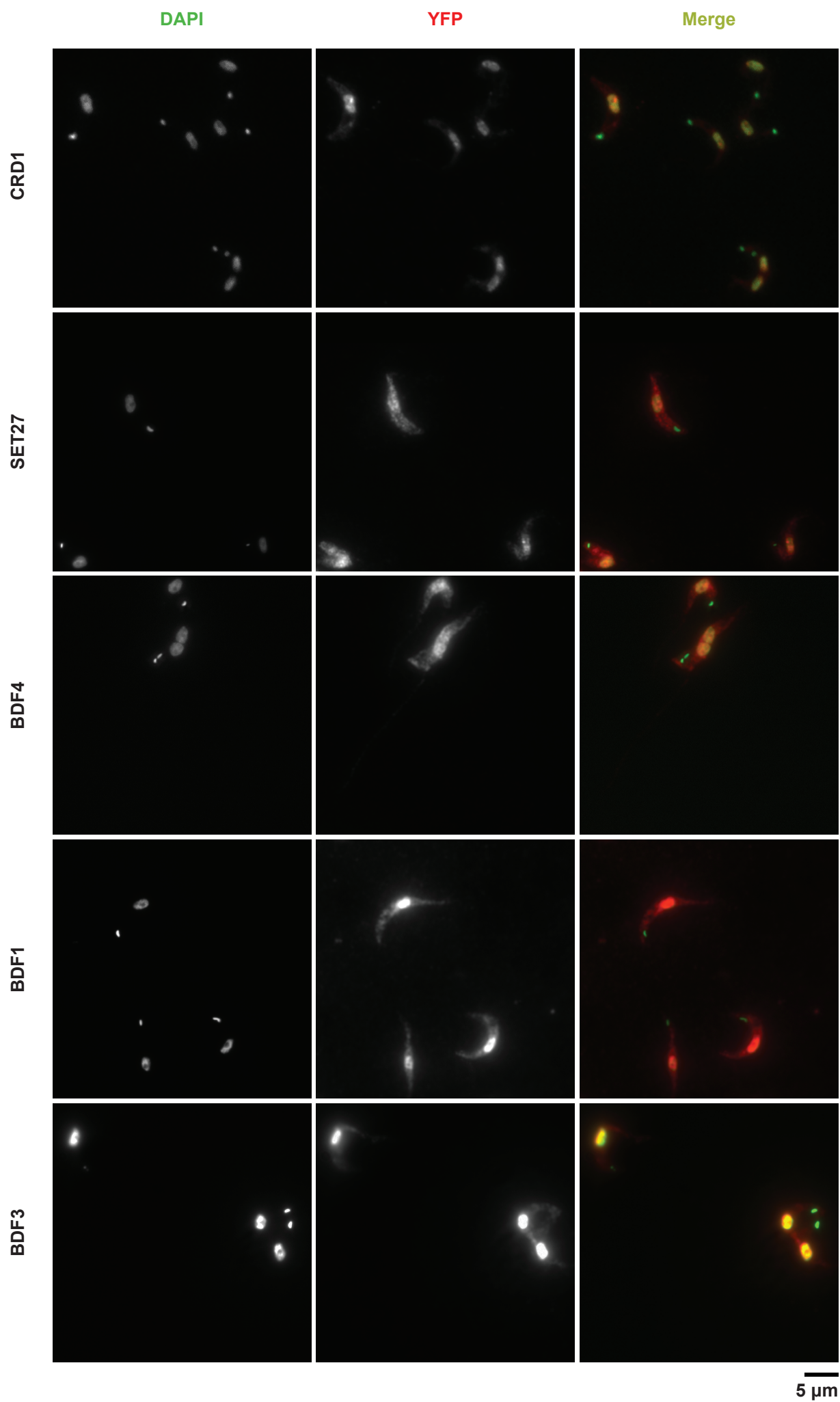
Supplemental Fig. S2. Cellular localization of *T. brucei* proteins with no specific genomic enrichment by ChIP-seq.

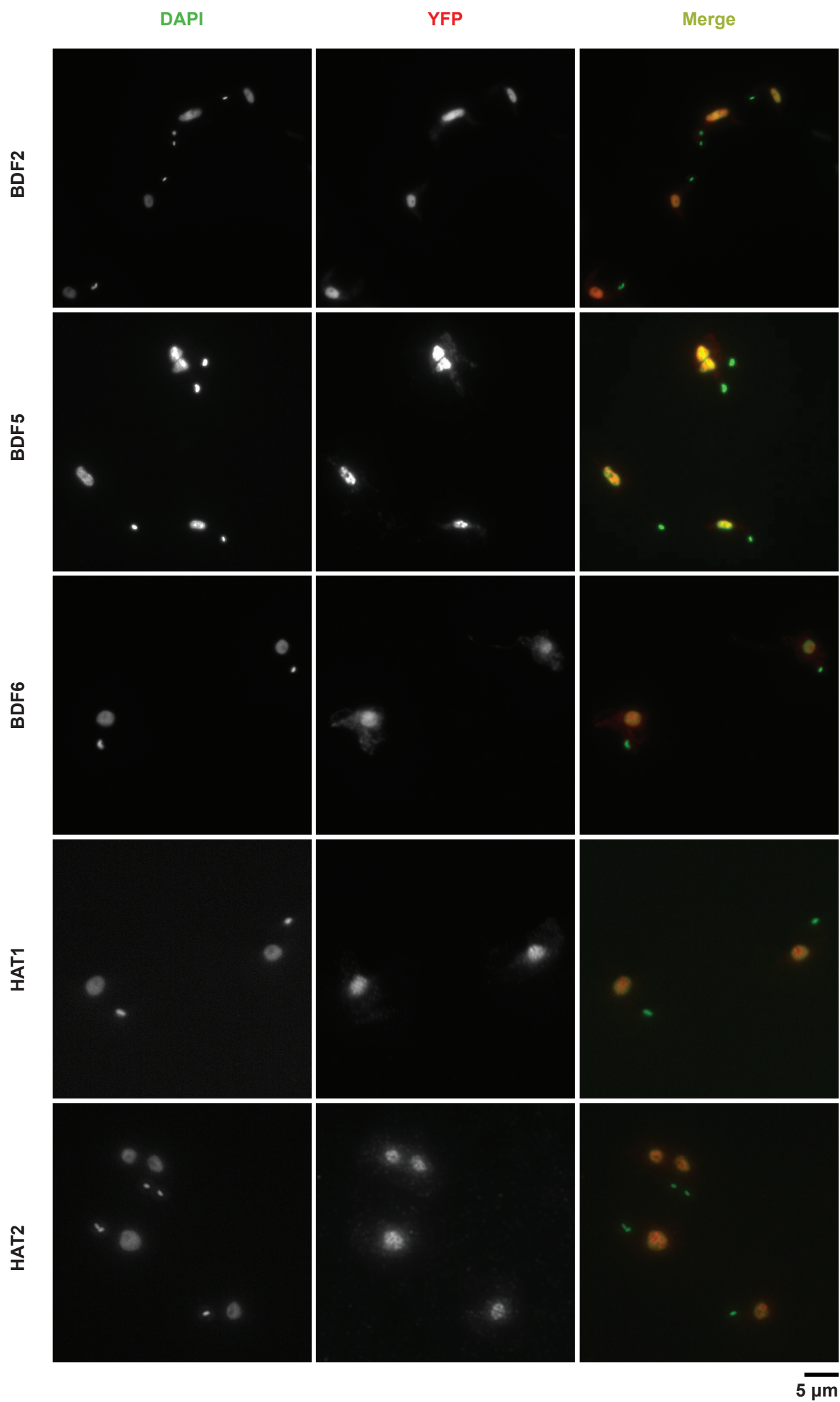
The indicated YFP-tagged proteins expressed in bloodstream Lister 427 cells from their endogenous genomic loci were detected with an anti-GFP primary antibody and an Alexa Fluor 568 labelled secondary antibody (red). Nuclear and kinetoplast (mitochondrial) DNA were stained with DAPI (green). Representative images are shown for each candidate protein for which no specific ChIP-seq pattern was detected. Images of KKT2, NOC1 and NUP110 are included as controls for nuclear localization patterns. AGO1 serves as a control for cytoplasmic localization. Scale bar = 5 μ m.

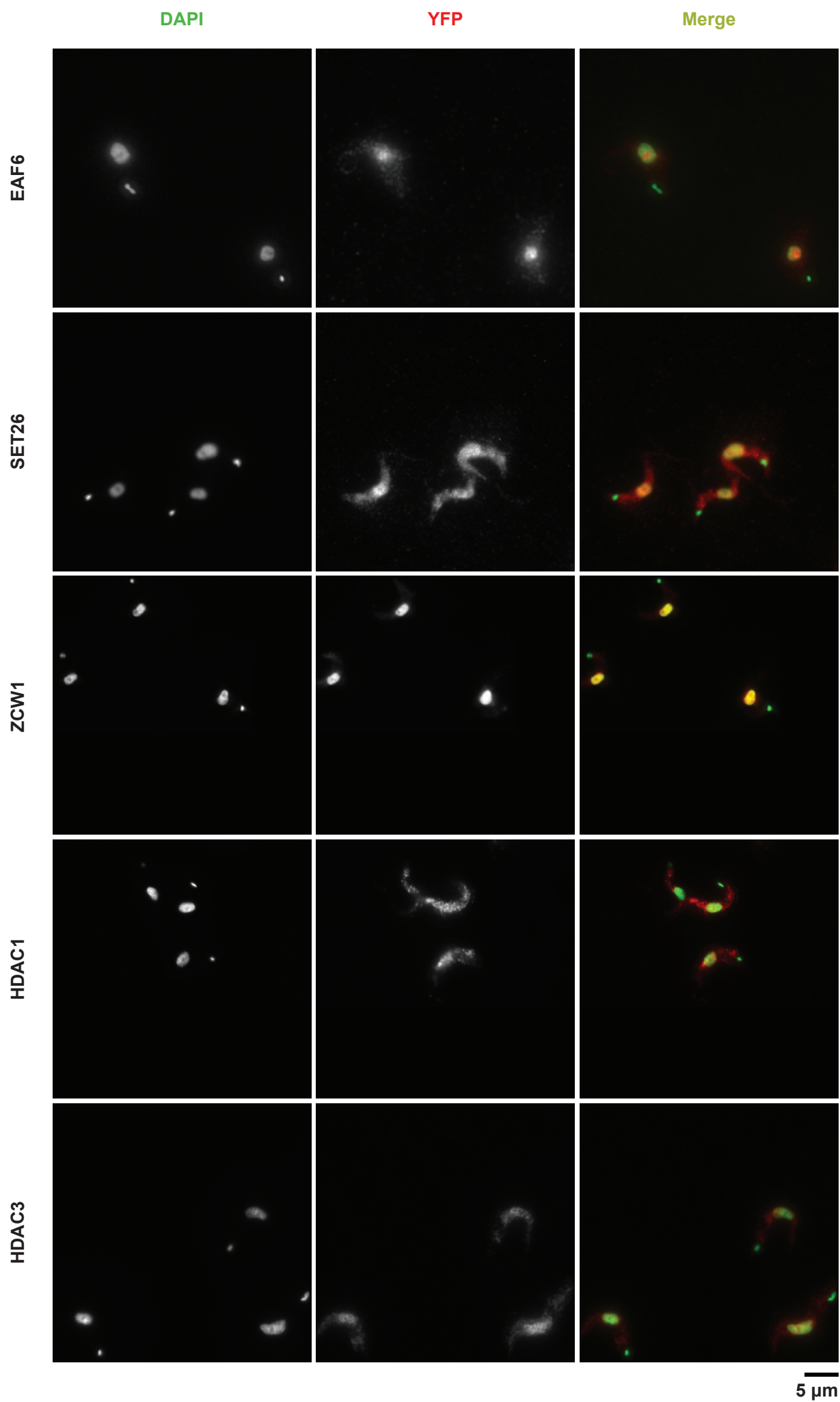


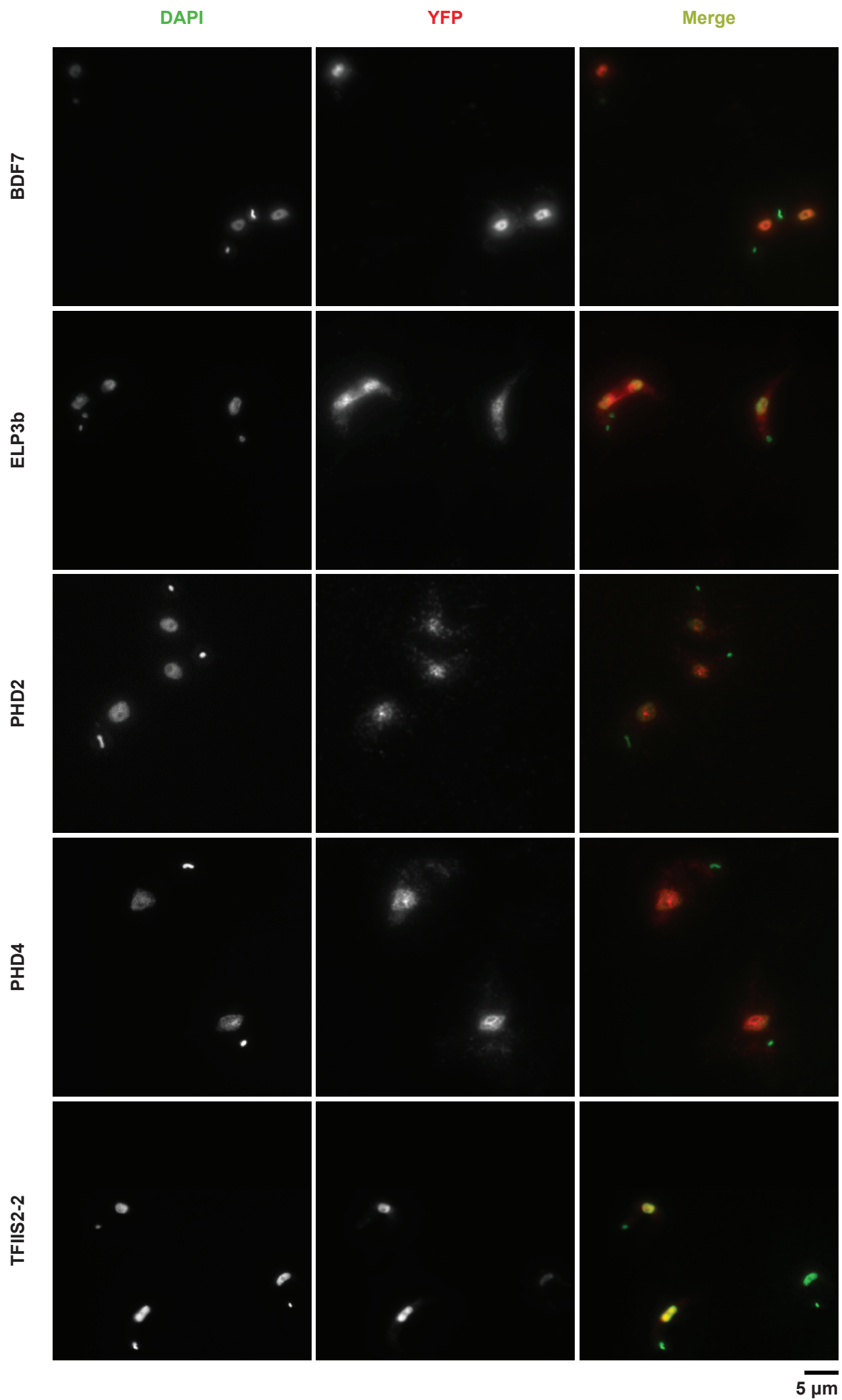
Supplemental Fig. S3. Comparison of different imaging procedures.

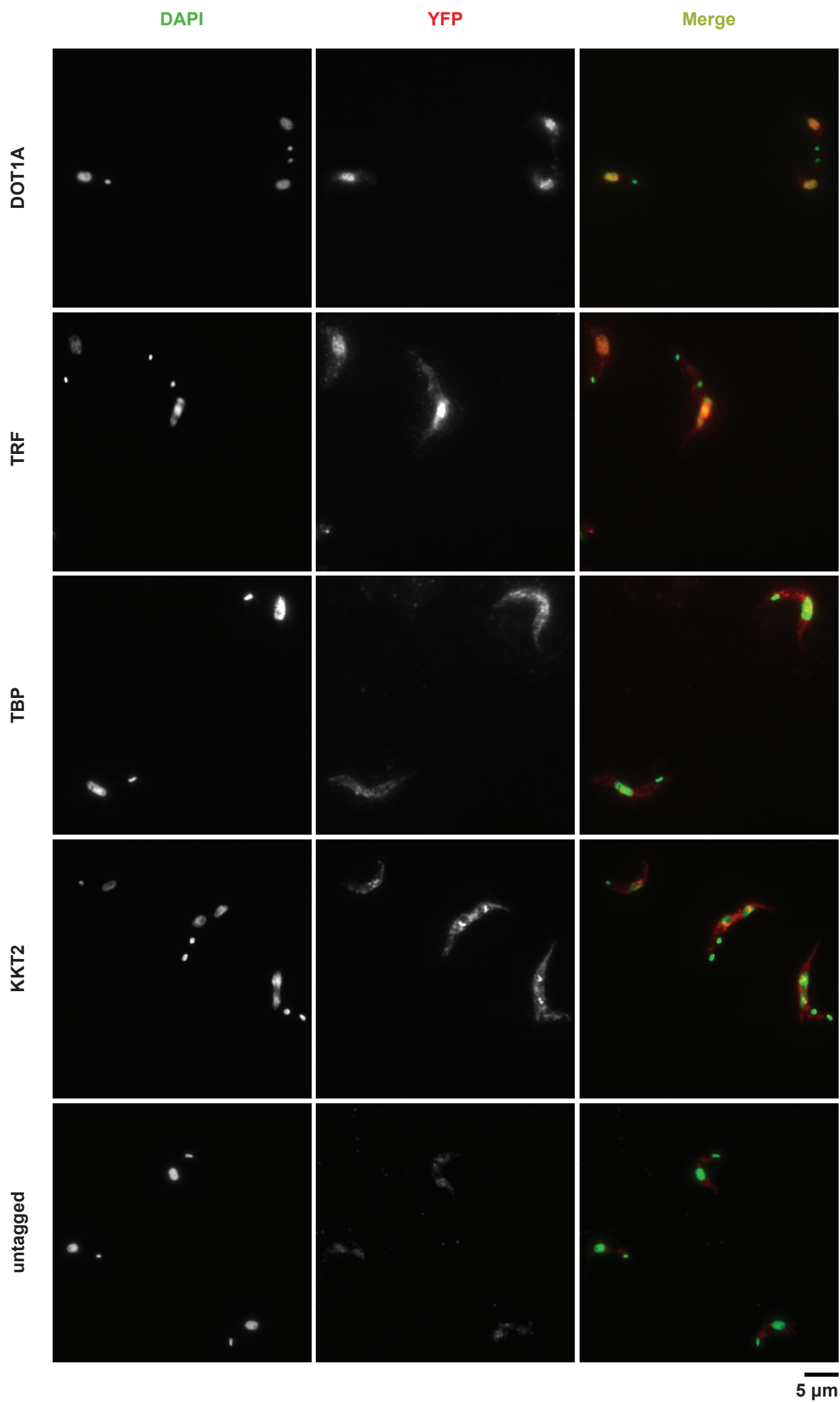
Selected YFP-tagged proteins and the untagged control were imaged with and without antibody staining. Cells stained only with DAPI (no Ab) were imaged at 100 ms (DAPI channel; blue) and 1000 ms (FITC channel where YFP is detected; green). Cells stained with DAPI, primary and secondary antibody (Ab) were imaged at 100 ms (DAPI channel) and 250, 500 or 1000 ms (TRITC channel where antibody-labelled YFP-tagged proteins are detected; red). AGO1 is used as a control for cytoplasmic localization whereas histones H2A and H3.V are controls for nuclear enrichment. Note that in some cases the inherent fluorescence of the expressed YFP fusion protein generates a signal even in the absence of secondary antibody. Scale bar = 5 μm .

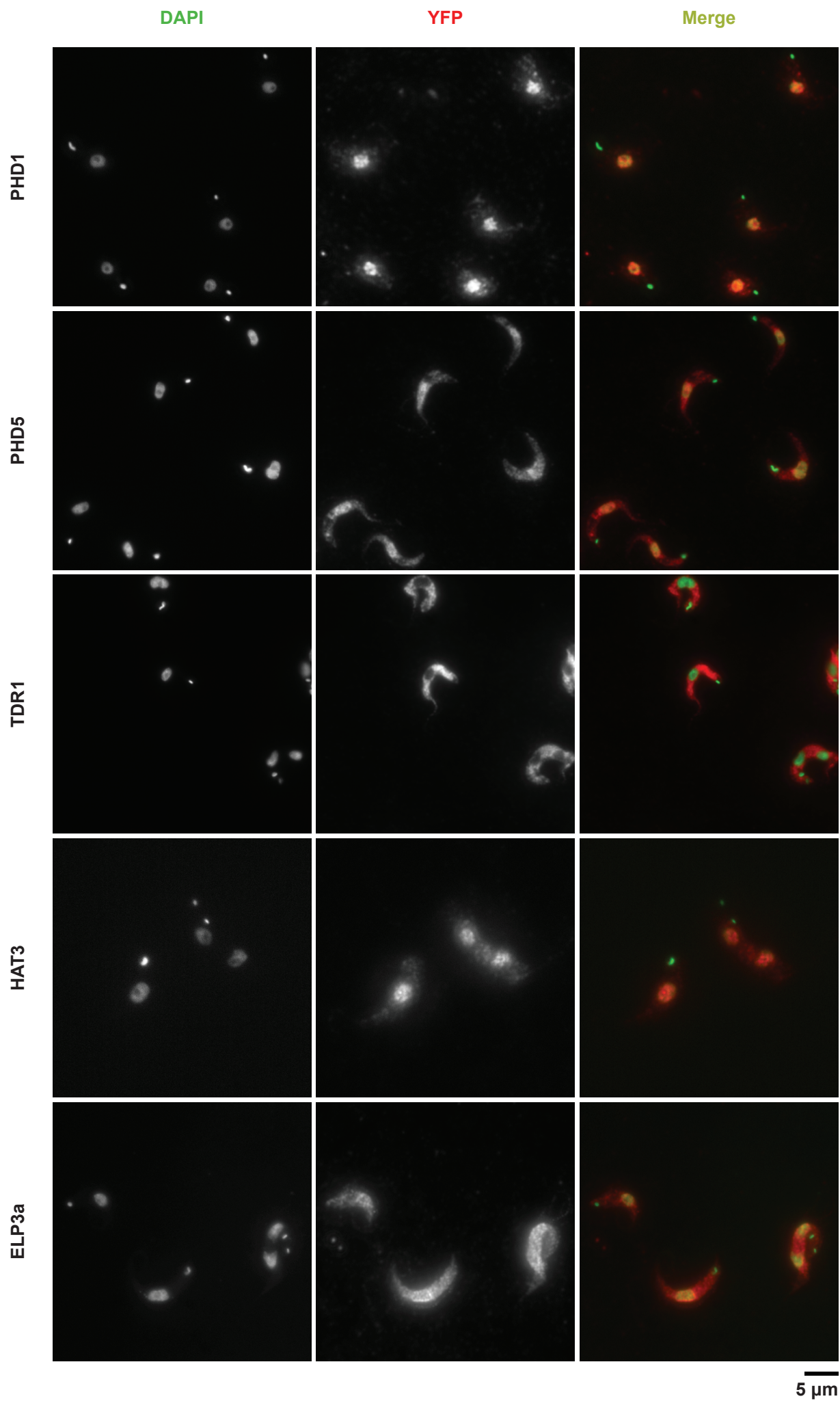


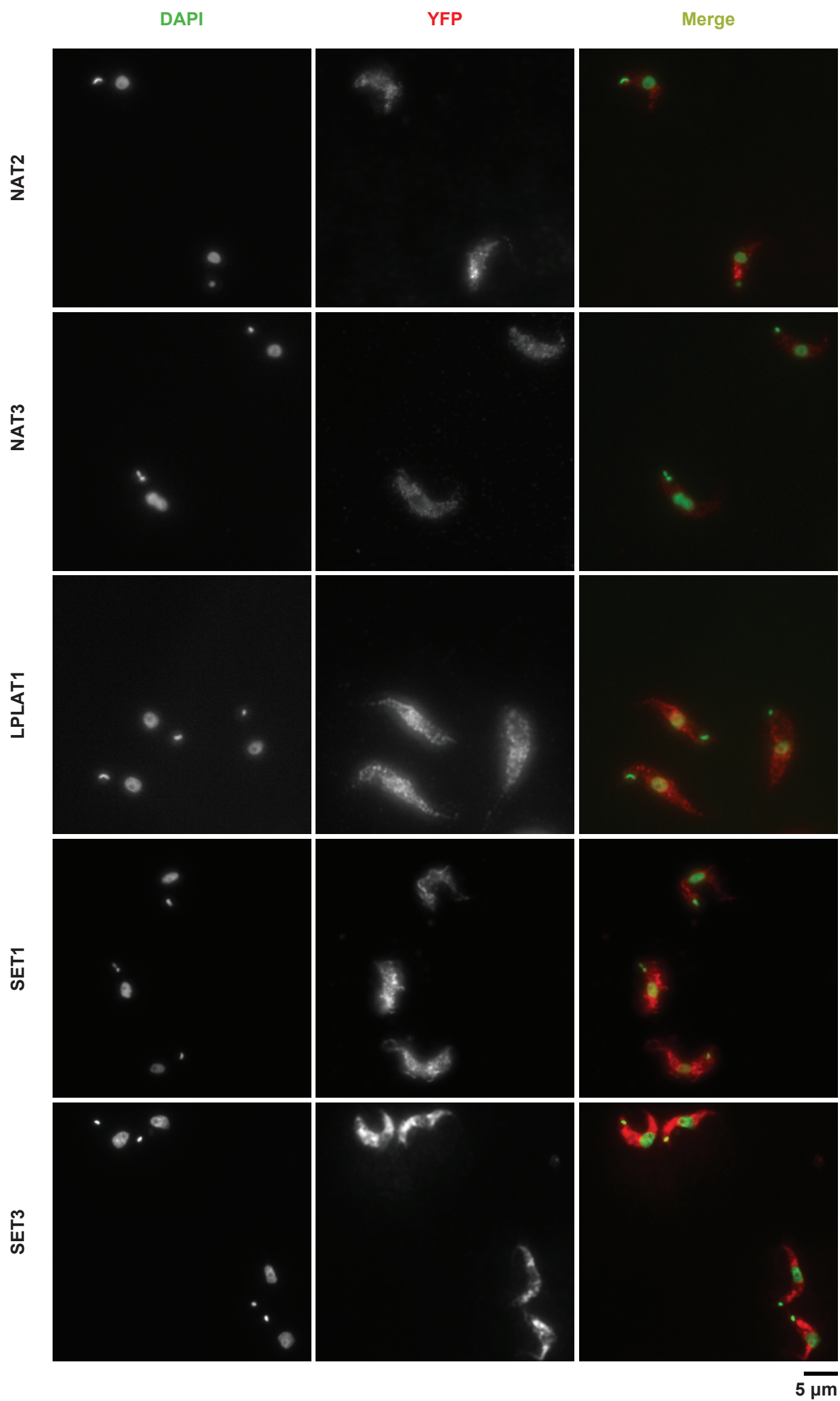




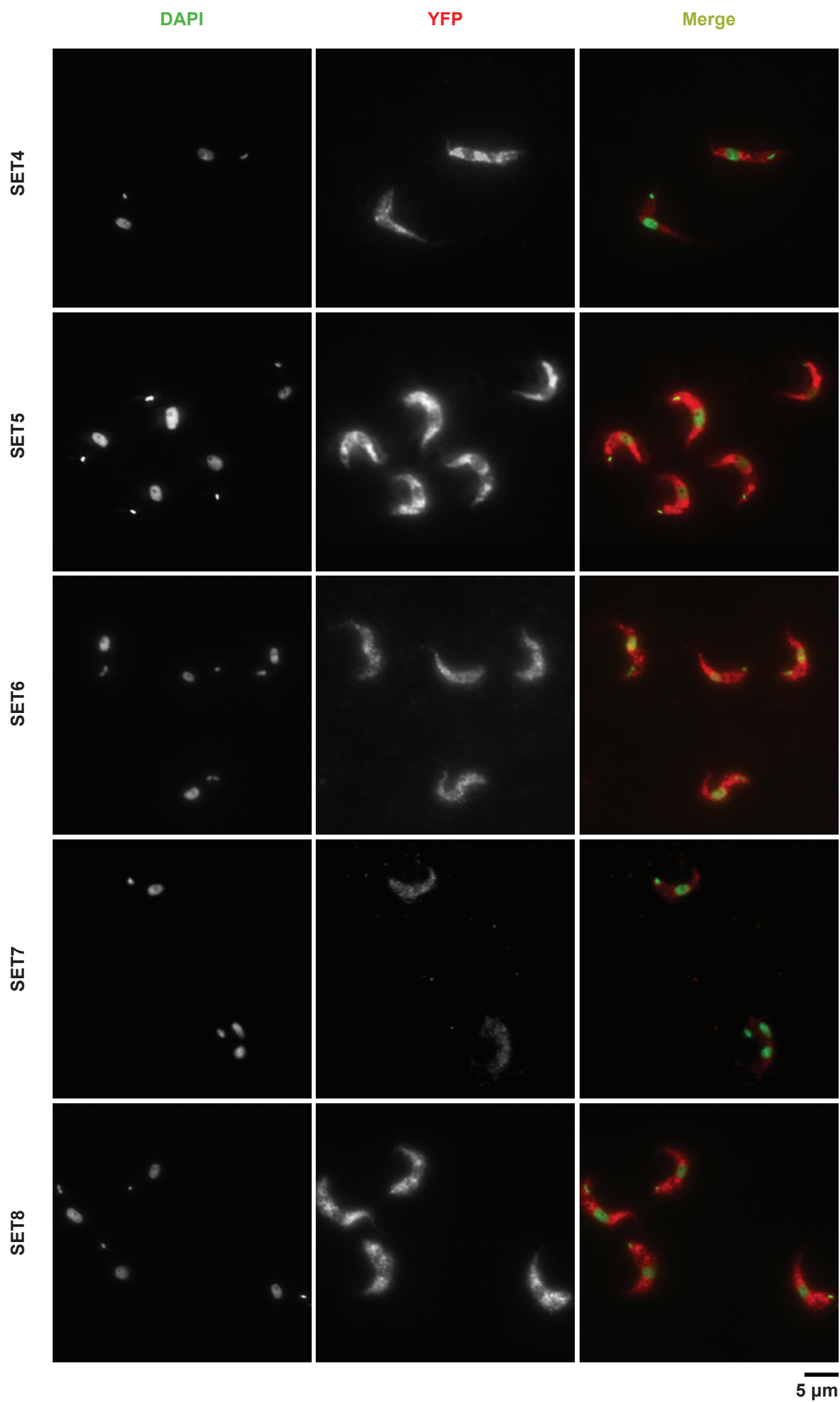




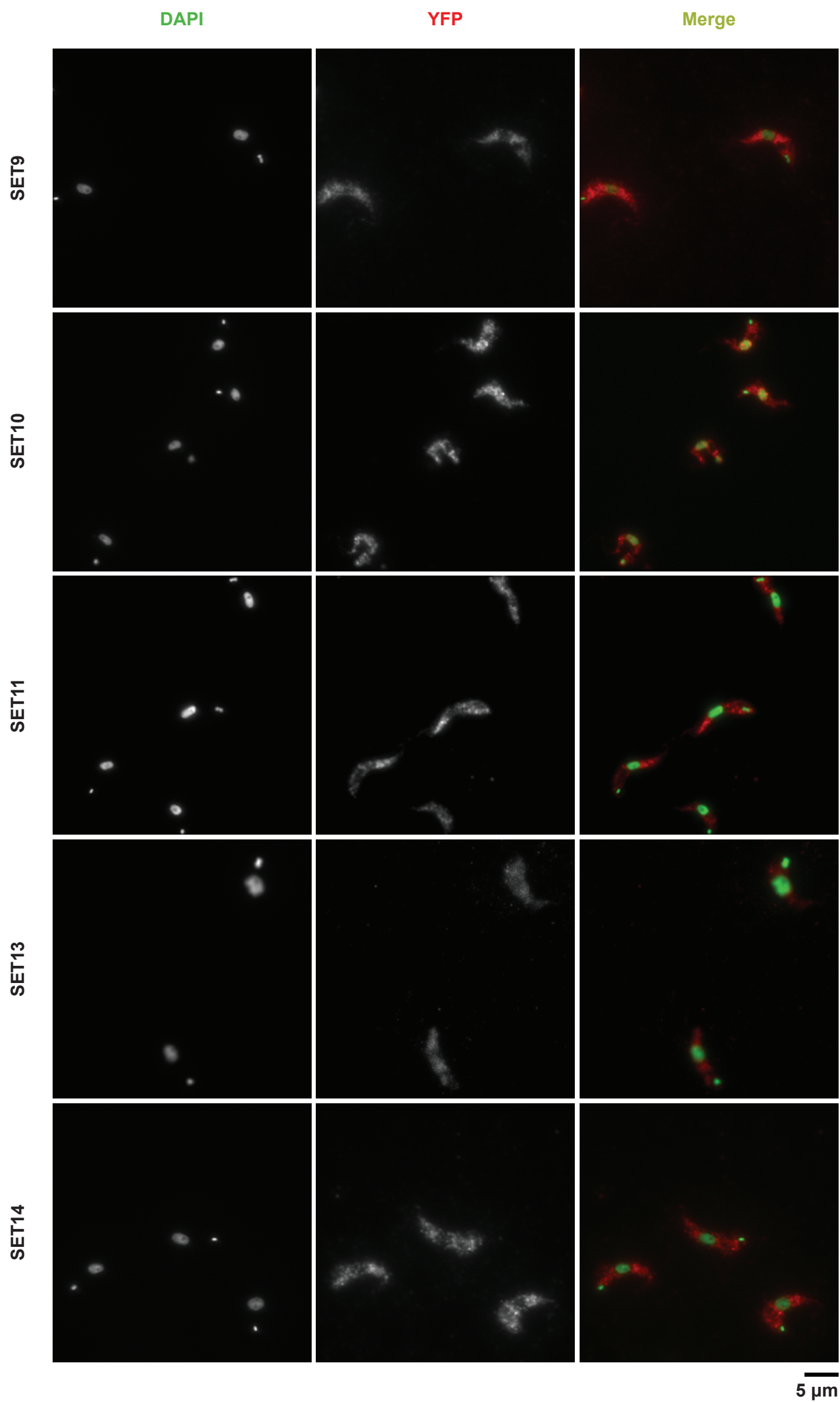


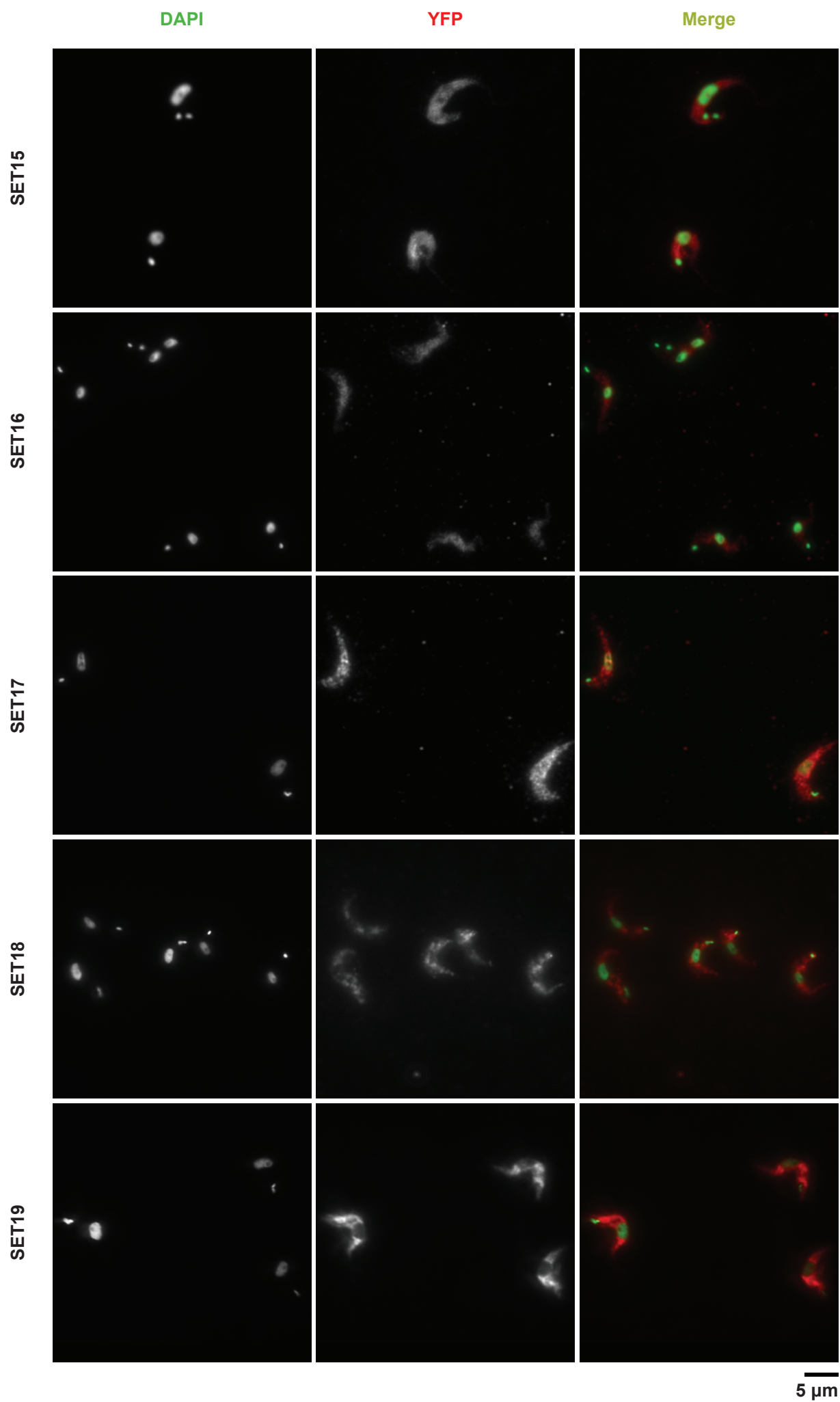


5 μ m

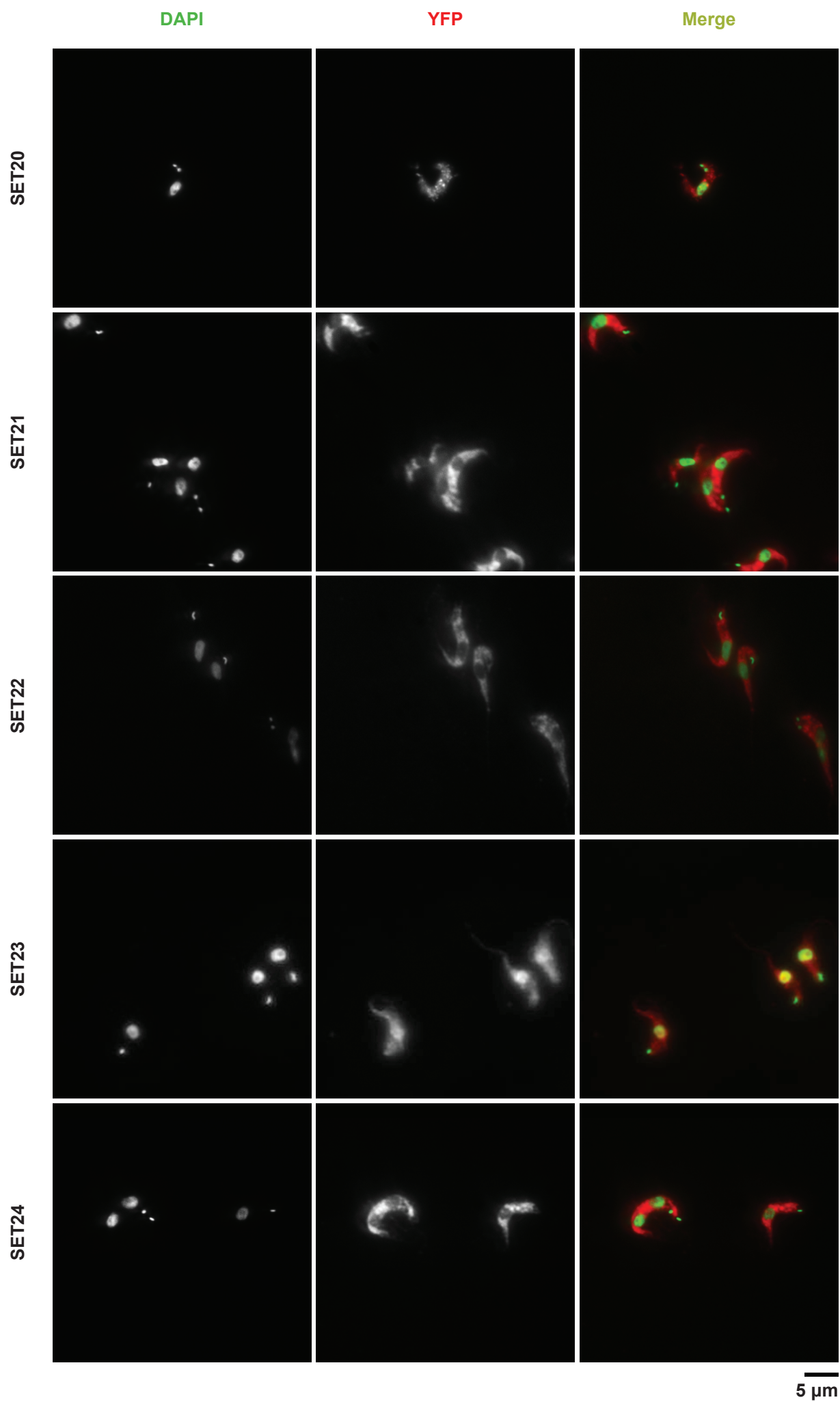


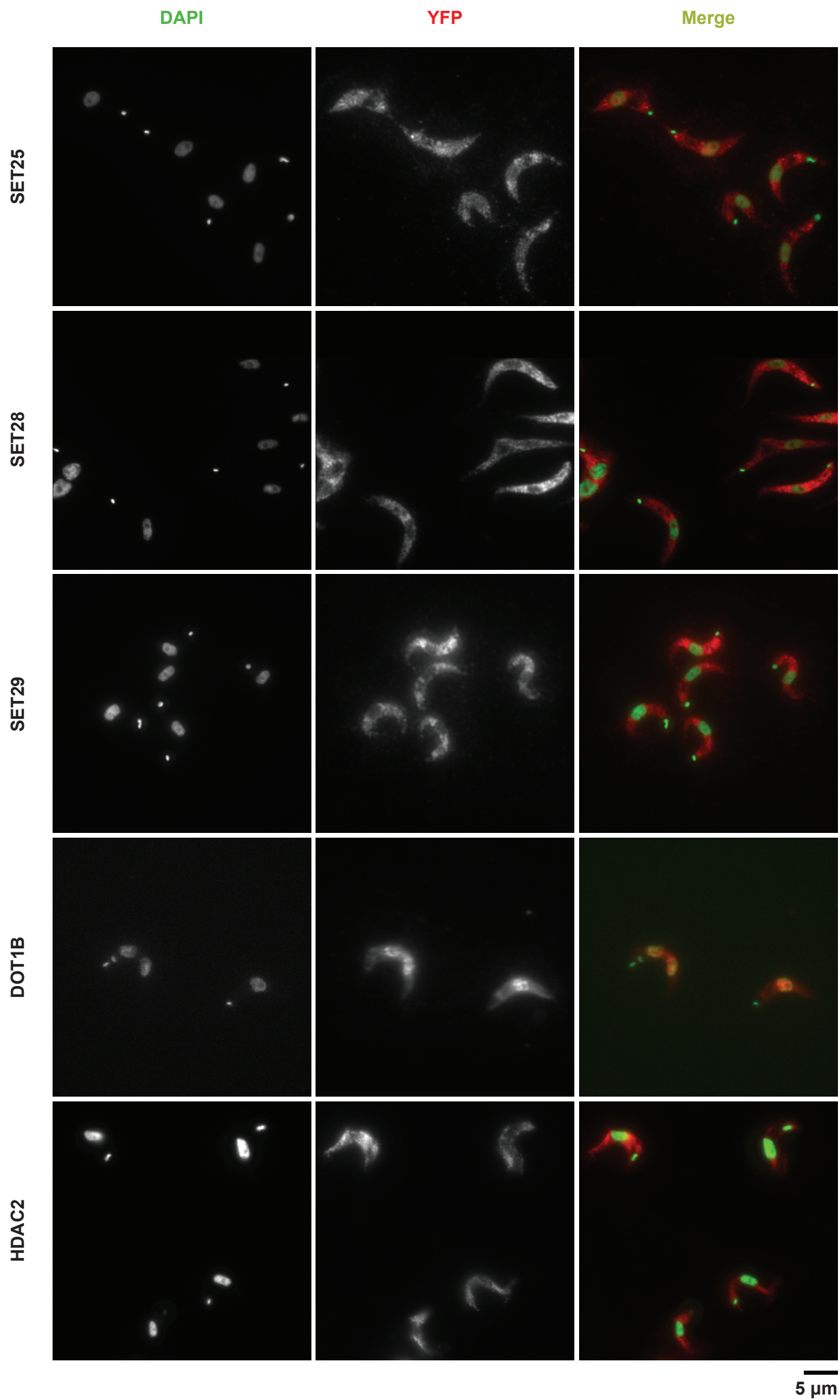
5 μ m

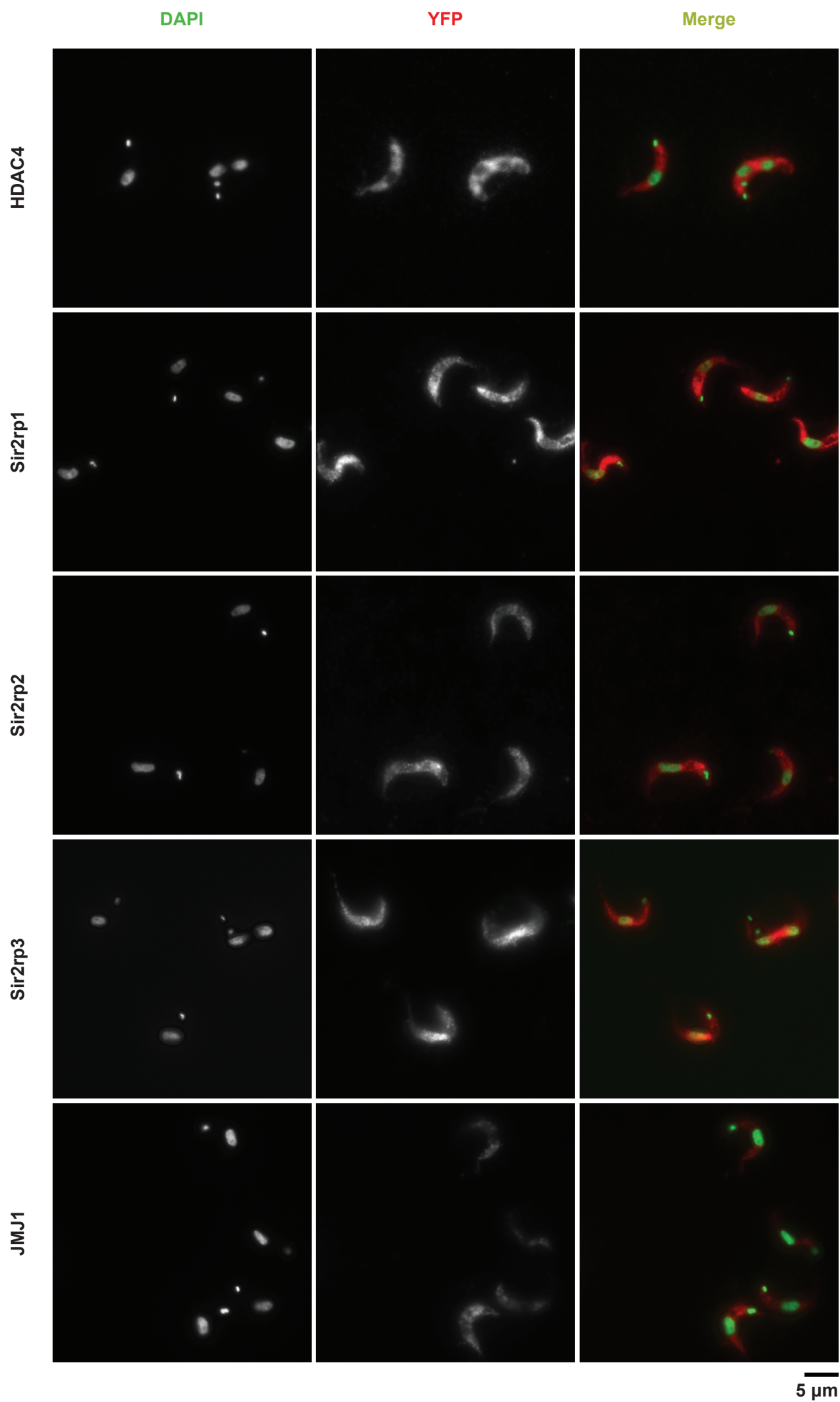


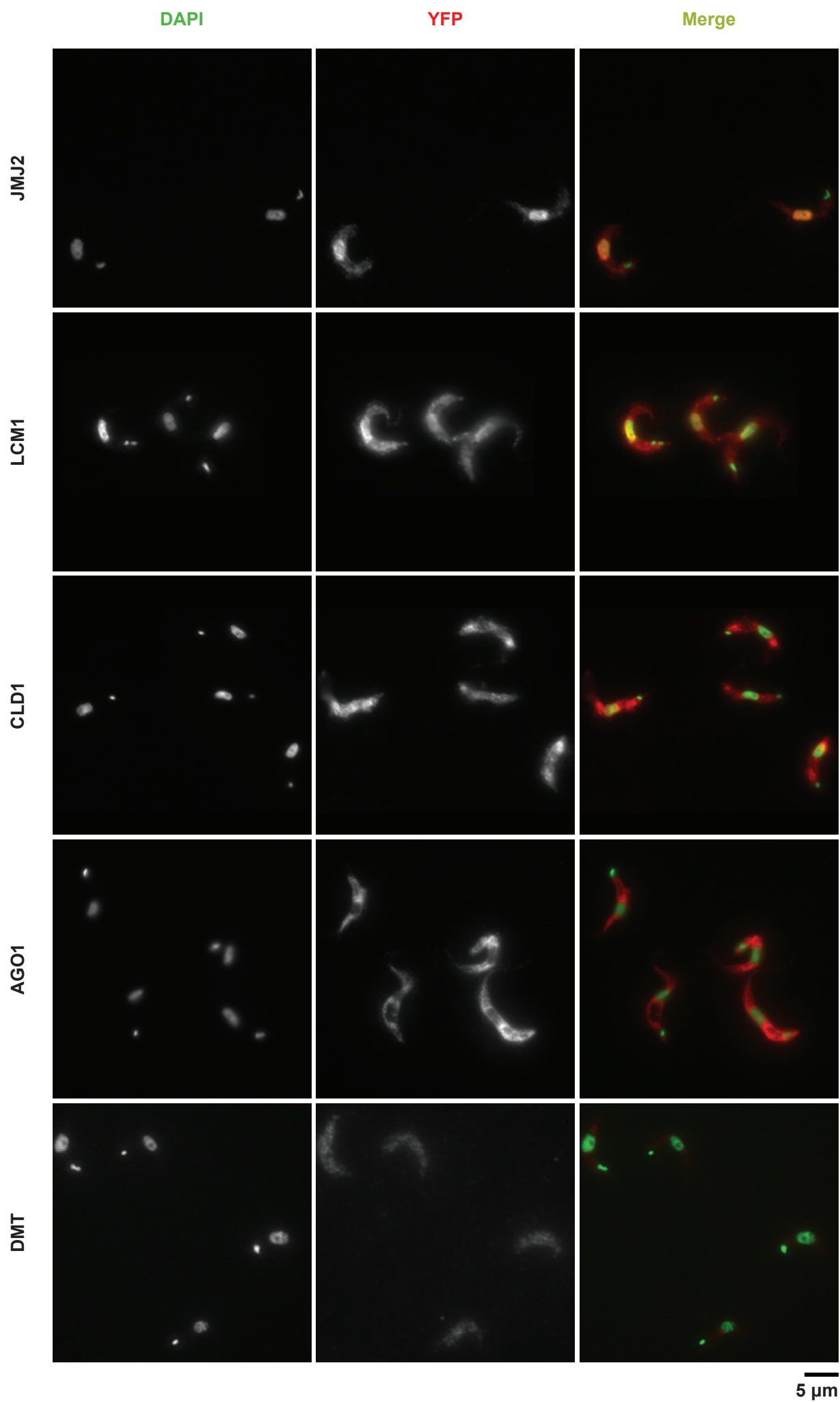


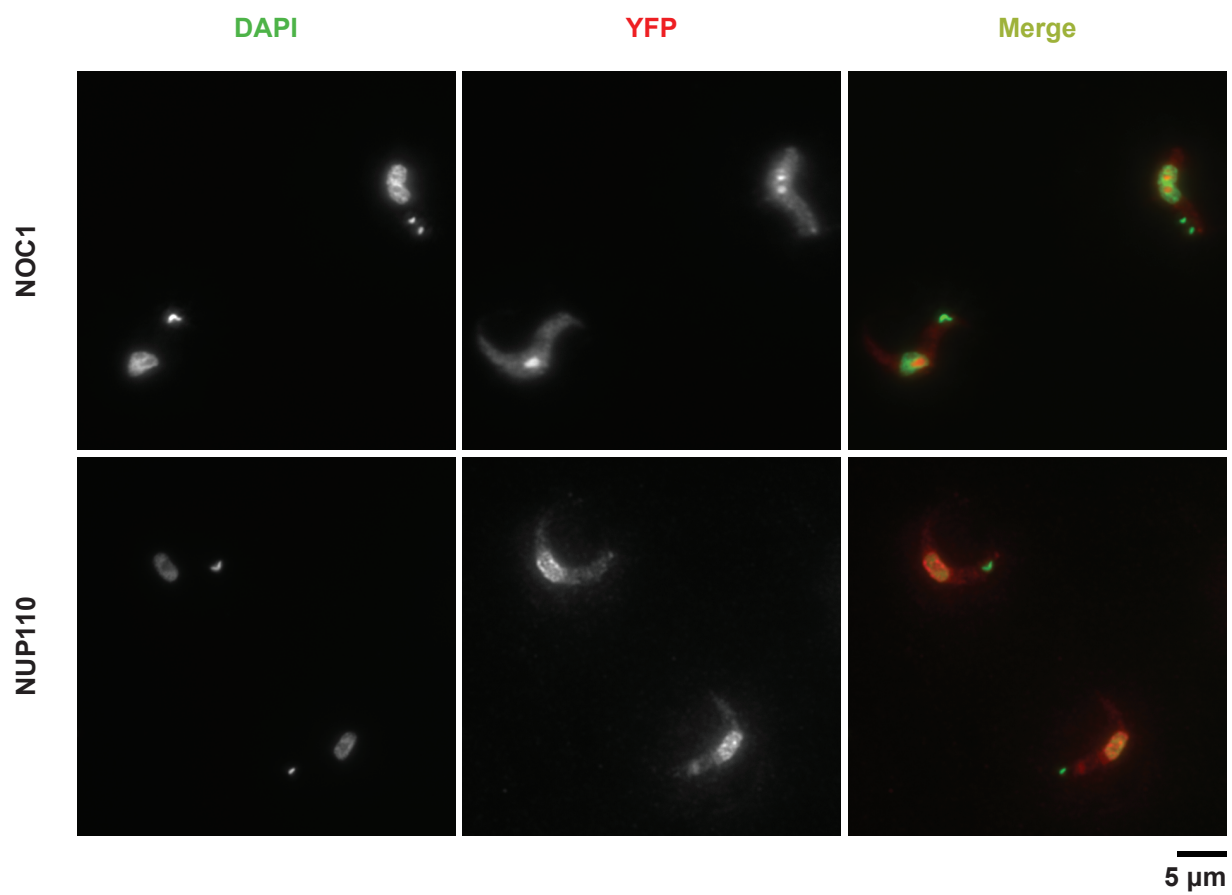
5 μ m











5 μ m

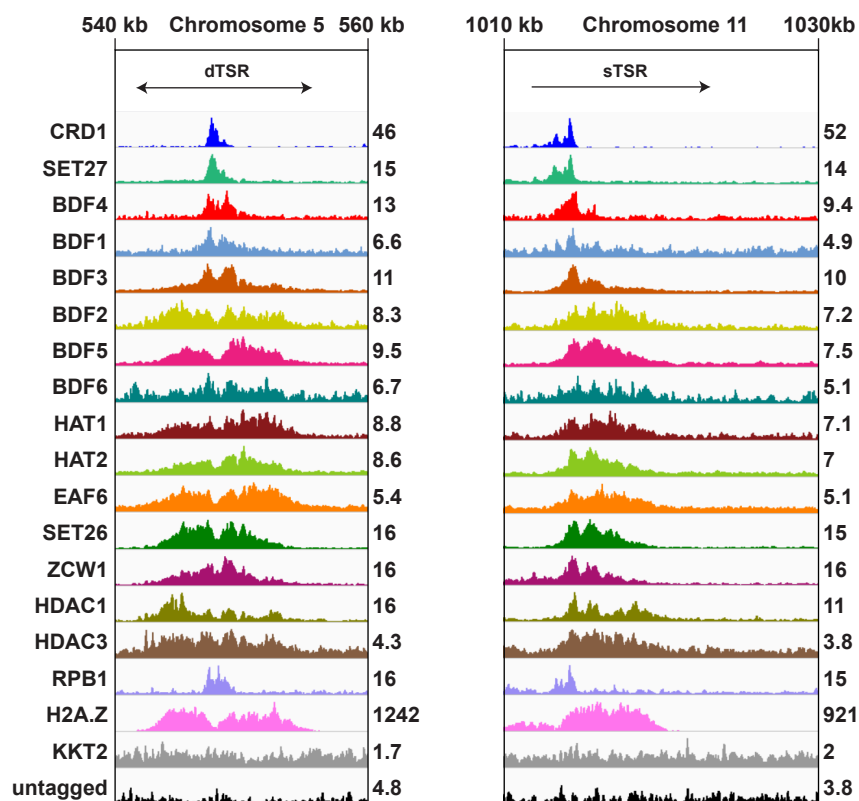
Supplemental Fig. S4. Localization of YFP-tagged proteins at different stages of the cell cycle.

The indicated YFP-tagged proteins expressed in bloodstream Lister 427 cells from their endogenous genomic loci were detected with an anti-GFP primary antibody and an Alexa Fluor 568 labelled secondary antibody (red). Nuclear and kinetoplast (mitochondrial) DNA were stained with DAPI (green). Scale bar = 5 μ m.

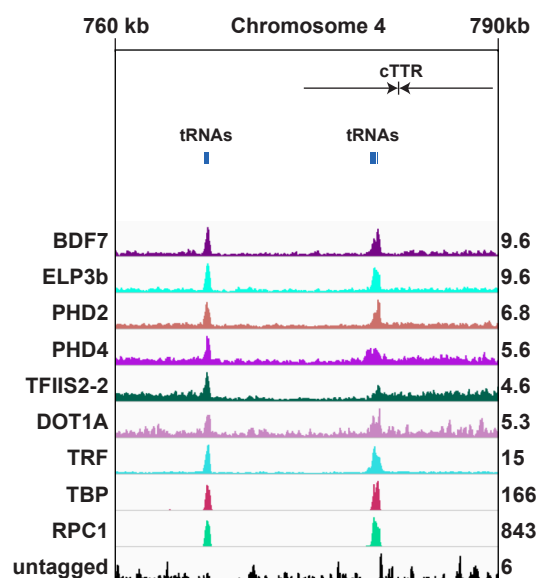
Supplemental Fig. S5. YFP-tagged proteins with no specific genomic enrichment by ChIP-seq.

Tracks show ChIP-seq reads of the indicated proteins for the region of Chromosome 7 shown in Fig. 2A and are scaled separately as reads per million (values shown at the end of each track). One replicate is included for each protein.

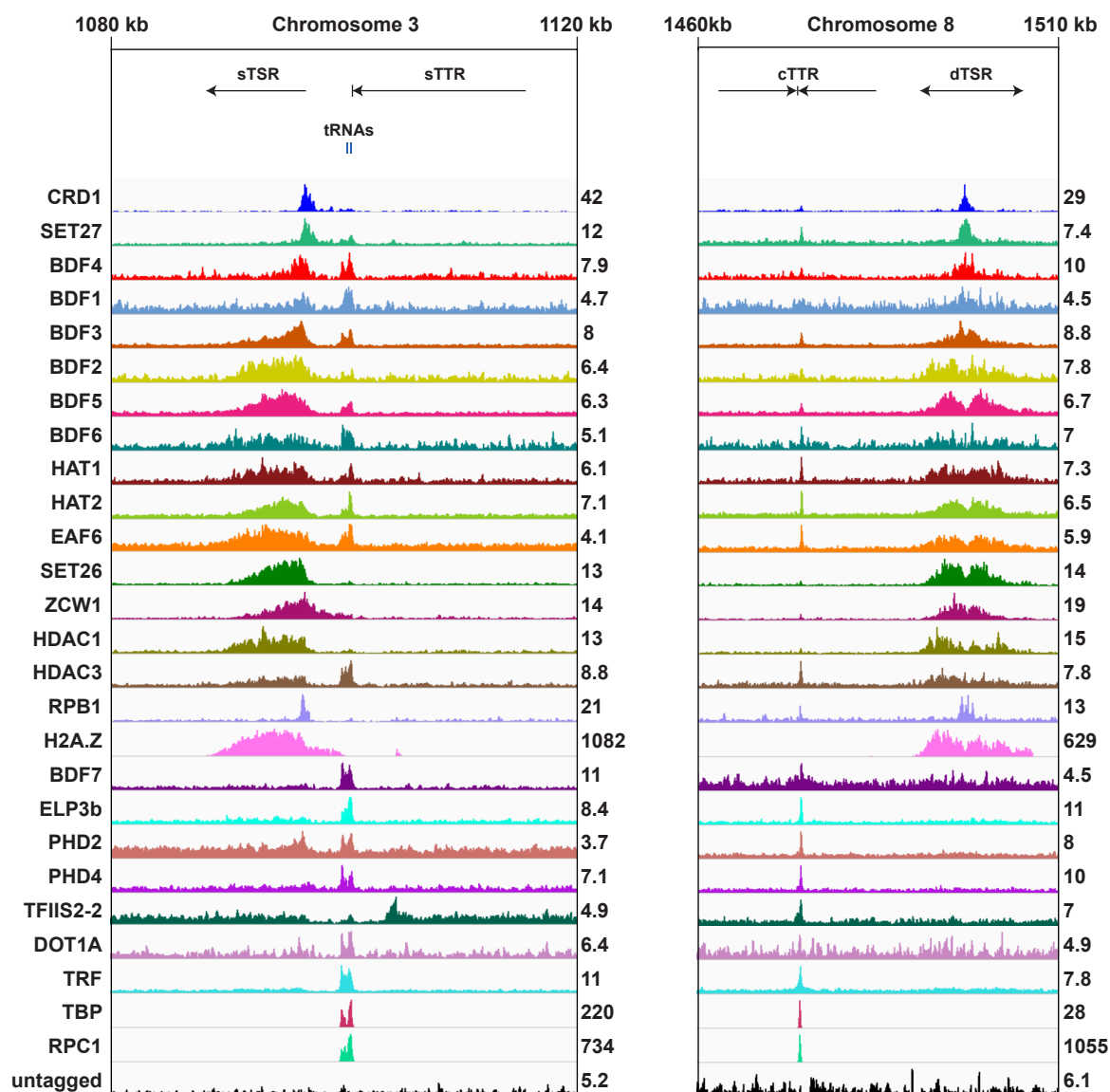
A



B



C



Supplemental Fig. S6. Examples of ChIP-seq enrichment of TSR- and RNAPIII/TTR-associated factors.

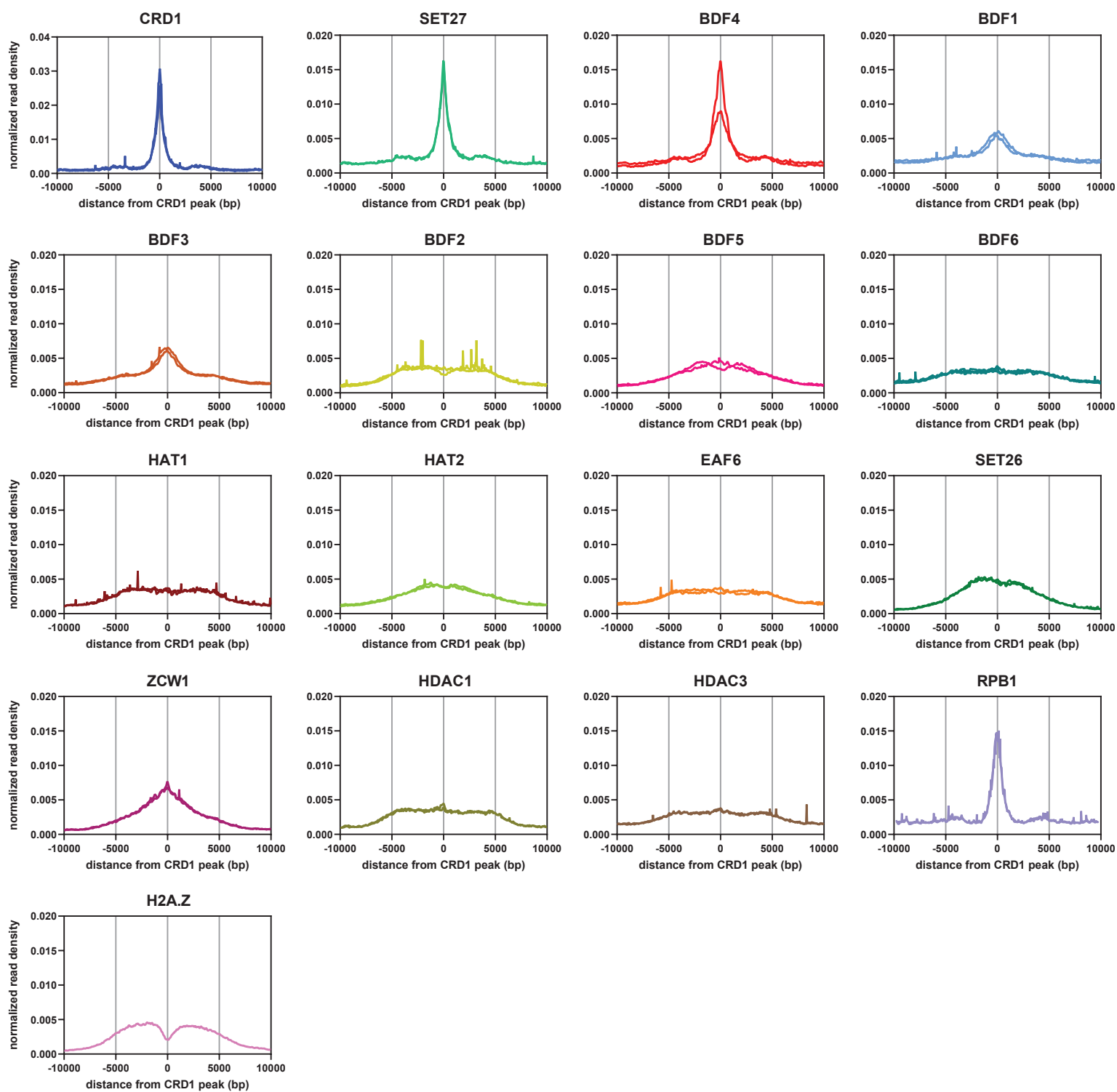
Tracks show one ChIP-seq replicate for each protein and are scaled separately as reads per million (values shown at the end of each track).

A. Left - enrichment at a bidirectional/divergent TSR.

Right - enrichment at a unidirectional/single TSR.

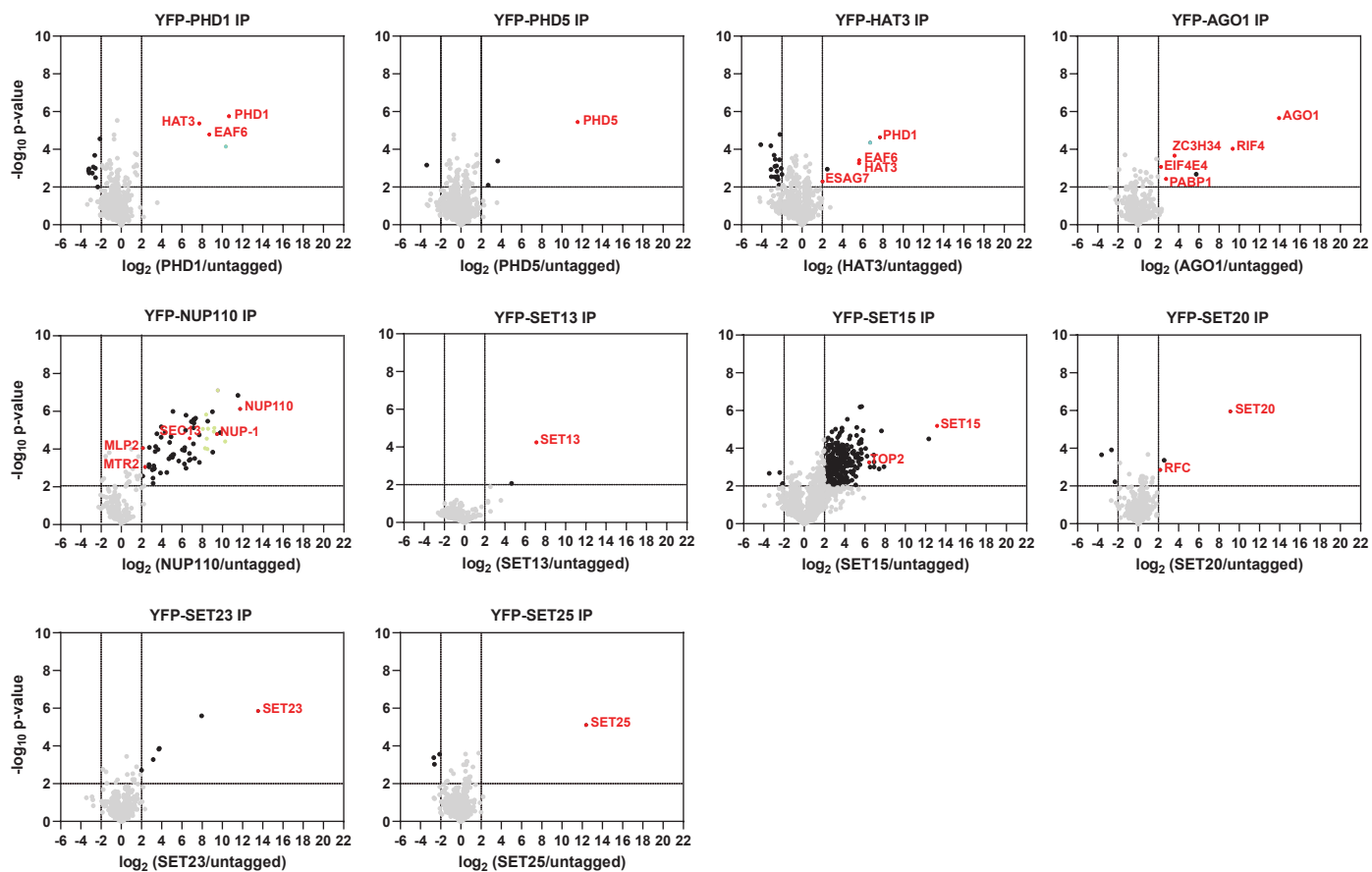
B. Enrichment at two tRNA clusters, one of which overlaps a convergent TTR.

C. Coincidence of peaks of the TSR- and RNAPIII/TTR-associated factors.



Supplemental Fig. S7. Average metagene plots.

For each protein, normalized reads around the CRD1 peak summits were averaged and plotted as density. Plots show separately data from individual ChIP-seq replicates. Note the different scale for CRD1.



Supplemental Fig. S8. Proteomic analysis of affinity selections for several non-chromatin associated proteins.

Data for each plot is based on three biological replicates. Cut-offs used for significance: $\log_2$ (tagged/untagged) > 2 or < -2 and $p < 0.01$ (Student's t -test). Enrichment scores for proteins identified in each affinity selection are presented in Supplemental Table S4. Significantly enriched proteins are indicated by black or coloured dots. Proteins of interest are indicated by red font. Blue dots correspond to some proteins identified also in the BDF6, HAT1 and EAF6 affinity purifications (see Fig. 4A). The ten most significant nucleoporins in the YFP-NUP110 affinity selection are marked in pale green.