Loss-of-function mutations in the dystonia gene THAP1 impair proteasome function by inhibiting PSMB5 expression

Dylan E. Ramage[1], Drew W. Grant[1], Richard T. Timms[1*]

[1]Cambridge Institute of Therapeutic Immunology and Infectious Disease, Department of Medicine, University of Cambridge, Puddicombe Way, Cambridge, CB2 0AW, UK

*Correspondence: rtt20@cam.ac.uk

ABSTRACT

The 26S proteasome is a multi-catalytic protease that serves as the endpoint for protein degradation via the ubiquitin-proteasome system. Proteasome function requires the concerted activity of 33 distinct gene products, but how the expression of proteasome subunits is regulated in mammalian cells remains poorly understood. Leveraging coessentiality data from the DepMap project, here we characterize an essential role for the dystonia gene *THAP1* in maintaining the basal expression of *PSMB5*. PSMB5 insufficiency resulting from loss of THAP1 leads to defects in proteasome assembly, impaired proteostasis and cell death. Exploiting the fact that the toxicity associated with loss of THAP1 can be rescued upon exogenous expression of PSMB5, we define the transcriptional targets of THAP1 through RNA-seq analysis and perform a deep mutational scan to systematically assess the function of thousands of single amino acid THAP1 variants. Altogether, these data identify THAP1 as a critical regulator of proteasome function and suggest that aberrant proteostasis may contribute to the pathogenesis of THAP1 dystonia.

INTRODUCTION

Regulated protein degradation is essential for cellular homeostasis. As the primary route through which the cell achieves selective protein degradation, the ubiquitin-proteasome system (UPS) plays an important role in essentially all critical cellular processes[1]. Proteins destined for degradation are typically identified by E3 ubiquitin ligases, which, following activation of ubiquitin by an E1 activating enzyme and its transfer to an E2 ubiquitin conjugating enzyme, catalyze the attachment of ubiquitin onto substrate proteins[2]. Addition of further ubiquitin moieties generates polyubiquitin chains, which can serve as a potent recognition signal for the 26S proteasome, a multi-catalytic protease. The importance of this pathway is underscored by the fact that dysregulation of the UPS is a hallmark of diseases such as cancer, autoimmunity and neurodegeneration[3].

The 26S proteasome is a large, multi-subunit complex comprising the 20S core particle and two 19S regulatory particles[4]. The regulatory subunits are responsible for the recognition and unfolding of ubiquitinated proteins, which are then threaded into the active site in the core particle formed by two rings of β-subunits[5]. The complex comprises three catalytic subunits: PSMB5 (also known as β5), PSMB6 (β1) and PSMB7 (β2). These exhibit trypsin-like, chymotrypsin-like and caspase-like activities respectively, resulting in the proteolysis of the polypeptide chain into short peptides fragments[6]. Each catalytic subunit harbors a catalytic threonine residue at its N-terminus[6], which is activated following autocatalytic processing of an N-terminal propeptide at a late stage in core particle assembly[7].

By performing genome-wide pooled CRISPR/Cas9 loss-of-function genetic screens across hundreds of cancer cell lines, the Broad Institute's Cancer Dependency Map (DepMap) project aims to systematically catalogue the essentiality of all protein-coding human genes[8]. A

key insight from these data is that whilst the dependency of different cell lines on any one particular gene may vary, genes which function in concert in a biological pathway often exhibit globally similar essentiality patterns[9]. Thus, by measuring gene dependency across hundreds of cell lines, genes exhibiting 'co-essential' relationships can be clustered into modules which may have the power to predict novel functions for genes. Indeed, multiple studies have exploited this dataset to provide new insights into gene function across a range of biological processes[9–14].

Here we leveraged insight from co-essentiality data to characterize an essential role for THAP1 in proteasome function. THAP1 is a ubiquitously expressed transcription factor which achieves sequence-specific DNA binding via an atypical THAP-type zinc finger domain located at its N-terminus[15]. Its target genes remain poorly defined, but THAP1 is thought to play an important role in DNA repair[16], cell cycle progression[17] and oligodendrocyte myelination[18,19]. Homozygous deletion of THAP1 leads to embryonic lethality [20,21]. Notably, a wide variety of autosomal dominant mutations located throughout the THAP1 coding sequence cause an early-onset form of the neurological disorder dystonia (DYT-THAP1, previously known as DYT-6), where progressive loss of motor function leads to sustained involuntary muscle contractions and abnormal posturing[22,23]. However, as the critical targets of THAP1 are poorly characterized, it remains unclear how the THAP1 mutations observed in dystonia patients result in disease.

Here, by interrogating the co-essential relationship between *THAP1* and *PSMB5*, we characterize an essential role for THAP1 in proteasome function. Exploiting a fluorescent reporter knocked into the endogenous *PSMB5* locus, we demonstrate that the co-essential relationship between THAP1 and PSMB5 is explained by an essential role for THAP1 in

activating *PSMB5* expression. THAP1 binds to cognate sites within the *PSMB5* promoter and is required for its basal expression, and hence loss of THAP1 results in insufficient PSMB5 expression, proteasome dysfunction and cell death. Finally, we leveraged our functional reporter assay to perform a deep mutational scan of THAP1, quantifying the activity of thousands of single amino acid variants to define the landscape of THAP1 mutations in dystonia.

RESULTS

The dystonia gene THAP1 exhibits a co-essential relationship with the proteasome subunits PSMB5 and PSMB6

Leveraging co-essentially data from the DepMap project[8], we set out to characterize novel roles for genes involved in the UPS. Focusing on a manually curated set of ~1000 genes implicated in UPS function, we examined co-essential gene relationships derived from genome-wide CRISPR-Cas9 screens across ~1100 different cancer cell lines. Supporting the utility of this approach to identify genetic relationships that are functionally relevant, many of the most significant positive co-essential relationships clustered genes whose products are known to act in multi-protein complexes to facilitate protein degradation **(Fig. S1A)**. For instance, the RNF126 E3 ubiquitin ligase cooperates with BAG6 to target hydrophobic sequences mislocalised to the cytosol for proteasomal degradation[24] **(Fig. 1A)**; Cul2, ElonginB/C and the von Hippel-Lindau (VHL) substrate adaptor comprise a Cullin-RING E3 ubiquitin ligase complex responsible for the degradation of hypoxia-inducible factor (HIF)-1α in normoxia **(Fig. 1B)**[25], and the CTLH complex is a multi-subunit E3 ligase orthologous to the yeast GID complex which degrades gluconeogenic enzymes[26] **(Fig. 1C)**. Furthermore, several of the most significant negative co-essential relationships define E3 ligase-substrate pairs: for example, MDM2 mediates the degradation of p53[27] **(Fig. S1B)** and the Cul4 substrate adaptor AMBRA1 targets cyclin D[28–30] **(Fig. S1C)**.

Our follow-up work focused on the most statistically significant uncharacterized co-essential relationship in the dataset: *THAP1* exhibits a highly significant positive association with both *PSMB5* and *PSMB6* **(Fig. 1D-F)**. THAP1 is a transcription factor which binds DNA in a sequence-specific manner using a THAP-type zinc-finger domain, while *PSMB5* and *PSMB6* encode catalytic subunits of the proteasome core particle. Thus, we set out to test the

hypothesis that the co-essential relationship between *THAP1* and *PSMB5/6* could be explained by an essential role for THAP1 in regulating the expression of catalytic proteasome subunits.

Loss of THAP1 abrogates *PSMB5* transcription

Lentiviral expression of Cas9 and CRISPR sgRNAs targeting *THAP1* in HEK-293T cells was extremely toxic **(Fig. 1G-H)**, consistent with DepMap data which demonstrates that knockout of THAP1 is broadly deleterious across cancer cell lines[8]. However, at day 5 post-transduction, before the onset of significant cell death, we found substantially reduced levels of PSMB5 transcripts by quantitative reverse transcription PCR (qRT-PCR) **(Fig. 1I)**. In contrast, we observed no reduction in the expression of either PSMB6 or PSMB7, the other catalytic subunits of the proteasome **(Fig. 1I)**. Concordantly, we also observed reduced abundance of PSMB5 protein as assessed by immunoblot **(Fig. 1J)**. Thus, these data suggest that THAP1 is required to maintain basal levels of *PSMB5* transcription.

Lethality resulting from loss of THAP1 can be rescued by exogenous PSMB5

Next, we sought to test the hypothesis that the essentiality of THAP1 is due to its role in activating *PSMB5* expression. Should this be the case, we reasoned that, irrespective of any reduction in the expression of endogenous PSMB5, an exogenous source of PSMB5 should rescue cell viability upon *THAP1* ablation. Strikingly, unlike their wild-type counterparts, HEK-239T cells transduced with a lentiviral vector expressing PSMB5 did not display any significant growth defect following CRISPR/Cas9-mediated targeting of *THAP1* **(Fig. 2A and Fig. S2A)**. Exogenous expression of PSMB6, however, was incapable of rescuing viability following *THAP1* ablation **(Fig. 2A and Fig. S2B)**. Thus, the toxicity that results from loss of THAP1 is due to insufficient PSMB5 expression, explaining the molecular basis for their co-essential relationship. In contrast, we found no evidence to support a direct relationship

between *THAP1* and *PSMB6*, suggesting that their co-essential relationship arises indirectly through their shared relationship with *PSMB5*.

DepMap data demonstrates that disruption of THAP1 is broadly lethal across cancer cell lines **(Fig. 2B)**. Thus, to generalize our finding that the essential requirement for THAP1 is to facilitate PSMB5 expression, we ablated THAP1 in three additional human cell lines: HeLa, A549 and THP-1. Mirroring our findings in HEK-293T cells, in A549 and HeLa we found that the toxicity observed upon loss of THAP1 could be ameliorated upon exogenous expression of PSMB5 **(Fig. 2C-D)**. THP-1 cells, however, did not exhibit reduced viability following *THAP1* disruption **(Fig. 2E)**. This prompted us to examine in more detail the nature of the cell lines in which THAP1 is not essential, which were strikingly enriched ($P < 1x10^{-9}$) for immune cells ('myeloid' or 'lymphoid' as defined by DepMap). Considering that the immunoproteasome is constitutively expressed by many immune cells[31,32], we reasoned that expression of PSMB8, the analogous counterpart of PSMB5 in the immunoproteasome, might relieve the essential requirement for THAP1. Indeed, there is a strong correlation between the essentiality of THAP1 and PSMB8 expression levels as measured by RNA-seq, wherein the cell lines in which THAP1 knockout has little or no impact on viability express the highest levels of PSMB8 **(Fig. 2F)**. Indeed, we found that THP-1 cells expressed high levels of PSMB8 by qRT-PCR **(Fig. 2G)**. We further validated these conclusions in HEK-293T cells, where, like PSMB5, exogenous expression of PSMB8 maintained the viability of THAP1 knockout cells **(Fig. 2H and Fig. S2C)**. Thus, sustained expression of either PSMB5 or its immunoproteasome counterpart PSMB8 can rescue the toxicity associated with loss of THAP1.

A fluorescent reporter at the endogenous *PSMB5* locus monitors THAP1 activity in live cells

We extended these findings by knocking-in a fluorescent reporter to the endogenous *PSMB5* locus, enabling us to monitor PSMB5 expression in live cells. Following transfection of HEK-293T cells with Cas9, an sgRNA targeting the transcriptional start site of *PSMB5* and a homology donor vector encoding the green fluorescent protein (GFP) variant mClover3[33] followed by a 2A peptide **(Fig. 3A)**, we were readily able to establish a population of cells (~10%) which were stably GFP-positive **(Fig. 3B)**. Single cell clones isolated from the GFP-positive population **(Fig. 3C)** harbored GFP at the intended site as validated by PCR from genomic DNA **(Fig. S2D)**. Furthermore, lentiviral expression of shRNAs targeting PSMB5 resulted in a reduction in GFP expression **(Fig. S2E)**, validating that the reporter clones could be used to quantitatively assess PSMB5 expression.

We exploited our findings above to generate viable THAP1 knockout (KO) reporter cells. Following CRISPR/Cas9-mediated disruption of THAP1 in P_{PSMB5}-GFP reporter cells (sustained by exogenous expression of PSMB5) **(Fig. 3D)**, we isolated single cell clones from the GFP^{dim} population by fluorescence-activated cell sorting (FACS) **(Fig. 3E)**. Disruption of the *THAP1* locus was confirmed by PCR from genomic DNA followed by Sanger sequencing **(Fig. S2F)**. Using primers specific to the 3' untranslated region of PSMB5 to allow for selective detection of the endogenous transcript, we confirmed a substantial reduction in PSMB5 expression in two THAP1 KO clones by qRT-PCR **(Fig. 3F)**. However, our attempts to validate efficient knockout of THAP1 by immunoblot were hampered by the paucity of effective commercial antibodies. In particular, we found that the Proteintech antibody (12584-1-AP) used in multiple previous studies detected a prominent band running around the expected molecular weight (~25 kDa), but whose abundance was not affected upon CRISPR/Cas9 targeting of *THAP1* **(Fig. 3G)**. However, this antibody could readily detect exogenous THAP1

as a separate band that migrated just slightly slower, and, upon prolonged exposure, was able to detect endogenous THAP1 in control cells but not in the THAP1 knockout clones **(Fig. 3G)**.

THAP1 acts through cognate binding sites located within the *PSMB5* promoter

Next, we sought to determine how THAP1 might regulate PSMB5 expression. In support of a direct effect, chromatin immunoprecipitation followed by sequencing (ChIP-seq) data from the ENCODE project[34] revealed THAP1 occupancy immediately upstream of the *PSMB5* transcription start site (TSS) **(Fig. 4A)**. Despite THAP1 binding to thousands of gene promoters **(Table S1)**, this property is not shared among the genes encoding proteasome subunits: *PSMD8* is the only other gene to exhibit THAP1 occupancy **(Fig. 4A)**. THAP1 contains a THAP-type zinc finger domain which mediates sequence-specific DNA binding, and competitive EMSA experiments[35] have defined the consensus binding sequence ("THABS" motif) as TNNNGGCA (where N represents any nucleotide) **(Fig. 4B)**. Strikingly, examination of the PSMB5 proximal promoter region revealed two perfect matches within 200 bp of the TSS, and a third near-perfect match a further ~500 bp upstream **(Fig. 4C).** Together, these data suggest that THAP1 binds cognate motifs in the PSMB5 promoter to activate its transcription.

We examined this hypothesis by engineering a lentiviral reporter system in which ~1 kb of the PSMB5 proximal promoter was placed upstream of GFP **(Fig. 4D)**. Single copy expression of this reporter construct in HEK-293T cells resulted in robust GFP expression **(Fig. S3A)**. This appeared to be due in part to the activity of THAP1, as combined deletion of all three THABS motifs from the PSMB5 promoter ("ΔTHABS") resulted in decreased GFP expression **(Fig. 4E)**. Importantly, CRISPR-mediated ablation of *THAP1* (performed following the introduction of exogenous PSMB5 to maintain cell viability) reduced GFP expression from

the reporter construct driven by the wild-type PSMB5 promoter, but did not further reduce expression from the PSMB5 promoter lacking all THABS sites **(Fig. 4F)**.

To investigate the relative importance of the three THABS motifs, we created an additional panel of mutant constructs in which each THABS motif was individually deleted. These data pointed to a critical role for site 2, as only the ΔTHABS2 construct exhibited decreased GFP expression relative to the level of the ΔTHABS construct **(Fig. S3B)**. Moreover, the ΔTHABS2 construct was the only one unaffected following ablation of THAP1, whereas a marked reduction in GFP expression was observed in cells expressing ΔTHABS1 and ΔTHABS3 **(Fig. S3C)**. Thus, THAP1 binding to a cognate motif (THABS2) immediately upstream of the PSMB5 TSS appears critical for PSMB5 expression.

Loss of THAP1 impairs proteasome function

As *PSMB5* encodes one of the three catalytic subunits of the constitutive 20S proteasome core particle, we set out to test the hypothesis that the toxicity associated with loss of THAP1 was due to proteasome dysfunction. First, we examined whether the catalytic activity of PSMB5 was required to sustain cell viability upon *THAP1* ablation. Whereas HEK-293T cells expressing wild-type PSMB5 did not exhibit any appreciable growth defect upon disruption of *THAP1*, cells expressing a catalytically-inactive PSMB5 mutant were not viable under these conditions **(Fig. 5A)**. Second, as PSMB5 is critical to facilitate the integration of the other catalytic β subunits into the 20S core particle during proteasome assembly[36–38], we assessed whether loss of THAP1 resulted in defects in proteasome assembly. As a result of impaired autocatalytic cleavage of their N-terminal propeptide, an inability of the catalytic subunits to incorporate into the 20S core particle causes an accumulation of the immature proteins which can be detected by immunoblot[38]. Supporting the idea of defective proteasome

assembly in the absence of THAP1, we found decreased abundance of the mature forms of PSMB6 and PSMB7, concomitant with the accumulation of immature (uncleaved) species (indicated by asterisks) which were absent from control cells **(Fig. 5B)**. Finally, we reasoned that proteasome dysfunction upon THAP1 loss should result in the stabilization of short-lived proteins. Exploiting the Global Protein Stability (GPS) two-color lentiviral reporter system[39] **(Fig. 5C)**, we found that CRISPR-mediated ablation of THAP1 stabilized two exogenous GFP-degron fusion proteins to a similar degree as shRNA-mediated knockdown of PSMB5 **(Fig. 5D)**. Similarly, *THAP1* disruption also resulted in increased abundance of endogenous HIF-1α, which is constitutively degraded by the proteasome in normoxia[25,40] **(Fig. 5E)**. Altogether, these data support a model whereby the death of cells lacking THAP1 is caused by defective proteasome function resulting from inadequate expression of PSMB5.

Defining the transcriptional targets of THAP1

The key transcriptional targets of THAP1 remain poorly defined, hampering our ability to understand the functional consequences of THAP1 mutations in disease. Armed with the knowledge that exogenous PSMB5 expression ameliorates the toxicity associated with THAP1 loss, we reasoned that we could leverage our findings to assess the impact of THAP1 deletion on the transcriptome. To avoid potential artefacts resulting from the analysis of single cell clones, we purified a population of THAP1 knockout cells. Following the CRISPR-mediated ablation of *THAP1* in PSMB5-GFP knock-in reporter cells (overexpressing PSMB5 to ensure viability), we isolated GFP^{dim} cells by FACS and performed RNA-seq analysis **(Fig. 6A)**. After discounting genes exhibiting differential expression between untransduced cells and cells expressing control sgRNAs, we identified 277 genes (220 downregulated, 57 upregulated) whose expression was significantly altered (FDR < 0.001, fold-change > 2) upon THAP1 knockout **(Fig. 6B and Table S2)**. Supporting the veracity of the dataset, the most significantly

downregulated gene was Shieldin complex subunit 1 (*SHLD1*, previously known as *C20orf196*), consistent with recent findings describing a role for THAP1 in DNA double strand break repair choice[16]. Furthermore, although the requirement for exogenous expression of PSMB5 precluded its identification as a differentially expressed gene, the abundance of intronic reads mapping to the endogenous PSMB5 locus was greatly reduced in the THAP1 knockout cells **(Fig. S4A-B)**.

To identify direct transcriptional targets of THAP1, we cross-referenced the differentially expressed genes identified through RNA-seq with THAP1 binding sites as defined by ChIP-seq. Among the differentially expressed genes, 42 exhibited THAP1 occupancy in their proximal promoter **(Fig. 6C)**; of these 42 direct targets, 19 were downregulated upon loss of THAP1 while 23 were upregulated, suggesting that THAP1 has the potential to act as either a repressor or activator of transcription depending on the genomic context. However, the primary role of THAP1 appeared to be as an activator, with several of its direct targets exhibiting marked downregulation in its absence **(Fig. 6B-C)**. We found no significant functional enrichment among the differentially expressed genes through GO term analysis, but their promoter sequences were enriched for transcription factor binding motifs for both THAP1 and YY1, a known THAP1 co-factor[16,41,42] **(Fig. 6D and Fig. S4C)**. None of these genes are currently associated with dystonia, but they include *ECH1*, an enzyme involved in fatty acid metabolism that has been previously identified as a THAP1 target[42], and *METTL3*, the N^6-methyladenosine methyltransferase, which is an attractive therapeutic target in cancer[43]. Across five of the THAP1 target genes that exhibited the greatest degree of downregulation upon loss of THAP1, we further validated these findings by qRT-PCR **(Fig. 6E-I)**. Altogether, these data define the genes directly targeted by the transcription factor activity of THAP1.

A deep mutagenic scan defines the landscape of THAP1 mutations in Dystonia

A wide range of autosomal dominant mutations distributed throughout the THAP1 coding sequence give rise to DYT-THAP1 dystonia[44–50], an early-onset neurological disorder characterized by involuntary muscle contractions and movements causing abnormal and painful posturing. Thus, we sought to exploit our P_{PSMB5}-GFP knock-in reporter clone to assess the functional impact of DYT-THAP1 mutations. Genetic complementation of THAP1 KO cells with wild-type THAP1 did result in a restoration in P_{PSMB5}-GFP expression, although this effect was partial and did not restore GFP fluorescence to the levels observed in the parental cells **(Fig. S5A)**. However, this assay was sufficiently sensitive to report on THAP1 activity, as expression of an inactive THAP1 mutant unable to bind DNA (C5A, which abrogates zinc chelation by the zinc finger motif[15]) did not restore P_{PSMB5}-GFP expression **(Fig. S5A)**.

With the goal of globally defining how mutations in THAP1 affect its function, we leveraged our phenotypic assay in P_{PSMB5}-GFP knock-in reporter cells to carry out a deep mutational scan. Through microarray oligonucleotide synthesis, we generated a library of mutant constructs in which each residue of THAP1 (with the exception of the initiator methionine, 212 amino acids total) was systematically replaced with all of the other 20 possible amino acids **(Fig. 7A-B)**. The resulting site-saturation mutagenesis library was packaged into lentiviral particles and introduced into THAP1 KO P_{PSMB5}-GFP reporter cells at low multiplicity of infection, ensuring single-copy expression. We then used FACS to partition the population into GFP^{dim} cells, in which no restoration of P_{PSMB5}-GFP expression was observed, and GFP^{bright} cells, in which P_{PSMB5}-GFP expression was restored, and quantified the THAP1 variants present in each population by Illumina sequencing **(Fig. 7B)**.

After an initial filtering step to remove variants with low read counts, we recovered data for 4002 of the 4240 possible single amino acid variants (94.3%). Overall, we observed high concordance between mutant performance across two replicate experiments **(Fig. S5B)**; however, we discarded 179 mutants which exhibited discordant behavior between the two replicate experiments, leaving a total of 3823 (90.2%) variants for analysis **(Table S3)**. The results are summarized as a heatmap in **Fig. 7C**, with the data normalized such that the mean performance of all the control (wild-type) constructs is centered at 1; thus, the darker the red color the more deleterious the impact of the mutation on THAP1 function, whereas blue cells indicate mutations which may enhance the THAP1-mediated activation of P_{PSMB5}-GFP expression.

We evaluated the quality of the dataset in several ways. First, we considered residues essential for zinc chelation and hence folding of the zinc finger motif[15]: C5, C10, C54 and H57. These critical residues were uniformly essential for THAP1 activity, as mutation to any other residue prevented activation of the P_{PSMB5}-GFP reporter **(Fig. 8A)**. Moreover, mutations across all residues previously determined to be important for DNA binding through biochemical assays[15] were extremely deleterious **(Fig. 8B)**. Second, the global landscape of THAP1 activity correlated well with the predicted structure of THAP1 **(Fig. 8C-D)**. In particular, mutation of most residues in the two structured regions, the THAP-type zinc finger (residues 2-81) and the predicted coiled-coil domain (residues 139-191) abrogated transcriptional activity, whereas most mutations targeting the unstructured central linker (residues 82-138) and C-terminus (residues 192-213) did not impair transcriptional activity **(Fig. 7C and Fig. 8E)**. A notable exception, however, was the DHNY motif (residues 134-137) lying at the end of the central unstructured linker, which was absolutely critical for THAP1 function **(Fig. 7C)**. Interestingly, this motif has been identified as the binding site for HCFC1[51], an essential cofactor for the

374 THAP1-mediated activation of *SHLD*[16]. Thus, the interaction with HCFC1 is also likely to be
375 critical for the THAP1-mediated activation of *PSMB5*. Indeed, we confirmed that deletion of
376 the DHNY motif and the coiled-coil domain, but not the disordered C-terminus, abolished the
377 ability of THAP1 to activate the P_{PSMB5}-GFP reporter **(Fig. 8F)**.
378
379 Many of the THAP1 mutations identified in dystonia patients remain of uncertain
380 significance[50]. Thus, we examined the utility of this dataset in classifying the functional effects
381 of THAP1 variants identified clinically **(Fig. 8G)**. The majority of these mutations strongly
382 impaired the ability of THAP1 to activate expression of the P_{PSMB5}-GFP reporter, consistent
383 with the prevailing view that disease-causing mutations represent loss-of-function alleles[19,41,50].
384 However, some mutants exhibited activity at or approaching the level of the wild-type protein,
385 suggesting that they might represent benign variants. To verify that the screen results could be
386 faithfully recapitulated in individual experiments, we selected eight patient mutations predicted
387 to abolish THAP1 activity (A7D, R13H, K24E, P26R, H57N, L72R, F81L and N136S) and
388 compared their performance to five mutants predicted not to affect THAP1 activity (I80V,
389 C83R, M143V, A166T and D192N). Validating the screen results, the eight inactive mutants
390 exhibited little or no ability to activate P_{PSMB5}-GFP reporter expression **(Fig. 8H and Fig. S5C-**
391 **E)**, whereas the five active mutants exhibited similar performance to wild-type THAP1 **(Fig.**
392 **8H)**. Altogether, these data illustrate structure-function relationships for THAP1 at high
393 resolution, enabling the functional classification of clinical THAP1 mutations.
394
395
396
397
398

DISCUSSION

Co-essential relationships identified through the DepMap project[8] represent a rich resource to characterize gene function. Here we explain the co-essential relationship between THAP1 and PSMB5 by demonstrating that THAP1 is essential for the basal expression of PSMB5. Insufficient PSMB5 expression resulting from loss of THAP1 results in proteasome dysfunction and cell death, which can be rescued through exogenous expression of PSMB5. We exploit this finding to generate viable THAP1 knockout cells and hence identify transcriptional targets of THAP1 by RNA-seq. Finally, leveraging a phenotypic assay to systematically assess the activity of THAP1 mutants at the endogenous *PSMB5* locus, we define the transcriptional activity of THAP1 mutants found in dystonia patients. Overall, these data identify THAP1 as a novel regulator of proteasome function and suggest that aberrant proteostasis could be a factor underlying the pathogenesis of THAP1 dystonias.

Two major players are known to regulate the expression of proteasome subunits. In yeast the transcription factor Rpn4 is responsible for both the basal expression of proteasome subunits, and, under conditions of proteasome insufficiency, feedback induction of proteasome subunit expression[52]. Rpn4 functions as part of a negative feedback loop that monitors proteasome activity: in unstressed cells Rpn4 is constitutively degraded, but it is rapidly stabilized upon proteasome dysfunction[53]. In mammalian cells Nrf1 acts in a similar manner to induce the expression of proteasome subunits under conditions of proteasome insufficiency[54,55], but appears not to have a major role in their basal expression[54]. Several other transcription factors have been implicated in proteasome gene expression[56–58], including NF-Y which regulates a set of proteasome genes which carry a CCAAT box motif in their promoters[58], but the factors which maintain the basal expression of proteasome subunits in human cells remain largely unknown.

Here we characterize THAP1 as an additional regulator of proteasome gene expression. In contrast to the master regulators Rpn4 and Nrf1, THAP1 appears to regulate only PSMB5. Why might THAP1 have evolved to exclusively regulate the expression of one single proteasome subunit? We speculate that perhaps there is a physiological circumstance wherein the downregulation of PSMB5 expression is beneficial, which could be achieved through the conditional inactivation of THAP1 activity. For example, it is plausible that transcriptional downregulation of *PSMB5* concomitant with upregulation of *PSMB8* might be beneficial upon viral infection, when immunoproteasomes are favored to increase the production of antigenic peptides[59,60], or during thymic development when PSMB11 (β5t) is incorporated in preference to PSMB5 and PSMB8 into the thymoproteasome[61]. Another interesting question for future studies will be to examine whether the expression of other individual proteasome subunits (such as PSMB6 and PSMB7) is also subject to specific regulatory mechanisms.

The mechanisms through which THAP1 mediates its effects on gene expression remain unclear. THAP1 does not possess an obvious activator or repressor domain, and so it is likely that it acts through the recruitment of co-factors to target genes. A complex of THAP1 with YY1 and HCFC1 has previously been shown to mediate activation at the *SHLD* promoter[16], and the F81L dystonia mutation is thought to disrupt YY1 binding and hence impair THAP1-mediated transcriptional activation[42]. Our data offer support to this notion: YY1 binding motifs were strongly enriched amongst the direct targets of THAP1 identified by RNA-seq, and an intact HCFC1-binding motif was critical for THAP1-mediated activation of the P_{PSMB5}-GFP reporter. However, ChIP-seq reveals thousands of THAP1 binding sites in promoters across the genome, many of which colocalize with HCFC1 and YY1[16], and yet we observed relatively few transcriptional changes by RNA-seq. Thus, we speculate that other factors must be

449 involved in determining whether THAP1 binding alters transcriptional activity. Our P_{PSMB5}-
450 GFP knock-in reporter cells may therefore serve as a useful resource for further genetic
451 interrogation of this pathway; for example, genome-wide CRISPR screens may identify
452 additional genes required for the THAP1-mediated activation of *PSMB5*.
453
454 We leveraged our genetic reporter to characterize the impact of single amino acid
455 variants on the ability of THAP1 to activate the expression of *PSMB5*. This dataset, covering
456 over 90% of all possible single amino acid variants, represents a rich resource for functional
457 classification of THAP1 mutations. Specifically, these data strongly support the notion that
458 disease-causing mutations in THAP1 generate loss-of-function alleles which are unable to
459 regulate target gene expression: 84% of the missense mutations in THAP1 identified in
460 dystonia patients exhibited performance at <50% of the wild-type protein. Thus, we propose
461 that the mutations which do not impair THAP1 activity are likely to represent benign variants.
462 Indeed, the clinical evidence supporting a pathogenic role for some THAP1 mutations remains
463 equivocal[50]; for example, a previous study concluded that the I80V mutation was very likely
464 to be benign, supported by the conservative nature of the substitution and the lack of any
465 functional defect in a reporter assay[50].
466
467 How does loss of THAP1 function result in Dystonia? As the most plausible
468 explanation is that dysregulated expression of one or more of its target genes leads to disease[62],
469 our data advance progress towards answering this question in two ways. First, the identification
470 of THAP1 as a critical activator of PSMB5 expression suggests that proteasome dysfunction
471 could underlie the pathogenesis of DYT-THAP1. Second, by exploiting exogenous PSMB5
472 expression to generate viable THAP1 knockout cells, we were able to rigorously identify
473 additional direct transcriptional targets of THAP1. However, as none of these genes are

currently associated with dystonia and aberrant proteostasis is a feature of many neurological disorders[63,64], these data highlight proteasome dysfunction as a candidate pathogenic mechanism underlying THAP1 dystonias.

ACKNOWLEDGEMENTS

We are grateful to Gabriela Grondys-Kotarba and Reiner Schulte at the Cambridge Institute for Medical Research Flow Cytometry Core Facility and to the NIHR Cambridge BRC Cell Phenotyping Hub. This work was supported by an Academy of Medical Sciences Springboard Grant (SBF007\100019) and an Isaac Newton Trust/Wellcome Trust ISSF/University of Cambridge Joint Research Grant to R.T.T. R.T.T. is a Pemberton-Trinity Fellow.

AUTHOR CONTRIBUTIONS

D.E.R. and R.T.T. conceived the study, and, along with D.W.G., performed the experiments and analyzed the data. D.E.R. and R.T.T. wrote the manuscript.

DATA AVAILABILITY

THAP1 RNA-seq data has been deposited at GEO (GSE264536) and Illumina sequencing data from the THAP1 deep mutagenic scan are available at SRA (PRJNA1102672). THAP1 ChIP-seq data was obtained from GEO (GSM803408). DepMap datasets are publicly available at https://depmap.org/portal/download/all/. Additional data and/or reagents that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Rape, M. Ubiquitylation at the crossroads of development and disease. *Nature Reviews Molecular Cell Biology 2017 19:1* **19**, 59–70 (2017).
2. Kwon, Y. T. & Ciechanover, A. The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy. *Trends Biochem Sci* **42**, 873–886 (2017).
3. Popovic, D., Vucic, D. & Dikic, I. Ubiquitination in disease pathogenesis and treatment. *Nat Med* **20**, 1242–1253 (2014).
4. Murata, S., Yashiroda, H. & Tanaka, K. Molecular mechanisms of proteasome assembly. *Nature Reviews Molecular Cell Biology 2009 10:2* **10**, 104–115 (2009).
5. Collins, G. A. & Goldberg, A. L. The Logic of the 26S Proteasome. *Cell* **169**, 792–806 (2017).
6. Tanaka, K. The proteasome: Overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci* **85**, 12 (2009).
7. Chen, P. & Hochstrasser, M. Autocatalytic subunit processing couples active site formation in the 20S proteasome to completion of assembly. *Cell* **86**, 961–972 (1996).
8. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e16 (2017).
9. Wainberg, M. *et al.* A genome-wide atlas of co-essential modules assigns function to uncharacterized genes. *Nature Genetics 2021 53:5* **53**, 638–649 (2021).
10. McDonald, E. R. *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170**, 577-592.e10 (2017).
11. Wang, T. *et al.* Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **168**, 890-903.e15 (2017).
12. Pan, J. *et al.* Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens. *Cell Syst* **6**, 555-568.e7 (2018).
13. Rauscher, B. *et al.* Toward an integrated map of genetic interactions in cancer cells. *Mol Syst Biol* **14**, (2018).
14. Kim, E. *et al.* A network of human functional gene interactions from knockout fitness screens in cancer cells. *Life Sci Alliance* **2**, (2019).
15. Bessière, D. *et al.* Structure-Function Analysis of the THAP Zinc Finger of THAP1, a Large C2CH DNA-binding Module Linked to Rb/E2F Pathways. *Journal of Biological Chemistry* **283**, 4352–4363 (2008).
16. Shinoda, K. *et al.* The dystonia gene THAP1 controls DNA double-strand break repair choice. *Mol Cell* **81**, 2611-2624.e10 (2021).
17. Cayrol, C. *et al.* The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. *Blood* **109**, 584–594 (2007).
18. Yellajoshyula, D. *et al.* The DYT6 Dystonia Protein THAP1 Regulates Myelination Within The Oligodendrocyte Lineage. *Dev Cell* **42**, 52 (2017).
19. Aguilo, F. *et al.* THAP1: Role in Mouse Embryonic Stem Cell Survival and Differentiation. *Stem Cell Reports* **9**, 92 (2017).
20. Ortiz-Virumbrales, M. *et al.* Dystonia type 6 gene product Thap1: identification of a 50 kDa DNA-binding species in neuronal nuclear fractions. *Acta Neuropathol Commun* **2**, (2014).
21. Ruiz, M. *et al.* Abnormalities of motor function, transcription and cerebellar structure in mouse models of THAP1 dystonia. *Hum Mol Genet* **24**, 7159–7170 (2015).
22. Almasy, L. *et al.* Idiopathic torsion dystonia linked to chromosome 8 in two Mennonite families. *Ann Neurol* **42**, 670–673 (1997).

23. Lange, L. M. *et al.* Genotype–Phenotype Relations for Isolated Dystonia Genes: MDSGene Systematic Review. *Movement Disorders* **36**, 1086–1103 (2021).
24. Rodrigo-Brenni, M. C., Gutierrez, E. & Hegde, R. S. Cytosolic Quality Control of Mislocalized Proteins Requires RNF126 Recruitment to Bag6. *Mol Cell* **55**, 227 (2014).
25. Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271–275 (1999).
26. Hämmerle, M. *et al.* Proteins of Newly Isolated Mutants and the Amino-terminal Proline Are Essential for Ubiquitin-Proteasome-catalyzed Catabolite Degradation of Fructose-1,6-bisphosphatase of Saccharomyces cerevisiae. *Journal of Biological Chemistry* **273**, 25000–25005 (1998).
27. Momand, J., Zambetti, G. P., Olson, D. C., George, D. & Levine, A. J. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. *Cell* **69**, 1237–1245 (1992).
28. Chaikovsky, A. C. *et al.* The AMBRA1 E3 ligase adaptor regulates the stability of cyclin D. *Nature* **592**, 794–798 (2021).
29. Maiani, E. *et al.* AMBRA1 regulates cyclin D to guard S-phase entry and genomic integrity. *Nature 2021 592:7856* **592**, 799–803 (2021).
30. Simoneschi, D. *et al.* CRL4AMBRA1 is a master regulator of D-type cyclins. *Nature* **592**, 789 (2021).
31. Eleuteri, A. M., Kohanski, R. A., Cardozo, C. & Orlowski, M. Bovine Spleen Multicatalytic Proteinase Complex (Proteasome): REPLACEMENT OF X, Y, AND Z SUBUNITS BY LMP7, LMP2, AND MECL1 AND CHANGES IN PROPERTIES AND SPECIFICITY. *Journal of Biological Chemistry* **272**, 11824–11831 (1997).
32. Noda, C., Tanahashi, N., Shimbara, N., Hendil, K. B. & Tanaka, K. Tissue Distribution of Constitutive Proteasomes, Immunoproteasomes, and PA28 in Rats. *Biochem Biophys Res Commun* **277**, 348–354 (2000).
33. Bajar, B. T. *et al.* Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Scientific Reports 2016 6:1* **6**, 1–12 (2016).
34. Dunham, I. *et al.* An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **489**, 57 (2012).
35. Marzo, M., Liu, D., Ruiz, A. & Chalmers, R. Identification of multiple binding sites for the THAP domain of the Galileo transposase in the long terminal inverted-repeats. *Gene* **525**, 84-89 (2013).
36. Li, X., Li, Y., Arendt, C. S. & Hochstrasser, M. Distinct Elements in the Proteasomal 5 Subunit Propeptide Required for Autocatalytic Processing and Proteasome Assembly *. *Journal of Biological Chemistry* **291**, 1991-2003 (2016).
37. Budenholzer, L., Cheng, C. L., Li, Y. & Hochstrasser, M. Proteasome Structure and Assembly. *J Mol Biol* **429**, 3500 (2017).
38. Shi, C. X. *et al.* Proteasome subunits differentially control Myeloma cell viability and proteasome inhibitor sensitivity. *Mol Cancer Res* **18**, 1453 (2020).
39. Yen, H. C. S., Xu, Q., Chou, D. M., Zhao, Z. & Elledge, S. J. Global protein stability profiling in mammalian cells. *Science (1979)* **322**, 918–923 (2008).
40. Salceda, S. & Caro, J. Hypoxia-inducible Factor 1α (HIF-1α) Protein Is Rapidly Degraded by the Ubiquitin-Proteasome System under Normoxic Conditions: ITS STABILIZATION BY HYPOXIA DEPENDS ON REDOX-INDUCED CHANGES. *Journal of Biological Chemistry* **272**, 22642–22647 (1997).

41. Domingo, A. *et al.* Dystonia-specific mutations in THAP1 alter transcription of genes associated with neurodevelopment and myelin. *The American Journal of Human Genetics* **108**, 2145–2158 (2021).
42. Yellajoshyula, D. *et al.* A pathogenic DYT-THAP1 dystonia mutation causes hypomyelination and loss of YY1 binding. *Hum Mol Genet* **31**, 1096–1104 (2022).
43. Zeng, C., Huang, W., Li, Y. & Weng, H. Roles of METTL3 in cancer: Mechanisms and therapeutic targeting. *J Hematol Oncol* **13**, 1–15 (2020).
44. Bressman, S. B. *et al.* Mutations in THAP1 (DYT6) in early-onset dystonia: a genetic screening study. *Lancet Neurol* **8**, 441–446 (2009).
45. Fuchs, T. *et al.* Mutations in the THAP1 gene are responsible for DYT6 primary torsion dystonia. *Nature Genetics 2009 41:3* **41**, 286–288 (2009).
46. Houlden, H. *et al.* THAP1 mutations (DYT6) are an additional cause of early-onset dystonia. *Neurology* **74**, 846 (2010).
47. Söhn, A. S. *et al.* Prevalence of THAP1 sequence variants in German patients with primary dystonia. *Movement Disorders* **25**, 1982–1986 (2010).
48. Xiao, J. *et al.* Novel THAP1 sequence variants in primary dystonia. *Neurology* **74**, 229–238 (2010).
49. LeDoux, M. S. *et al.* Genotype–phenotype correlations in THAP1 dystonia: Molecular foundations and description of new cases. *Parkinsonism Relat Disord* **18**, 414–425 (2012).
50. Lohmann, K. *et al.* Identification and functional analysis of novel THAP1 mutations. *European Journal of Human Genetics 2012 20:2* **20**, 171–175 (2011).
51. Mazars, R. *et al.* The THAP-Zinc Finger Protein THAP1 Associates with Coactivator HCF-1 and O-GlcNAc Transferase: A LINK BETWEEN DYT6 AND DYT3 DYSTONIAS*. *J Biol Chem* **285**, 13364 (2010).
52. Dohmen, R. J., Willers, I. & Marques, A. J. Biting the hand that feeds: Rpn4-dependent feedback regulation of proteasome function. *Biochim Biophys Acta* **1773**, 1599–1604 (2007).
53. Xie, Y. & Varshavsky, A. RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: a negative feedback circuit. *Proc Natl Acad Sci U S A* **98**, 3056–3061 (2001).
54. Radhakrishnan, S. K. *et al.* Transcription factor Nrf1 mediates the proteasome recovery pathway after proteasome inhibition in mammalian cells. *Mol Cell* **38**, 17–28 (2010).
55. Steffen, J., Seeger, M., Koch, A. & Krüger, E. Proteasomal degradation is transcriptionally controlled by TCF11 via an ERAD-dependent feedback loop. *Mol Cell* **40**, 147–158 (2010).
56. Vilchez, D. *et al.* FOXO4 is necessary for neural differentiation of human embryonic stem cells. *Aging Cell* **12**, 518 (2013).
57. Vangala, J. R., Dudem, S., Jain, N. & Kalivendi, S. V. Regulation of PSMB5 Protein and β Subunits of Mammalian Proteasome by Constitutively Activated Signal Transducer and Activator of Transcription 3 (STAT3): POTENTIAL ROLE IN BORTEZOMIB-MEDIATED ANTICANCER THERAPY *. *Journal of Biological Chemistry* **289**, 12612–12622 (2014).
58. Xu, H. *et al.* The CCAAT box-binding transcription factor NF-Y regulates basal expression of human proteasome genes. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 818–825 (2012).
59. Chapiro, J. *et al.* Destructive cleavage of antigenic peptides either by the immunoproteasome or by the standard proteasome results in differential antigen presentation. *J Immunol* **176**, 1053–1061 (2006).

60. Kincaid, E. Z. *et al.* Mice completely lacking immunoproteasomes show major changes in antigen presentation. *Nature Immunology 2011 13:2* **13**, 129–135 (2011).
61. Murata, S., Takahama, Y., Kasahara, M. & Tanaka, K. The immunoproteasome and thymoproteasome: functions, evolution and human disease. *Nature Immunology 2018 19:9* **19**, 923–931 (2018).
62. Cheng, F. *et al.* DYT6 mutated THAP1 is a cell type dependent regulator of the SP1 family. *Brain* **145**, 3968–3984 (2022).
63. Wilson, D. M. *et al.* Hallmarks of neurodegenerative diseases. *Cell* **186**, 693–714 (2023).
64. Tseng, C. S., Chao, Y. W., Liu, Y. H., Huang, Y. S. & Chao, H. W. Dysregulated proteostasis network in neuronal diseases. *Front Cell Dev Biol* **11**, (2023).
65. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nature Communications 2018 9:1* **9**, 1–15 (2018).
66. Clouaire, T. *et al.* The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. *Proc Natl Acad Sci U S A* **102**, 6907–6912 (2005).
67. Timms, R. T. *et al.* A glycine-specific N-degron pathway mediates the quality control of protein N-myristoylation. *Science* **365**, (2019).
68. Koren, I. *et al.* The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons. *Cell* **173**, 1622-1635.e14 (2018).
69. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res* **47**, W191–W198 (2019).

MATERIALS & METHODS

Cell culture. HEK-293T, HeLa and A549 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Merck #D6429); Jurkat and THP-1 cells were grown in Roswell Park Memorial Institute Medium (RPMI, Merck #R8758). Both were supplemented with 10% fetal bovine serum (ThermoFisher Scientific, #A5256701) plus penicillin and streptomycin (ThermoFisher Scientific, #15140122) and incubated at 37°C plus 5% CO_2. All cells were routinely checked for mycoplasma contamination.

Antibodies. Primary antibodies used in this study were: rabbit anti-THAP1 (Proteintech, #12584-1-AP), mouse anti-V5 tag (Abcam, #AB27671), rabbit anti-PSMB5 (Enzo Life Sciences, #BML-PW8895), mouse anti-PSMB6 (Enzo Life Sciences, #BML-PW8140), mouse anti-PSMB7 (Enzo Life Sciences, #BML-PW8145), mouse anti-HIF-1α (BD, #610959), mouse anti-Vinculin (Sigma, #V9131) and mouse anti-β-actin (Sigma, #A2228). HRP-conjugated donkey anti-mouse IgG and donkey anti-rabbit IgG secondary antibodies were obtained from Jackson ImmunoResearch.

Plasmids. Proteasome 20S core particle β-subunits were exogenously expressed from the pHRSIN-P_{SFFV}-GFP-WPRE-P_{PGK}-$Blast^R$/$Hygro^R$ lentiviral vectors (a gift from Paul Lehner), with constructs cloned in place of GFP via the Gibson assembly method using the NEBuilder HiFi Cloning Kit (NEB, #E5520S). GPS lentiviral vectors encoding the N-terminus of PTGS1 and the C-terminus of TNNC2 fused to GFP were gifts from Stephen Elledge. CRISPR sgRNA sequences were selected from the Brunello genome-wide library[65] and synthesized as top and bottom strand oligonucleotides (IDT). Oligos were phosphorylated (T4 PNK; NEB #M0201), annealed by heating to 95°C followed by slow cooling to room temperature, and then inserted (T4 ligase; NEB #M0202) into lentiCRISPRv2 (Addgene

731 #52961). shRNAs were cloned in an analogous manner into the pHR-SIREN-P_{U6}-shRNA-
732 WPRE-P_{PGK}-Puro lentiviral vector (a gift from Paul Lehner) using the BamHI and EcoRI sites.
733 Top strand oligonucleotide sequences used were:
734 sg1-Control (targets FOXP1 intron): caccgTGGGAACAGGATGAGGAAGG
735 sg2-Control (targets ATP1A1 intron): caccGATGGGCAAGAAGGAAGCAG
736 sg1-THAP1: caccgCTGCAAGAACCGCTACGACA
737 sg2-THAP1: caccGAAAACTGAGAGATTAACAG
738 sg3-THAP1: caccgCTGTGACCACAACTATACTG
739 shControl:gattcGTTATAGGCTCGCAAAAGGTTCAAGAGACCTTTTGCGAGCC
740 TATAACTTTTTTg
741 shPSMB5:gattcCAATGTCGAATCTATGAGCTTCTCGAGAAGCTCATAGATTC
742 GACATTGTTTTTTg
743
744 **Lentivirus production.** Lentiviral stocks were generated through the transfection of
745 HEK-293T cells with the specific lentiviral vector plus a mix of packaging plasmids encoding
746 Gag-Pol, Rev, Tat and VSV-G. HEK-293T cells seeded at 70-90% confluence were transfected
747 using PolyJet *In Vitro DNA* Transfection Reagent (SignaGen Laboratories, #SL100688)
748 according to the manufacturer protocol. The media was replaced 24 hours post-transfection and
749 the viral supernatant was collected at 48 hours post-transfection, centrifuged at 800 x *g* for 5
750 minutes to remove cellular debris, and either applied immediately to target cells or stored at -
751 80°C in single-use aliquots.
752
753 **CRISPR/Cas9-mediated gene knock-in.** A four-fragment Gibson assembly reaction
754 was used to generate the homology donor vector. 5' and 3' homology arms (~1 kb) were
755 amplified from genomic DNA, and were assembled together with a fragment encoding

mClover3 followed by a 2A peptide and a pUC plasmid digested with PciI (NEB, #R0655) and SbfI (NEB, #R3642). The resulting plasmid was transfected into HEK-239T cells along with a PX459 (Addgene #48139, kindly deposited by Feng Zhang) plasmid encoding Cas9 and an sgRNA (CTTTCTGCCCACACTAGACA) targeting the start of the PSMB5 coding sequence. Transfected cells were selected with puromycin for 48 h commencing 24 h post-transfection. Two weeks later, cells that remained GFP^+ were single cell cloned by FACS.

Flow cytometry and FACS. Analysis of cells by flow cytometry was performed using either an LSR-II or Fortessa instrument (BD Biosciences), collecting a minimum of 10,000 cells per sample. All flow cytometry data were collected through FACSDiva software and subsequently analyzed using FlowJo. Cell sorting was carried out using an Influx instrument (BD Biosciences).

Immunoblotting. Cells were lysed in 1% SDS plus 1:200 Benzonase (Merck, #E1014) for 20 minutes at room temperature. Following the addition of Laemmli buffer (Bio-Rad, #161-0747), lysates were heated to 70°C for 10 minutes. Proteins were separated by SDS-PAGE using 4-12% Bis-Tris gels (Merck, #MP41G12) and transferred onto an activated PVDF membrane (Merck, #IPFL00010). Membranes were blocked for a minimum of 30 min in 5% Skim Milk Powder (Merck, #70166) in PBS + 0.1% Tween-20 (PBS-T) (Merck, #P1379). Membranes were incubated with primary antibodies overnight at 4°C, washed at least three times in PBS-T, and then incubated with HRP-conjugated secondary antibodies for 40 minutes at room temperature. Following a further five washes in PBS-T, reactive bands were visualized using SuperSignal West Detection Reagents (ThermoFisher Scientific, #32106, #34580 and #34076) and images collected on a ChemiDoc Imaging System (Bio-Rad). Raw images were processed using GNU Image Manipulation Platform (GIMP) version 2.10.34.

781

782 **Imaging.** HEK-293T cells were imaged on a Zeiss Primovert Inverted Phase Contrast
783 Microscope Ph1/0.3 at 10x magnification using the NexYZ 3-axis Universal Smartphone
784 Adapter (Celestron).

785

786 **qRT-PCR.** Total RNA was extracted from ~1 million cells using the RNeasy Mini Kit
787 (Qiagen, #74104) with QIAshdredder Mini Spin Columns (Qiagen, #79656) as per the
788 manufacturer's protocol, including on-column DNaseI digestion using the RNase-Free DNase
789 Set (Qiagen, #79254). Reverse transcription was performed with 1 µg of RNA using one-step
790 reaction using LunaScript RT SuperMix Kit (NEB, #E3010) as indicated by the manufacturer.
791 For subsequent analysis by qPCR, 1 µl of cDNA template, 0.5 µl of each primer (10 µM) and
792 12.5 µl Luna Universal Probe qPCR Master Mix (NEB, M3004) were mixed in a final volume
793 of 25 µl; thermocycling was performed on a QuantStudio 7 Flex Real-Time PCR system
794 (ThermoFisher Scientific). Relative expression was quantified using the ΔΔCt method relative
795 to RPS18; data are expressed as mean ± standard deviation and *P* values calculated using a
796 one-tailed unpaired t-test. All qPCR amplicons were verified using agarose gel electrophoresis.
797 Primer sequences are listed in **Table S4**.

798

799 **RNA-seq.** RNA extracted as above was sent to Azenta for strand-specific polyA$^+$
800 Illumina library preparation and sequencing. Raw sequence reads were trimmed of adaptor
801 sequence using Cutadapt (version 4.1), aligned using HISAT2 (version 2.2.1) to the human
802 genome (GRCh38 genome_tran index), and further analyzed using SeqMonk (version 1.48.1).

803

804 **Deep mutational scan.** The THAP1 coding sequence was divided into six segments
805 for mutagenesis (encompassing residues 2-37, 38-73, 74-109, 110-145, 146-181 and 182-213).

Using pKLV2-$P_{EF1\alpha}$-THAP1-P_{PGK}-Puro-2A-BFP-WPRE as the starting point, six vectors were generated in which 'stuffer' regions flanked by BbsI restriction sites replaced the sequence encoding each segment. An oligonucleotide pool encoding the mutant alleles was synthesized by Twist Bioscience: for each segment, each amino acid was systematically exchanged to all 20 possible amino acids. Each of the six mutant segments were amplified from the oligonucleotide pool by PCR (Q5, NEB #M0491L), gel purified (QIAEX II Gel Extraction Kit, Qiagen #20021), and then cloned into their respective 'stuffer' vector cut with BbsI (NEB, #R3539S) using the Gibson assembly method (NEB, #E5520S). The reaction products were electroporated into DH10β cells (ThermoFisher Scientific, #18290015) and grown on LB plates with ampicillin overnight at 30°C; the next morning, plasmid DNA was extracted (GenElute HP Plasmid Midiprep Kit, Merck #NA0200-1KT) from all of the *E. coli* and verified by Sanger sequencing (Azenta).

The six mutant pools were combined into three for screening (1-2, 3-4 and 5-6). These were packaged into lentiviral particles and, in duplicate, introduced into GFP^{dim} THAP1 KO P_{PSMB5}-GFP reporter cells at a multiplicity of infection of ~0.3 (~30% BFP^+ cells). Five days post-transduction, the BFP^+ cells were partitioned into GFP^{dim} (THAP1 inactive) and GFP^{bright} (THAP1 active) populations by FACS. Genomic DNA was extracted from the sorted cells (Gentra Puregene Cell Kit, Qiagen #158767) and the exogenous THAP1 sequences in each sample amplified by PCR (Q5, NEB #M0493L), using primers annealing to invariant regions flanking each mutagenized segment. PCR products were purified using a spin column (QIAquick PCR Purification Kit, Qiagen #28104), and then used as a template for a second PCR reaction using primers to add the Illumina adaptors and indexes. Products were purified using a spin column, quantified using a Nanodrop spectrophotometer, and mixed evenly; the final pool was purified from a 2% agarose gel (QIAEX II Gel Extraction Kit, Qiagen #20021).

All steps were performed at sufficient scale so as to maintain at least 200-fold representation of the library. Sequencing was performed on an Illumina NovaSeq 6000 instrument using 150 bp paired-end reads. Count tables quantifying the abundance of each mutant in each sorting bin were generated by trimming the raw sequence reads of constant flanking sequence using Cutadapt (version 4.1) and aligning them to a reference index using Bowtie 2 (version 2.4.5).

FIGURE LEGENDS

Figure 1 | Transcriptional regulation of PSMB5 by THAP1 explains their co-essential relationship.

(A-C) Co-essential relationships involving UPS genes predict biological relationships, as exemplified by three E3 ligase complexes: the BAG6 complex (A), $Cul2^{VHL}$ (B) and the CTLH complex (C). Network diagrams were produced using NetworkX; numbers annotating the edges indicate pairwise correlation coefficients as calculated in[9].

(D-F) *THAP1* exhibits a strong positive co-essential relationship with both *PSMB5* and *PSMB6* across DepMap data.

(G-H) *THAP1* disruption is toxic in HEK-293T cells. Cells were transduced with a lentiviral vector expressing Cas9 and the indicated sgRNAs, followed by puromycin selection to eliminate untransduced cells commencing 48 hours later. A further 48 hours later, cells were counted, plated in equal numbers, and their viability assessed by counting (G) and brightfield microscopy (H). (Scale bar = 100 μm)

(I-J) Ablation of *THAP1* decreases PSMB5 expression. HEK-293T expressing Cas9 and the indicated sgRNAs were analyzed by qRT-PCR (I) and immunoblot (J). (*P < 0.05, t-test; ns, not significant)

Figure 2 | Lethality resulting from THAP1 loss can be rescued by exogenous expression of PSMB5.

(A) Exogenous expression of PSMB5 rescues cell viability upon *THAP1* ablation. HEK-293T cells were first transduced with lentiviral vectors expressing either PSMB5 or PSMB6; then, following transduction with Cas9 and the indicated sgRNAs, cell numbers were monitored over time.

(B) Loss of THAP1 is broadly toxic across cell types. CRISPR/Cas9-mediated ablation of *THAP1* adversely affects the viability of 946/1100 cancer cell lines (blue dots, representing effect scores < -0.25) examined by DepMap.

(C-D) The toxicity associated with *THAP1* ablation is rescued by exogenous PSMB5 expression in HeLa cells (C) and A549 cells (D).

(E-H) Like PSMB5, expression of PSMB8 also protects against the toxic effects of THAP1 loss. THP-1 cells do not exhibit any substantial growth defect following *THAP1* ablation (E). High levels of PSMB8 expression are observed in the cell lines whose growth is not significantly affected by *THAP1* loss (orange dots, representing effect scores >-0.25) in DepMap data (F), and THP-1 cells strongly express PSMB8 as assessed by qRT-PCR (G). Exogenous expression of PSMB8 can rescue the viability of HEK-293T cells following *THAP1* disruption (H).

Figure 3 | A fluorescent reporter measures endogenous PSMB5 expression in live cells.

(A-C) CRISPR/Cas9-mediated knock-in of GFP into the endogenous PSMB5 locus. A schematic representation of the procedure is shown in (A). Transfection of HEK-293T cells with Cas9, an sgRNA targeting PSMB5 and a donor template resulted in ~10% GFP-positive cells (B), which were purified by FACS and single cell cloned (C).

(D-G) THAP1 ablation reduces PSMB5 expression. CRISPR/Cas9-mediated disruption of *THAP1* in a P_{PSMB5}-GFP reporter clone (expressing exogenous PSMB5 to ensure viability) reduced P_{PSMB5}-GFP expression (D), permitting the derivation of GFP^{dim} THAP1 KO clones (E). These THAP1 KO clones exhibited greatly reduced expression of PSMB5 by qRT-PCR (using primers annealing to the 3'UTR to ensure selective amplification of the endogenous transcripts) (F), and an absence of THAP1 protein by immunoblot (a non-specific band is indicated by an asterisk) (G). (***P < 0.001, t-test; ns, not significant)

Figure 4 | THAP1 binds cognate motifs within the PSMB5 promoter to regulate its expression.

(A) THAP1 binds the *PSMB5* promoter. THAP1 ChIP-seq data in K562 cells[34] reveals an intense peak representing THAP1 occupancy at the *PSMB5* transcription start site (TSS) (top). PSMD8 is the only other proteasome subunit at which concordant binding of THAP1 is observed (bottom).

(B) Consensus THAP1 binding site (THABS) motif (adapted from[66]).

(C) Schematic representation of the three consensus THABS motifs located near the *PSMB5* TSS.

(D-F) THAP1 targets cognate sites in the PSMB5 promoter to activate gene expression. **(D)** Schematic representation of a lentiviral reporter system in which ~1 kb of the PSMB5 promoter drives the expression of GFP. Removal of the three THABS motifs ("ΔTHABS") reduced GFP expression (E); this effect was mediated through THAP1, as *THAP1* ablation decreased GFP expression driven by the wild-type promoter but not the ΔTHABS promoter (F).

Figure 5 | PSMB5 insufficiency resulting from THAP1 loss impairs proteasome function.

(A) The catalytic activity of PSMB5 is required to rescue viability upon THAP1 loss. Exogenous expression of wild-type PSMB5, but not a catalytically-inactive mutant, restored the viability of HEK-293T cells following *THAP1* ablation.

(B) Loss of THAP1 impairs proteasome assembly. *THAP1* ablation decreases the abundance of mature, processed PSMB6 and PSMB7 as assessed by immunoblot, but leads to the accumulation of the uncleaved proproteins (indicated with asterisks).

(C-E) THAP1 impairs the proteasomal degradation of short-lived proteins. **(C)** Schematic representation of the lentiviral Global Protein Stability (GPS) two-color fluorescent reporter system to monitor protein stability. **(D)** Stabilization of two model GFP-degron fusion proteins upon ablation of *THAP1*, as assessed by flow cytometry; the N-terminal peptide derived from PTGS1 harbors an N-terminal degron targeted by UBR-family E3 ligases[67], while the C-terminal peptide derived from TNNC2 harbors a C-terminal degron targeted by $Cul4^{DCAF12}$ [68].

(E) Increased abundance of endogenous HIF-1α upon THAP1 disruption, as assayed by immunoblot.

Figure 6 | Defining the transcriptional targets of THAP1.

(A) Schematic representation of the RNA-seq experiment.

(B-C) Identifying the transcriptional targets of THAP1. A summary of the RNA-seq dataset is shown in (B): genes exhibiting differential expression between sgControl and sgTHAP1 cells (n = 277) are highlighted in yellow, with the subset of those genes that display THAP1 occupancy as assessed by ChIP-seq (n = 42) colored red. Expression changes amongst all differentially expressed genes are summarized in (C), with circles representing genes bound by THAP1.

(D) Consensus binding sites for YY1 (purple) and THAP1 (red) are enriched amongst the promoters of the 42 direct THAP1 target genes. Bubble size is proportional to the number of the gene promoters containing the binding site. See also Fig. S4C.

(E-I) Validation of THAP1 target genes. Five target genes directly activated by THAP1 binding are shown: ChIP-seq data indicating THAP1 occupancy is shown on the left; the RNA-seq expression data is summarized in the center, and qRT-PCR validation is shown on the right. (***P < 0.001, t-test)

Figure 7 | A deep mutational scan defines the functional landscape of THAP1 mutations.

(A) Schematic representation of the domain architecture of the THAP1 protein.

(B) Schematic representation of the deep mutational scan, designed to interrogate the ability of all possible single amino acid variants of THAP1 to activate expression of the endogenous P_{PSMB5}-GFP knock-in allele.

(C) Site-saturation mutagenesis reveals critical residues for THAP1 function. Each cell represents the performance of a single THAP1 mutant: the mean performance of all the wild-type proteins is centered at 1 (light gray), with red colors indicating mutants which abrogate activation of the P_{PSMB5}-GFP reporter and blue colors indicating mutants which may enhance the activation of the P_{PSMB5}-GFP reporter. Dark gray cells indicate mutants for which insufficient data was available for analysis.

Figure 8 | Determining the functional effects of THAP1 mutations found in Dystonia patients.

(A-B) The deep mutational scan successfully identifies THAP1 residues critical for DNA binding. **(A)** Performance of all THAP1 variants targeting zinc-chelating residues of the zinc finger domain. Bars indicate the mean of two replicate experiments. Constructs which encode the wild-type protein are indicated in bold; the mean activity exhibited across all the wild-type THAP1 constructs is set at 1 (dotted line). **(B)** Distribution of activity scores across all the THAP1 variants targeting residues previously determined[15] to be important for DNA binding by THAP1.

(C-F) Defining structure-function relationships for THAP1. Mean activity scores from the genetic screen were mapped onto the predicted structure of THAP1 (C) or the experimental structure of the THAP1 zinc finger domain[15] (D). Overall, residues predicted to lie in ordered regions of the protein (AlphaFold pLDDT > 60) were much less tolerant of mutations than residues predicted to lie in disordered regions (E). Individual validation of the screen results was performed using flow cytometry: deletion of the coiled-coil and HCFC1-binding motif abrogated THAP1 function, whereas deletion of the disordered C-terminus did not (F).

(G-H) Profiling the activity of THAP1 mutations found in Dystonia patients. **(G)** Performance of all missense variants identified in Dystonia patients, displayed as in (A). **(H)** Individual validation experiments measuring the activity of THAP1 mutants predicted to be inactive (top row) and THAP1 mutants predicted to be active (bottom row) by flow cytometry. See also Fig. S5C-E.

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Figure 1 | Co-essentiality analysis defines functional relationships amongst UPS genes.

(A) Co-essential relationships involving UPS genes reflect functional relationships. The table illustrates the strongest pairwise relationships involving at least one gene known to function as part of the UPS, based on previous analysis of DepMap data[9]. The co-essential relationship between PSMB5 and THAP1 (highlighted in green) represents the most significant unexplained relationship across this dataset.

(B-C) Negative co-essential relationships highlight known E3 ligase-substrate pairs, including p53 which is targeted by MDM2[27] (B) and cyclin D which is degraded by $Cul4^{AMBRA1}$ [28–30] (C).

Supplementary Figure 2 | Validation of HEK-293T cell lines.

(A-C) Validation of exogenous expression of proteasome subunits. Lentiviral expression of PSMB5 (A), PSMB6 (B) and PSMB8 (C) was validated by qRT-PCR.

(D) Validation of P_{PSMB5}-GFP reporter clones by PCR from genomic DNA.

(E) P_{PSMB5}-GFP reporter cells faithfully monitor PSMB5 expression. A P_{PSMB5}-GFP knock-in clone was transduced with two shRNAs targeting PSMB5 and P_{PSMB5}-GFP expression was measured by flow cytometry.

(F) Genomic characterization of THAP1 KO clone #1. PCR amplification of the sgRNA target site from genomic DNA followed by TOPO cloning and Sanger sequencing revealed three mutated THAP1 alleles.

Supplementary Figure 3 | THAP1 acts through a cognate site in the PSMB5 promoter.

(A) Schematic depiction of the gating process to assess GFP expression from the reporter vector. Following transduction with the P_{PSMB5}-GFP lentiviral reporter vector at single copy, GFP^+ cells were gated for downstream analysis.

(B-C) THAP1 activates the PSMB5 promoter through binding to THABS site 2. Deletion of THABS2, but not THABS1 or THABS3, reduced expression from P_{PSMB5}-GFP lentiviral reporter (B); furthermore, THAP1 disruption decreased expression from the ΔTHABS1 and ΔTHABS3 vectors, but did not further decrease expression from the ΔTHABS2 vector (C).

Supplementary Figure 4 | Analysis of THAP1 target genes identified by RNA-seq analysis.

(A-B) Transcriptional downregulation of PSMB5 upon loss of THAP1. Owing to the necessity to overexpress PSMB5 to maintain the viability of THAP1 knockout cells, PSMB5 was not identified as a differentially expressed gene by RNA-seq analysis; however, reads mapping to intronic sequences (which are not present in the exogenous PSMB5 expression construct) show that endogenous PSMB5 expression was markedly reduced in THAP1 knockout cells. Raw sequence reads are shown in (A) and quantified in (B).

(C) Transcription factor binding sites enriched in the promoters of THAP1 target genes. Analysis using g:Profiler[69] identified significant enrichment for the indicated motifs; the presence of the motif in the promoter of the direct THAP1 targets genes is indicated by the green cells.

Supplementary Figure 5 | A deep mutagenic scan defines the functional impact of THAP1 mutations found in Dystonia patients.

(A) Genetic complementation of THAP1 knockout cells. A THAP1 knockout P_{PSMB5}-GFP reporter clone was transduced with a lentiviral vector expressing either wild-type or mutant (C5A, which disrupts zinc coordination by the THAP-type zinc finger) THAP1, and restoration of PSMB5 reporter expression assayed by flow cytometry.

(B) Assessing concordance between replicate experiments. The THAP1 scan was performed in duplicate: scatterplots compare the activity of each individual THAP1 mutant construct (blue dots) between the two replicates.

(C) Validation of the scan results. Individual validation by flow cytometry for three additional THAP1 mutants is shown, all of which the screen suggests should be inactive.

(D-E) Validation of the expression of a panel of inactive mutants. For seven inactive mutants, we generated constructs appending a C-terminal V5 epitope tag to allow detection by immunoblot. Following single-copy transduction of HEK-293T cells (~10% BFP^+) in each case (D), protein abundance was assessed by immunoblot (E). With the exception of the C5A and P26R mutants, all were expressed at approximately the same level as the wild-type protein.

SUPPLEMENTARY TABLE LEGENDS

Supplementary Table 1 | THAP1 ChIP-seq peaks. THAP1 ChIP-seq data from the ENCODE project was obtained from GSM803408. Peaks were identified using the implementation of the MACS peak caller in SeqMonk, using a P-value cutoff of 10^{-16} and a fragment size of 150 bp.

Supplementary Table 2 | Effect of THAP1 loss on the transcriptome as assessed by RNA-seq.

Supplementary Table 3 | A deep mutational scan of THAP1.

Supplementary Table 4 | Primer sequences.

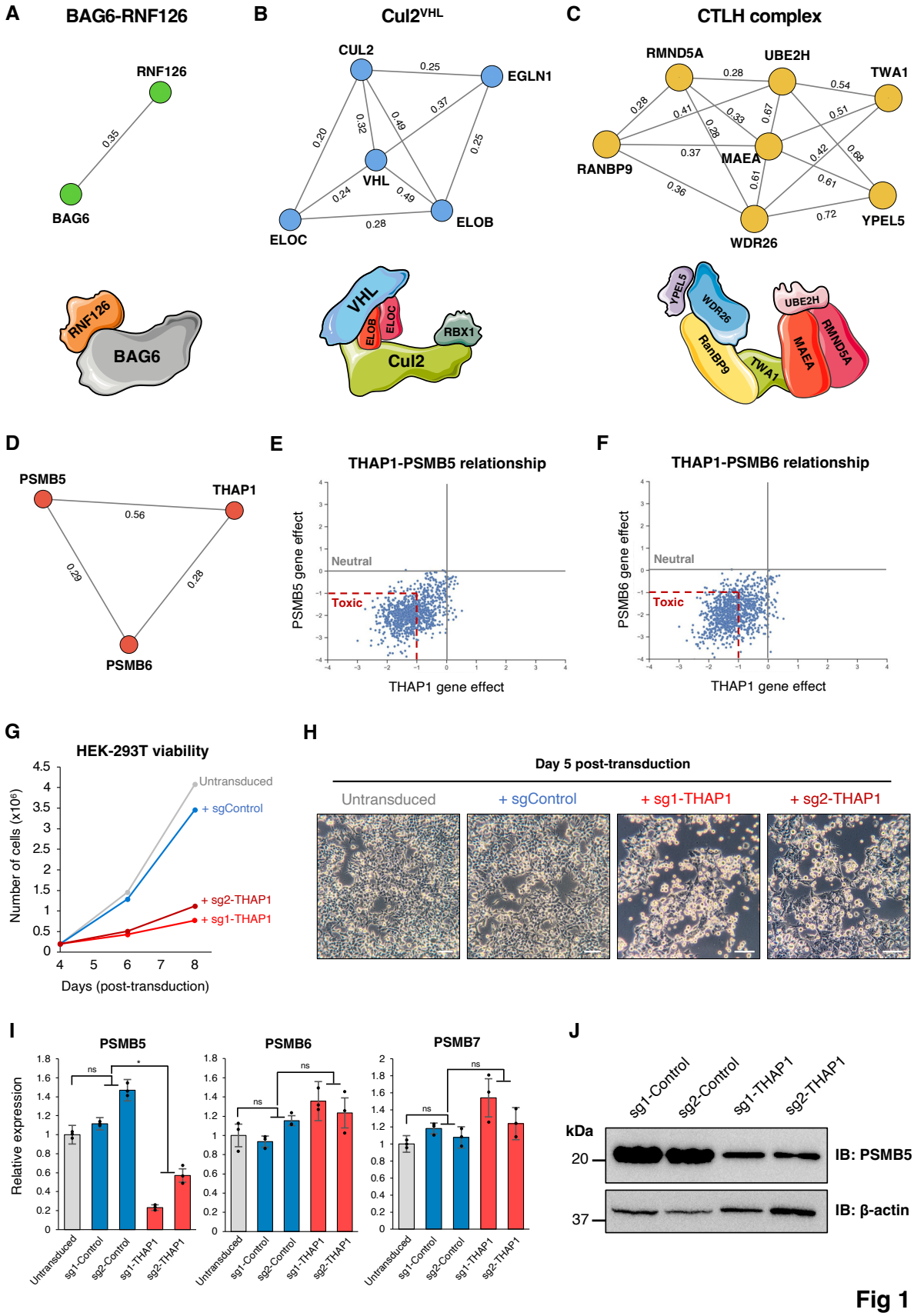

A
BAG6-RNF126
RNF126
BAG6
0.35
B
Cul2VHL
CUL2
EGLN1
VHL
ELOC
ELOB
0.25
0.37
0.49
0.32
0.20
0.25
0.24
0.49
0.28
RBX1
Cul2
C
CTLH complex
RMND5A
UBE2H
TWA1
RANBP9
MAEA
WDR26
YPEL5
0.28
0.54
0.41
0.33
0.67
0.51
0.37
0.42
0.68
0.36
0.61
0.61
0.72
D
PSMB5
THAP1
PSMB6
0.56
0.29
0.28
E
THAP1-PSMB5 relationship
PSMB5 gene effect
THAP1 gene effect
Neutral
Toxic
F
THAP1-PSMB6 relationship
PSMB6 gene effect
THAP1 gene effect
Neutral
Toxic
G
HEK-293T viability
Number of cells (x10⁶)
Days (post-transduction)
Untransduced
+ sgControl
+ sg2-THAP1
+ sg1-THAP1
H
Day 5 post-transduction
Untransduced
+ sgControl
+ sg1-THAP1
+ sg2-THAP1
I
PSMB5
PSMB6
PSMB7
Relative expression
ns
*
Untransduced
sg1-Control
sg2-Control
sg1-THAP1
sg2-THAP1
J
sg1-Control
sg2-Control
sg1-THAP1
sg2-THAP1
kDa
20
37
IB: PSMB5
IB: β-actin

Fig 1

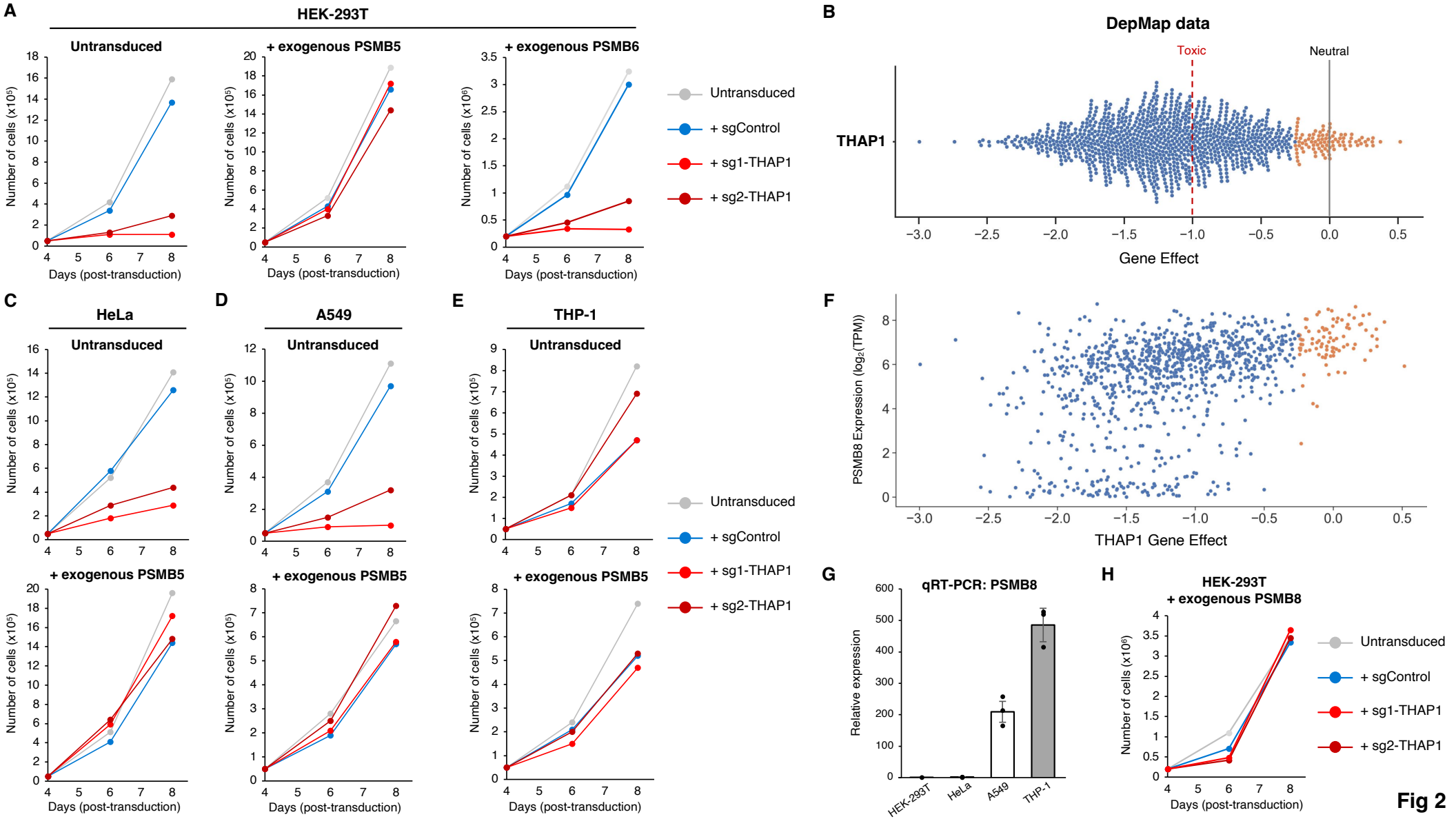

A
HEK-293T
Untransduced
+ exogenous PSMB5
+ exogenous PSMB6
Number of cells (x10⁵)
Number of cells (x10⁶)
Days (post-transduction)
Untransduced
+ sgControl
+ sg1-THAP1
+ sg2-THAP1
B
DepMap data
Toxic
Neutral
THAP1
Gene Effect
C
HeLa
D
A549
E
THP-1
F
PSMB8 Expression (log₂(TPM))
THAP1 Gene Effect
G
qRT-PCR: PSMB8
Relative expression
HEK-293T
HeLa
A549
THP-1
H
HEK-293T
+ exogenous PSMB8

Fig 2

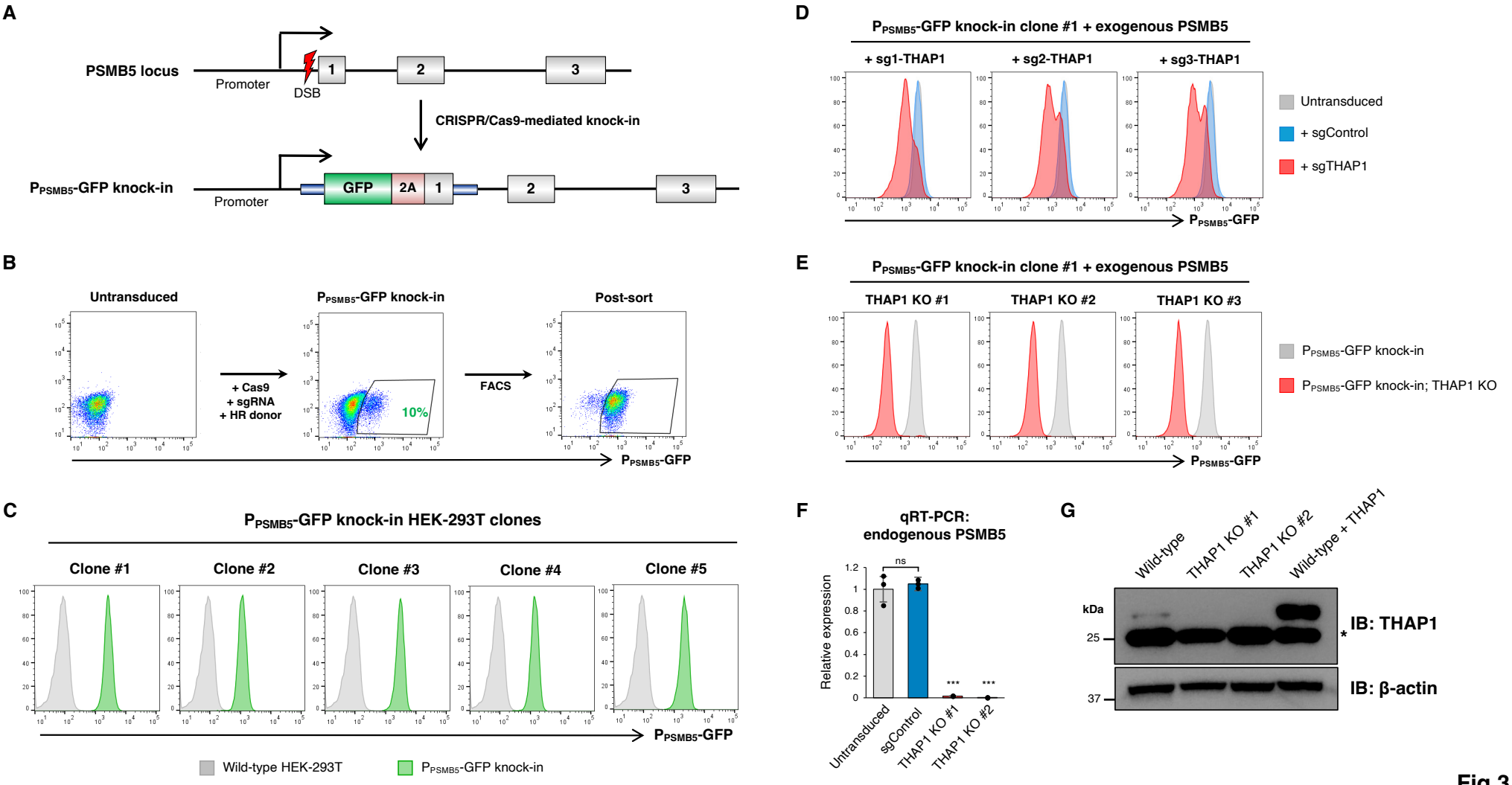
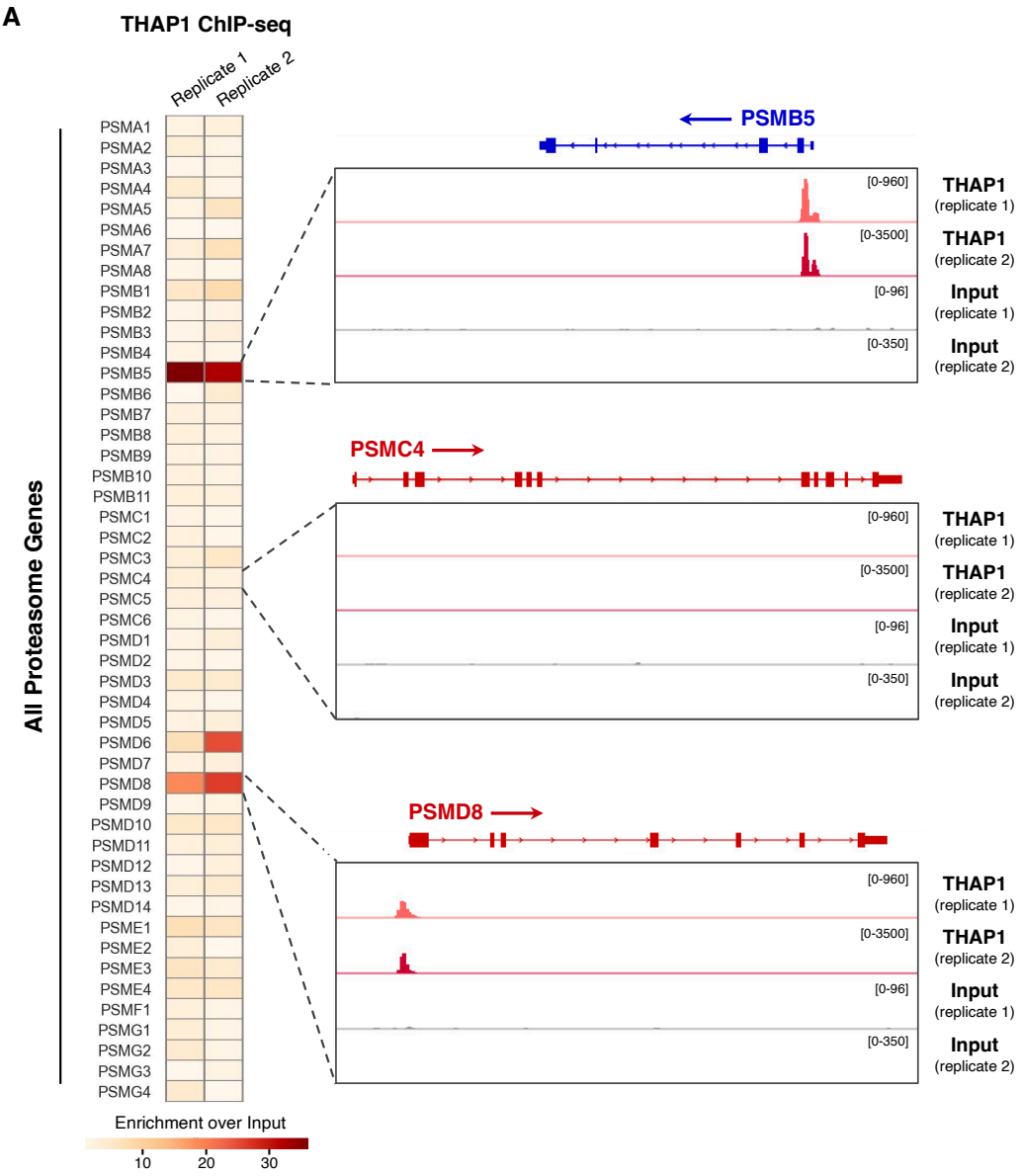
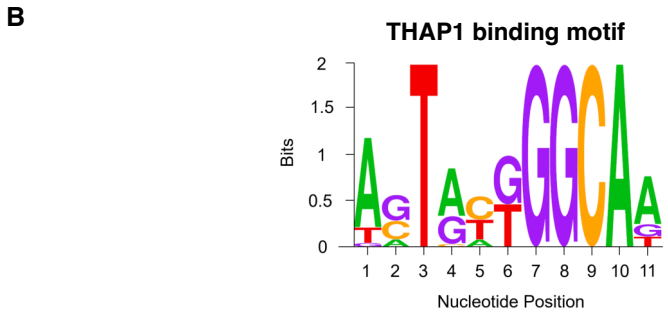
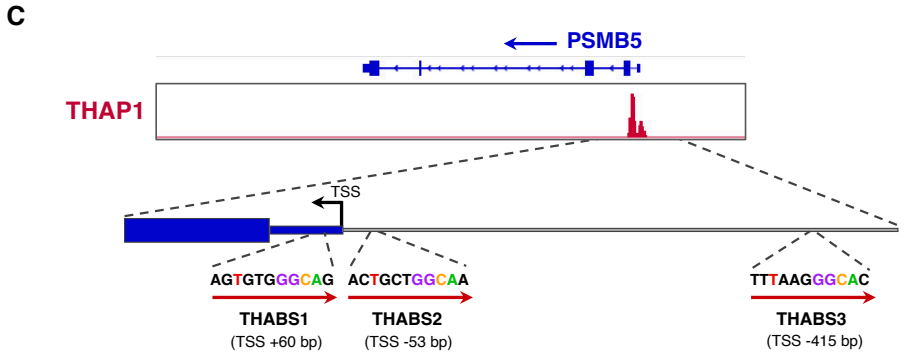
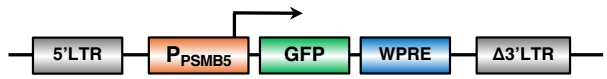

A
PSMB5 locus
Promoter
DSB
1
2
3
CRISPR/Cas9-mediated knock-in
PPSMB5-GFP knock-in
Promoter
GFP
2A
1
2
3
B
Untransduced
PPSMB5-GFP knock-in
Post-sort
+ Cas9
+ sgRNA
+ HR donor
10%
FACS
PPSMB5-GFP
C
PPSMB5-GFP knock-in HEK-293T clones
Clone #1
Clone #2
Clone #3
Clone #4
Clone #5
PPSMB5-GFP
Wild-type HEK-293T
PPSMB5-GFP knock-in
D
PPSMB5-GFP knock-in clone #1 + exogenous PSMB5
+ sg1-THAP1
+ sg2-THAP1
+ sg3-THAP1
PPSMB5-GFP
Untransduced
+ sgControl
+ sgTHAP1
E
PPSMB5-GFP knock-in clone #1 + exogenous PSMB5
THAP1 KO #1
THAP1 KO #2
THAP1 KO #3
PPSMB5-GFP
PPSMB5-GFP knock-in
PPSMB5-GFP knock-in; THAP1 KO
F
qRT-PCR:
endogenous PSMB5
Relative expression
ns
***
***
Untransduced
sgControl
THAP1 KO #1
THAP1 KO #2
G
Wild-type
THAP1 KO #1
THAP1 KO #2
Wild-type + THAP1
kDa
25
37
IB: THAP1
IB: β-actin

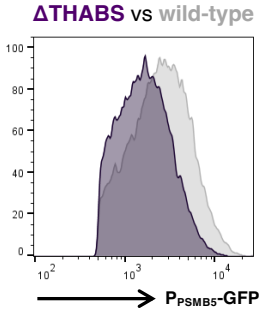
Fig 3


A
THAP1 ChIP-seq
Replicate 1
Replicate 2
All Proteasome Genes
PSMA1
PSMA2
PSMA3
PSMA4
PSMA5
PSMA6
PSMA7
PSMA8
PSMB1
PSMB2
PSMB3
PSMB4
PSMB5
PSMB6
PSMB7
PSMB8
PSMB9
PSMB10
PSMB11
PSMC1
PSMC2
PSMC3
PSMC4
PSMC5
PSMC6
PSMD1
PSMD2
PSMD3
PSMD4
PSMD5
PSMD6
PSMD7
PSMD8
PSMD9
PSMD10
PSMD11
PSMD12
PSMD13
PSMD14
PSME1
PSME2
PSME3
PSME4
PSMF1
PSMG1
PSMG2
PSMG3
PSMG4
Enrichment over Input
10
20
30
PSMB5
PSMC4
PSMD8
[0-960]
[0-3500]
[0-96]
[0-350]
THAP1 (replicate 1)
THAP1 (replicate 2)
Input (replicate 1)
Input (replicate 2)


B
THAP1 binding motif
Bits
2
1.5
1
0.5
0
1 2 3 4 5 6 7 8 9 10 11
Nucleotide Position


C
PSMB5
THAP1
TSS
AGTGTGGGCAG
ACTGCTGGCAA
TTTAAGGGCAC
THABS1
(TSS +60 bp)
THABS2
(TSS -53 bp)
THABS3
(TSS -415 bp)


D
5'LTR
P PSMB5
GFP
WPRE
Δ3'LTR


E
ΔTHABS vs wild-type
100
80
60
40
20
0
P PSMB5-GFP

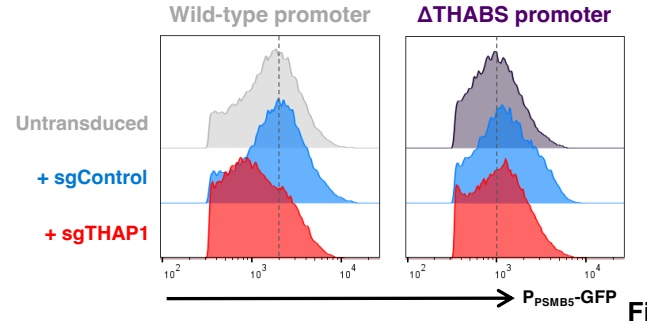

F
Wild-type promoter
ΔTHABS promoter
Untransduced
+ sgControl
+ sgTHAP1
P PSMB5-GFP

Fig 4

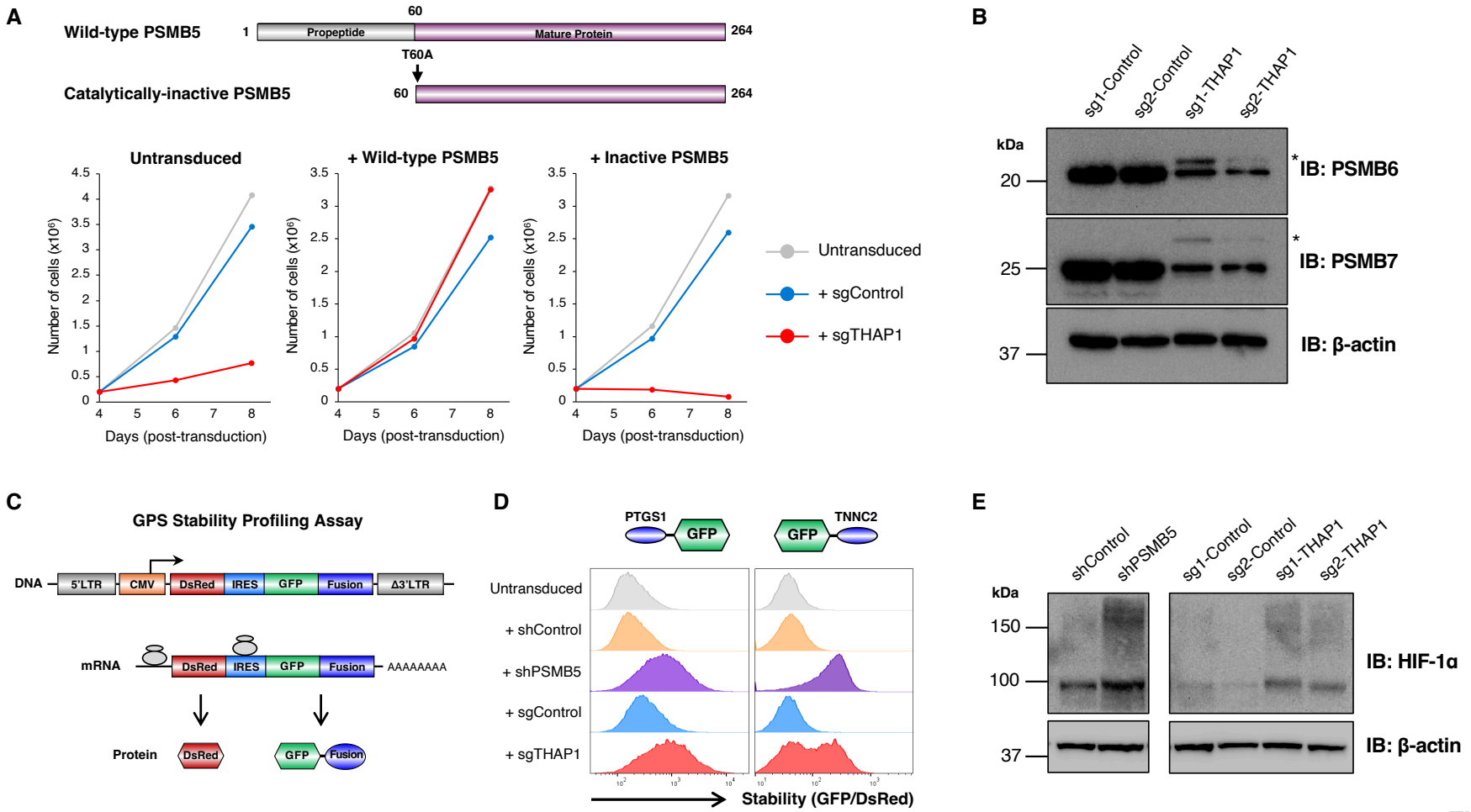

A
Wild-type PSMB5
1
60
Propeptide
Mature Protein
264
T60A
Catalytically-inactive PSMB5
60
264
Untransduced
+ Wild-type PSMB5
+ Inactive PSMB5
Number of cells (x10⁶)
Days (post-transduction)
Untransduced
+ sgControl
+ sgTHAP1
B
sg1-Control
sg2-Control
sg1-THAP1
sg2-THAP1
kDa
20
25
37
IB: PSMB6
IB: PSMB7
IB: β-actin
C
GPS Stability Profiling Assay
DNA
5'LTR
CMV
DsRed
IRES
GFP
Fusion
Δ3'LTR
mRNA
AAAAAAAA
Protein
D
PTGS1
GFP
TNNC2
GFP
Untransduced
+ shControl
+ shPSMB5
+ sgControl
+ sgTHAP1
Stability (GFP/DsRed)
E
shControl
shPSMB5
sg1-Control
sg2-Control
sg1-THAP1
sg2-THAP1
kDa
150
100
37
IB: HIF-1α
IB: β-actin

Fig 5

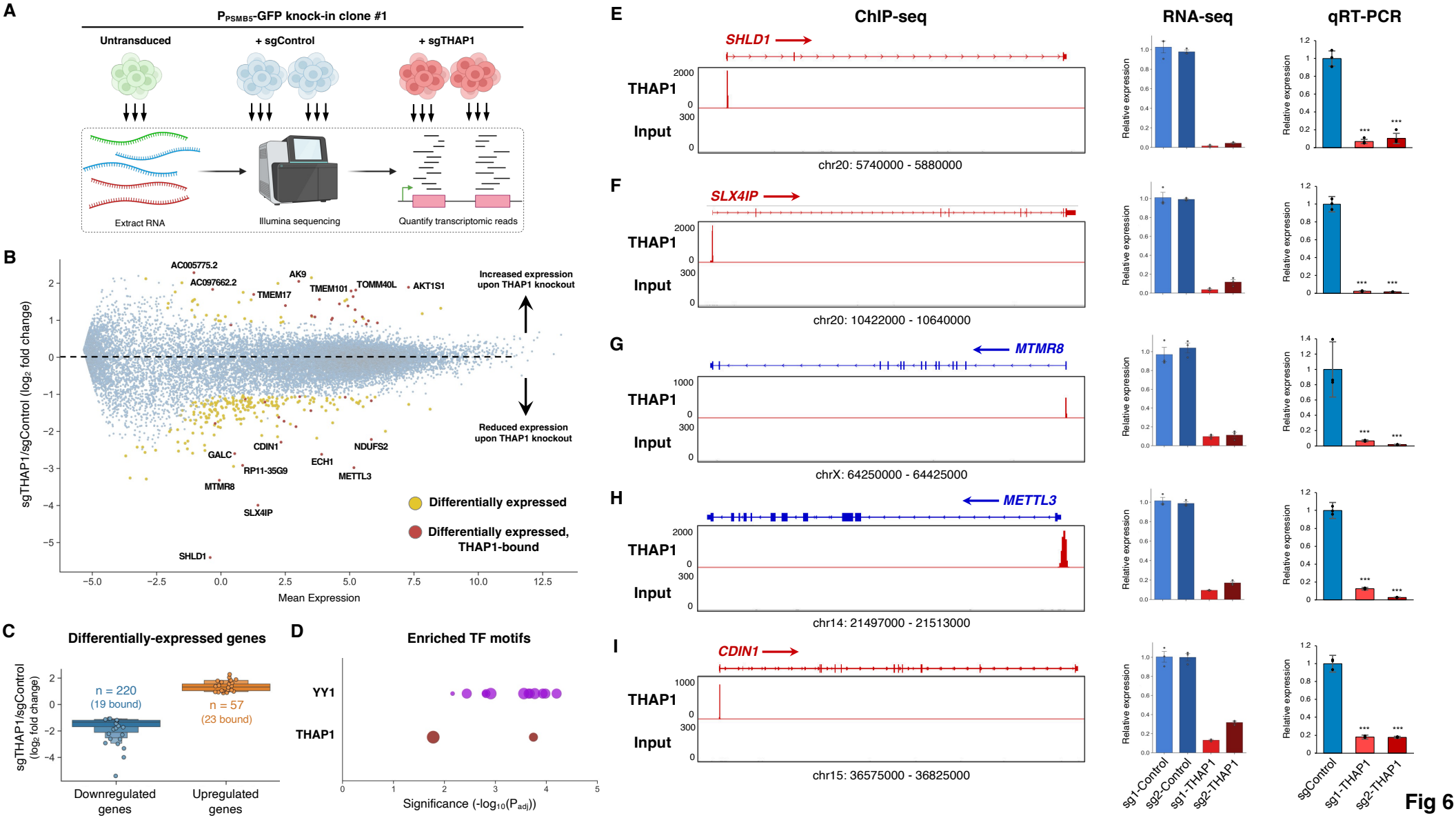

A
P_PSMB5-GFP knock-in clone #1
Untransduced
+ sgControl
+ sgTHAP1
Extract RNA
Illumina sequencing
Quantify transcriptomic reads
B
sgTHAP1/sgControl (log2 fold change)
Mean Expression
AC005775.2
AC097662.2
AK9
TMEM17
TMEM101
TOMM40L
AKT1S1
Increased expression upon THAP1 knockout
Reduced expression upon THAP1 knockout
GALC
CDIN1
RP11-35G9
MTMR8
SLX4IP
SHLD1
ECH1
METTL3
NDUFS2
Differentially expressed
Differentially expressed, THAP1-bound
C
Differentially-expressed genes
sgTHAP1/sgControl (log2 fold change)
n = 220
(19 bound)
n = 57
(23 bound)
Downregulated genes
Upregulated genes
D
Enriched TF motifs
YY1
THAP1
Significance (-log10(Padj))
E
ChIP-seq
RNA-seq
qRT-PCR
SHLD1
THAP1
Input
chr20: 5740000 - 5880000
F
SLX4IP
chr20: 10422000 - 10640000
G
MTMR8
chrX: 64250000 - 64425000
H
METTL3
chr14: 21497000 - 21513000
I
CDIN1
chr15: 36575000 - 36825000
Relative expression
sg1-Control
sg2-Control
sg1-THAP1
sg2-THAP1
sgControl
sg1-THAP1
sg2-THAP1

Fig 6

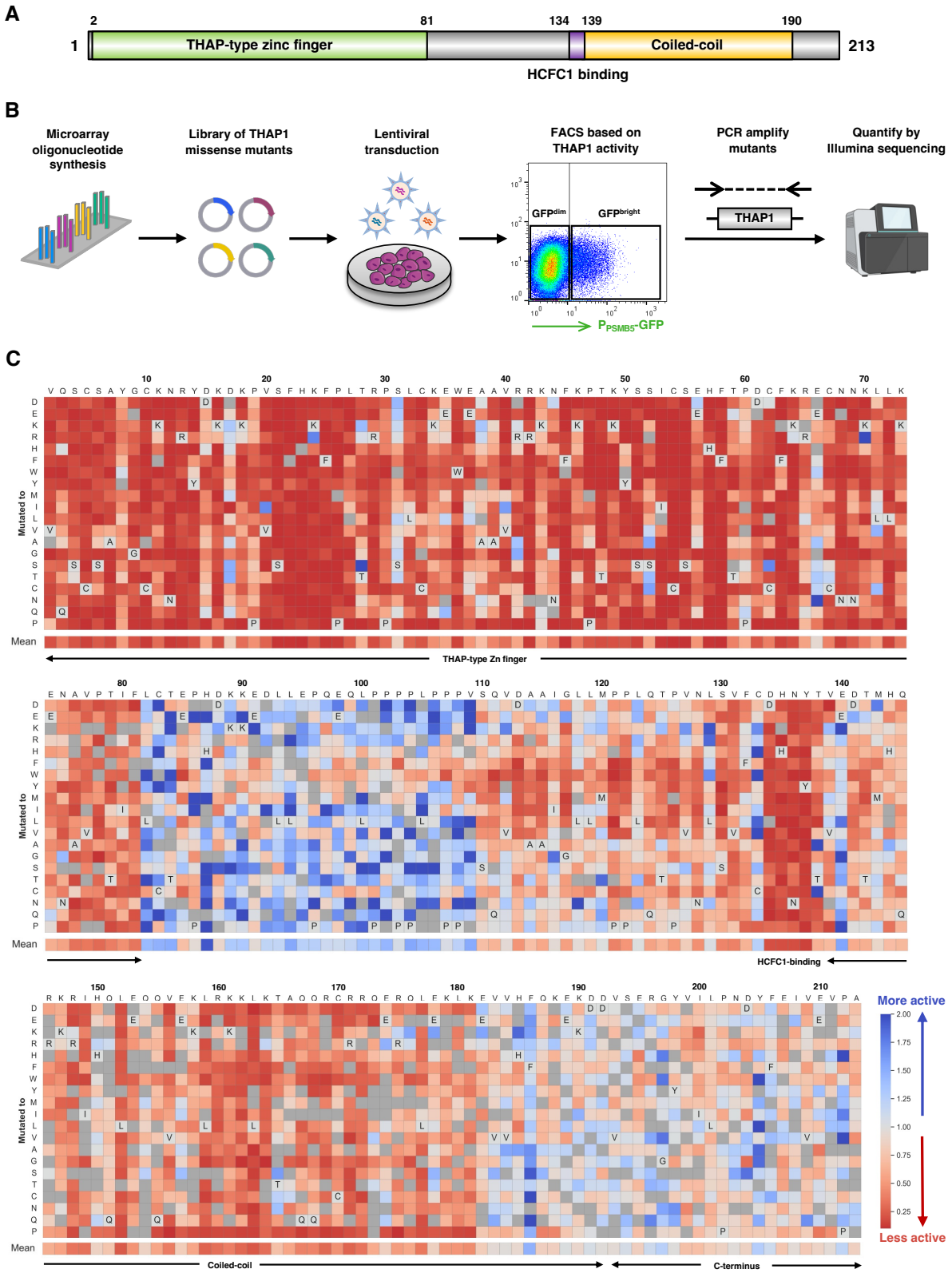

A
2
81
134
139
190
1
THAP-type zinc finger
Coiled-coil
213
HCFC1 binding
B
Microarray oligonucleotide synthesis
Library of THAP1 missense mutants
Lentiviral transduction
FACS based on THAP1 activity
PCR amplify mutants
Quantify by Illumina sequencing
GFPdim
GFPbright
PPSMB5-GFP
THAP1
C
Mutated to
Mean
THAP-type Zn finger
HCFC1-binding
Coiled-coil
C-terminus
More active
Less active
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25

Fig 7

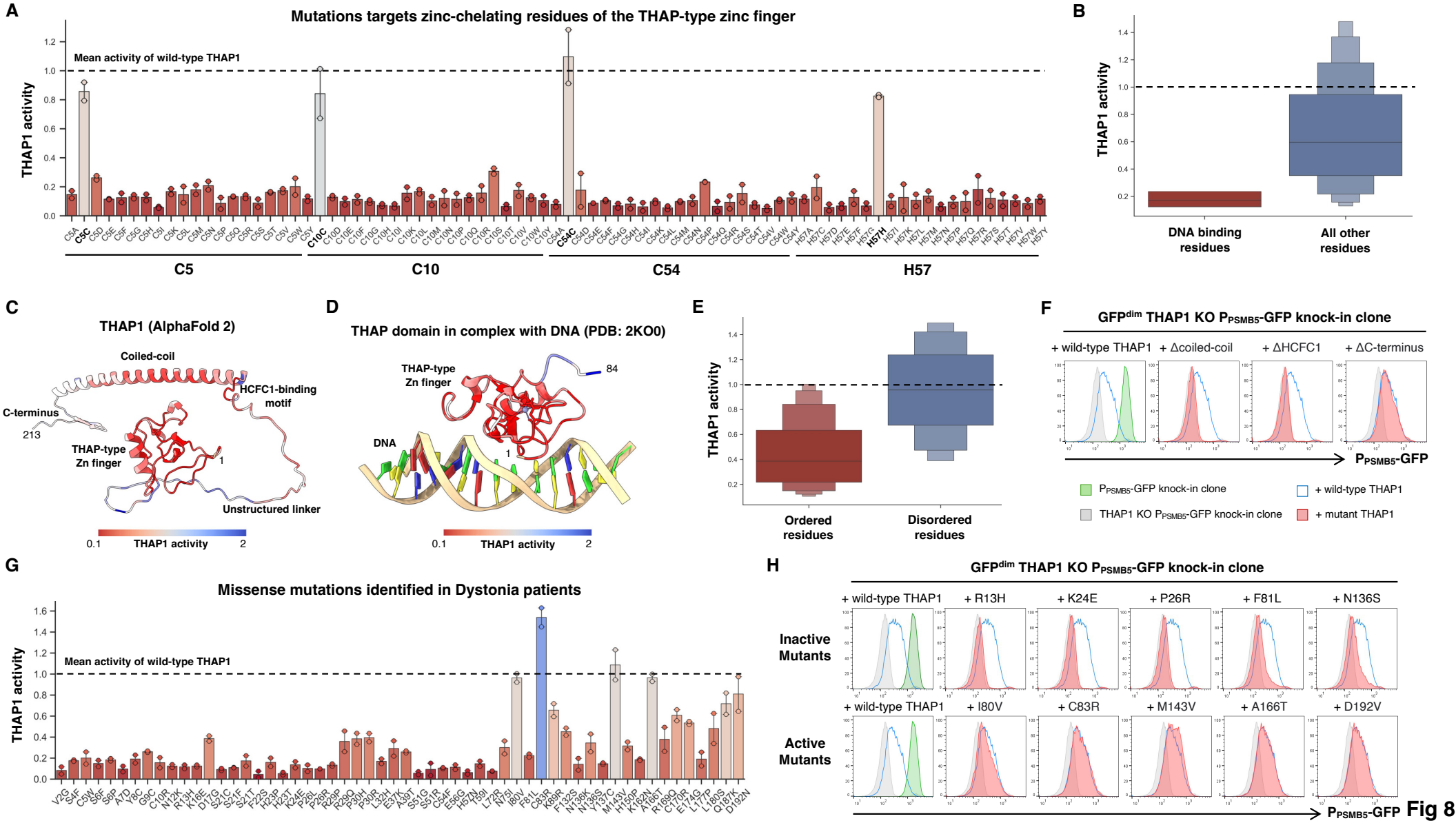
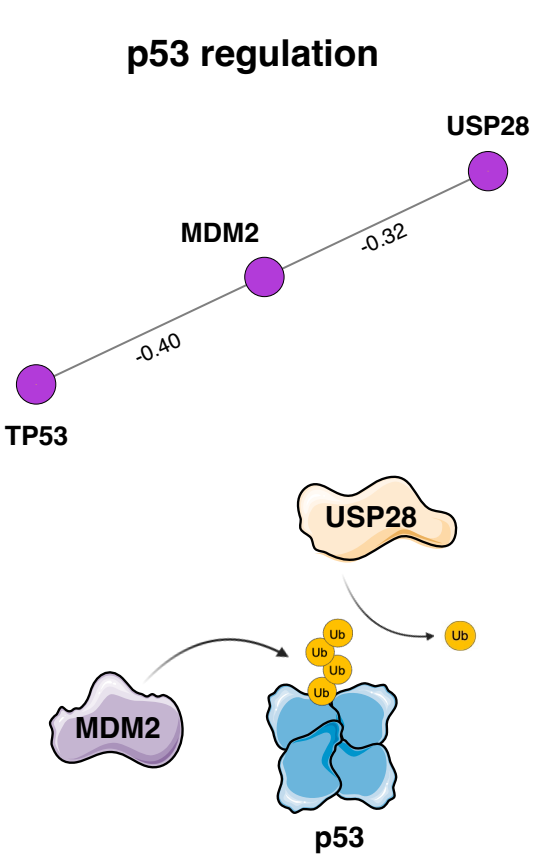

A
Mutations targets zinc-chelating residues of the THAP-type zinc finger
THAP1 activity
Mean activity of wild-type THAP1
C5
C10
C54
H57
B
THAP1 activity
DNA binding residues
All other residues
C
THAP1 (AlphaFold 2)
Coiled-coil
HCFC1-binding motif
C-terminus
213
THAP-type Zn finger
1
Unstructured linker
0.1
THAP1 activity
2
D
THAP domain in complex with DNA (PDB: 2KO0)
THAP-type Zn finger
84
DNA
1
0.1
THAP1 activity
2
E
THAP1 activity
Ordered residues
Disordered residues
F
GFPdim THAP1 KO PPSMB5-GFP knock-in clone
+ wild-type THAP1
+ Δcoiled-coil
+ ΔHCFC1
+ ΔC-terminus
PPSMB5-GFP
PPSMB5-GFP knock-in clone
+ wild-type THAP1
THAP1 KO PPSMB5-GFP knock-in clone
+ mutant THAP1
G
Missense mutations identified in Dystonia patients
THAP1 activity
Mean activity of wild-type THAP1
H
GFPdim THAP1 KO PPSMB5-GFP knock-in clone
Inactive Mutants
+ wild-type THAP1
+ R13H
+ K24E
+ P26R
+ F81L
+ N136S
Active Mutants
+ wild-type THAP1
+ I80V
+ C83R
+ M143V
+ A166T
+ D192V
PPSMB5-GFP

Fig 8

A

Rank	Gene1	Gene2	GLS_F	GLS_R	P-value	Direction	Annotation
#1	UBA5	UFM1	0.9157	0.6459	4.00E-95	+	UFMylation
#2	PSMG1	PSMG2	0.6580	0.8519	1.00E-87	+	Proteasome
#3	UFM1	UFL1	0.5439	0.9653	2.00E-79	+	UFMylation
#4	DET1	RFWD2	0.7773	0.6574	2.00E-76	+	COP1 E3 ligase complex
#5	FBXO42	CCDC6	0.5732	0.8861	8.00E-76	+	Cul1 complex
#6	UBA5	UFL1	0.6229	0.7796	4.00E-71	+	UFMylation
#7	UBR4	KCMF1	0.6678	0.7239	1.00E-70	+	SIFI E3 ligase complex
#8	WDR26	YPEL5	0.9263	0.5127	6.00E-69	+	CTLH E3 ligase complex
#9	UBE2H	YPEL5	0.8087	0.5596	1.00E-64	+	CTLH E3 ligase complex
#10	MAEA	UBE2H	0.7447	0.6044	3.00E-64	+	CTLH E3 ligase complex
#11	ARIH2	CUL5	0.5902	0.7623	4.00E-64	+	Cul5 complex
#12	DDI2	NGLY1	0.5733	0.7698	2.00E-62	+	Nrf1 pathway
#13	UFSP2	UBA5	0.7353	0.5992	2.00E-62	+	UFMylation
#14	WDR24	MIOS	0.6474	0.6778	4.00E-62	+	GATOR2 complex
#15	URM1	CTU2	0.7035	0.6156	6.00E-61	+	UFMylation
#16	FANCL	FANCG	0.7221	0.5848	5.00E-59	+	Fanconi anemia pathway
#17	UFM1	UFSP2	0.4916	0.8554	8.00E-59	+	UFMylation
#18	ARIH2	UBE2F	0.6943	0.5985	8.00E-58	+	Cul5 complex
#19	FANCL	FANCI	0.7055	0.5889	9.00E-58	+	Fanconi anemia ID2 complex
#20	ARNT	AHR	0.6797	0.6107	9.00E-58	+	Aryl hydrocarbon receptor complex
#21	UBE2H	WDR26	0.5684	0.7106	1.00E-55	+	CTLH E3 ligase complex
#22	UBE2F	CUL5	0.5186	0.7770	1.00E-55	+	Cul5 complex
#23	PEX12	PEX6	0.7792	0.5152	2.00E-55	+	Peroxisome biogenesis
#24	UFL1	UFSP2	0.6358	0.6233	2.00E-54	+	UFMylation
#25	UBA5	C1orf27	0.4456	0.8620	3.00E-52	+	UFMylation
#26	UBE2T	FANCI	0.6883	0.5468	5.00E-51	+	Fanconi anemia pathway
#27	WDR26	MAEA	0.6147	0.6058	2.00E-50	+	CTLH E3 ligase complex
#28	UFSP2	C1orf27	0.4852	0.7649	3.00E-50	+	UFMylation
#29	PEX2	PEX6	0.7560	0.4835	3.00E-49	+	Peroxisome biogenesis
#30	SKP2	CKS1B	0.4598	0.7892	8.00E-49	+	Cul1 complex
#31	VPS41	VPS39	0.5880	0.6061	7.00E-48	+	HOPS complex
#32	WDR48	USP1	0.4315	0.8176	3.00E-47	+	DUB complex
#33	WDR59	MIOS	0.8302	0.4191	2.00E-46	+	GATOR2 complex
#34	UFC1	UBA5	0.8057	0.4311	3.00E-46	+	UFMylation
#35	FANCL	FANCF	0.7247	0.4796	3.00E-46	+	Fanconi anemia pathway
#36	MAEA	YPEL5	0.7820	0.4392	1.00E-45	+	CTLH E3 ligase complex
#37	DDI2	NFE2L1	0.4513	0.7419	2.00E-44	+	Nrf1 pathway
#38	UFL1	C1orf27	0.4610	0.7124	3.00E-43	+	UFMylation
#39	PEX2	PEX5	0.5828	0.5590	6.00E-43	+	Peroxisome biogenesis
#40	PEX10	PEX6	0.7032	0.4621	7.00E-43	+	Peroxisome biogenesis
#41	BIRC6	UBA6	0.4346	0.7460	1.00E-42	+	Negative regulation of autophagy
#42	UFC1	UFM1	0.9245	0.3489	2.00E-42	+	UFMylation
#43	RNF31	MAP3K7	0.7035	0.4528	9.00E-42	+	LUBAC complex
#44	PSMB5	THAP1	0.5436	0.5845	1.00E-41	+	This paper

B



C

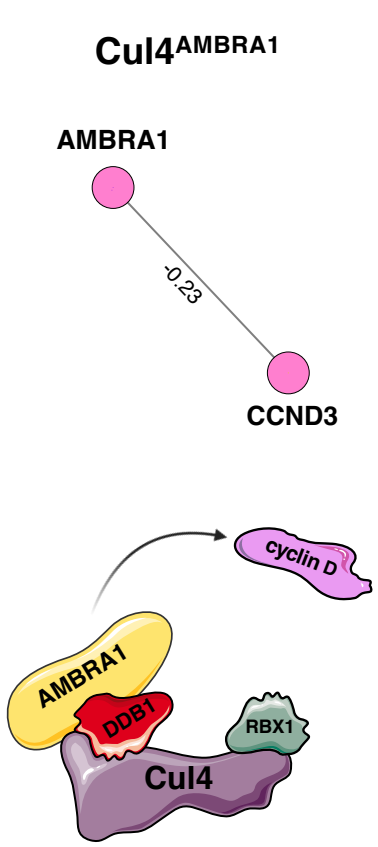


Fig S1

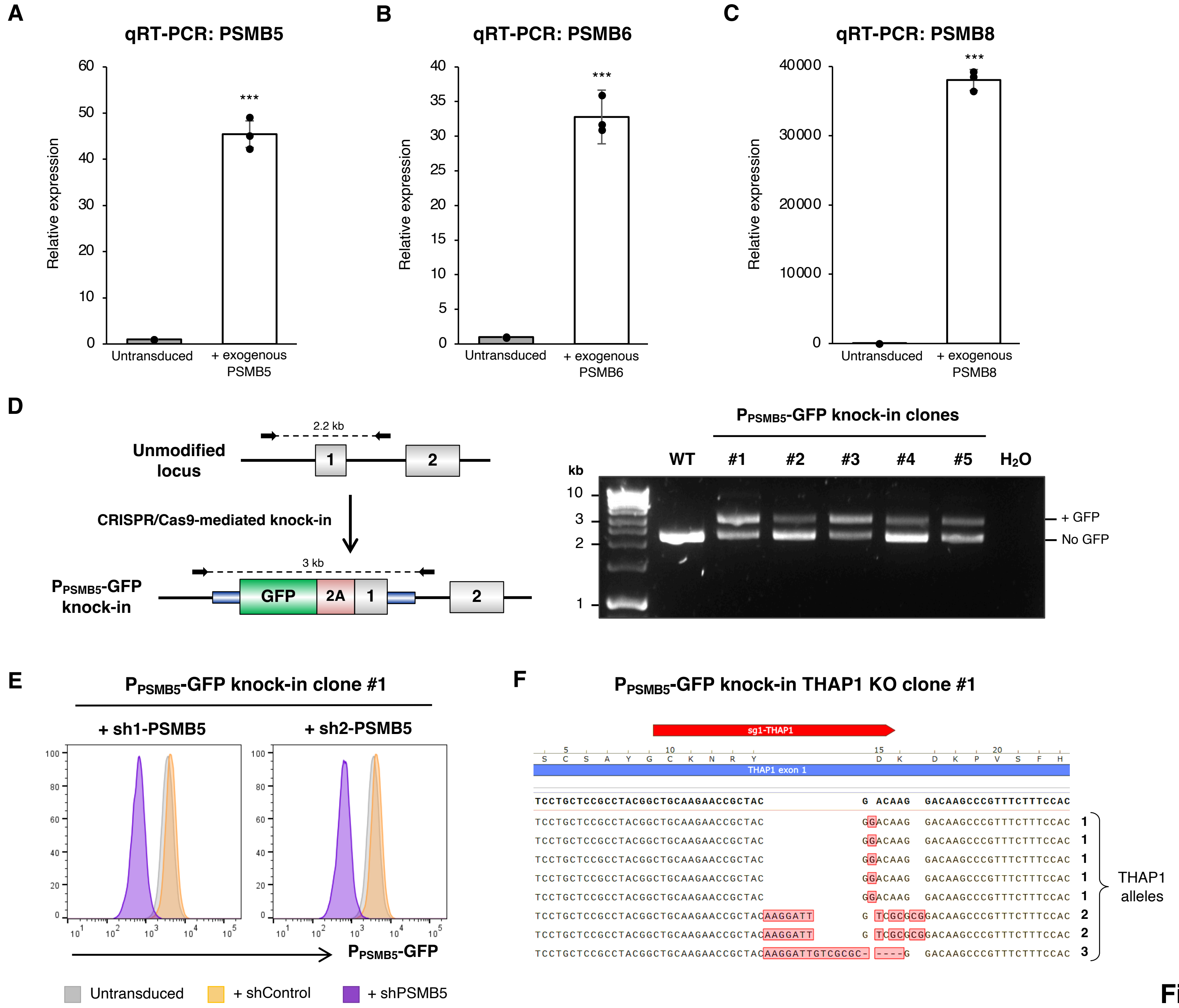
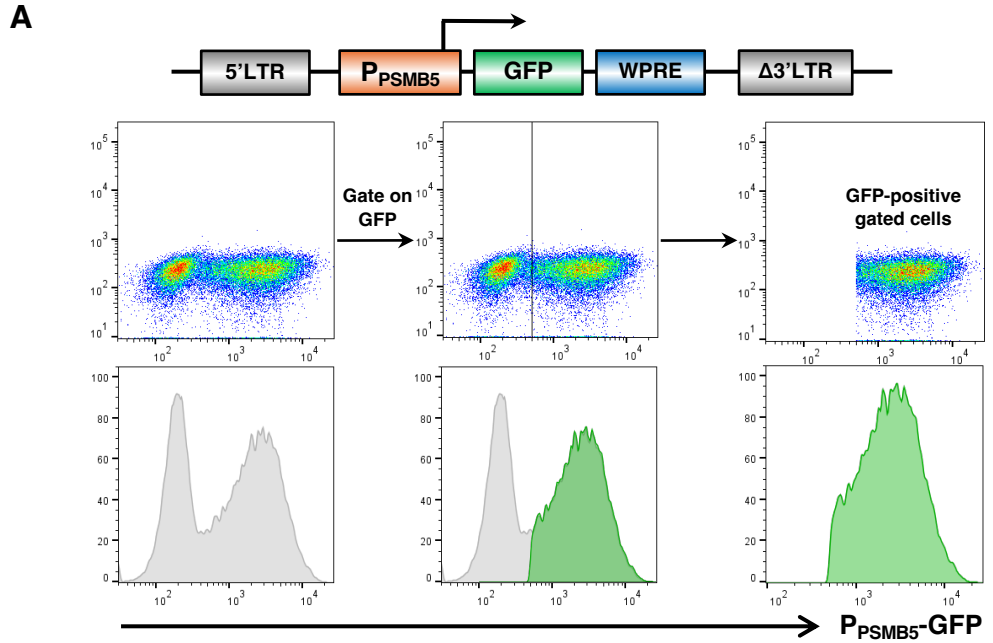
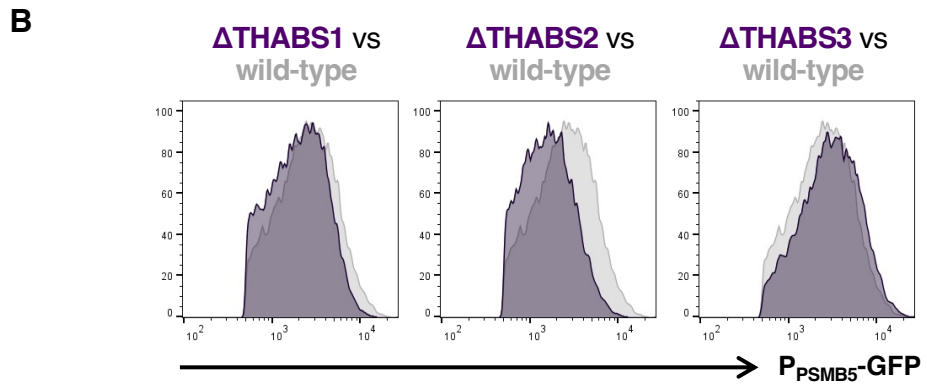

A
qRT-PCR: PSMB5
Relative expression
60
50
40
30
20
10
0
***
Untransduced
+ exogenous PSMB5
B
qRT-PCR: PSMB6
Relative expression
40
35
30
25
20
15
10
5
0
***
Untransduced
+ exogenous PSMB6
C
qRT-PCR: PSMB8
Relative expression
40000
30000
20000
10000
0
***
Untransduced
+ exogenous PSMB8
D
2.2 kb
Unmodified locus
1
2
CRISPR/Cas9-mediated knock-in
3 kb
PPSMB5-GFP knock-in
GFP
2A
1
2
PPSMB5-GFP knock-in clones
kb
WT
#1
#2
#3
#4
#5
H2O
10
3
2
1
+ GFP
No GFP
E
PPSMB5-GFP knock-in clone #1
+ sh1-PSMB5
+ sh2-PSMB5
100
80
60
40
20
0
PPSMB5-GFP
Untransduced
+ shControl
+ shPSMB5
F
PPSMB5-GFP knock-in THAP1 KO clone #1
sg1-THAP1
THAP1 exon 1
TCCTGCTCCGCCTACGGCTGCAAGAACCGCTAC G ACAAG GACAAGCCCGTTTCTTTCCAC
THAP1 alleles
1
1
1
1
1
2
2
3

Fig S2


A
5'LTR
P PSMB5
GFP
WPRE
Δ3'LTR
Gate on GFP
GFP-positive gated cells
P PSMB5-GFP


B
ΔTHABS1 vs wild-type
ΔTHABS2 vs wild-type
ΔTHABS3 vs wild-type
P PSMB5-GFP

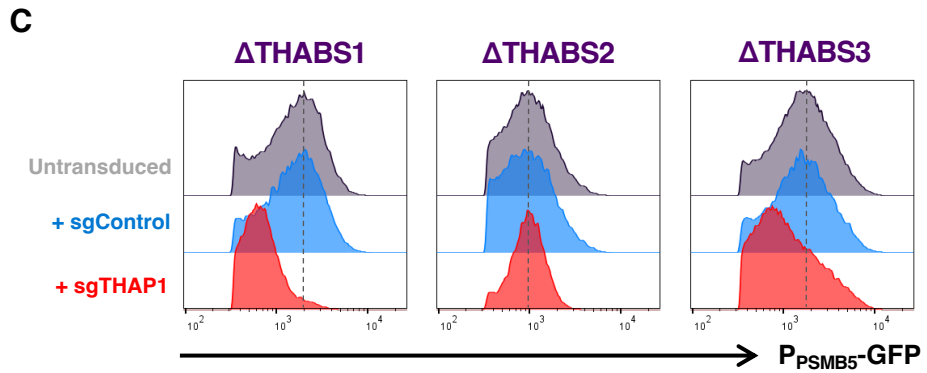

C
ΔTHABS1
ΔTHABS2
ΔTHABS3
Untransduced
+ sgControl
+ sgTHAP1
P PSMB5-GFP

Fig S3

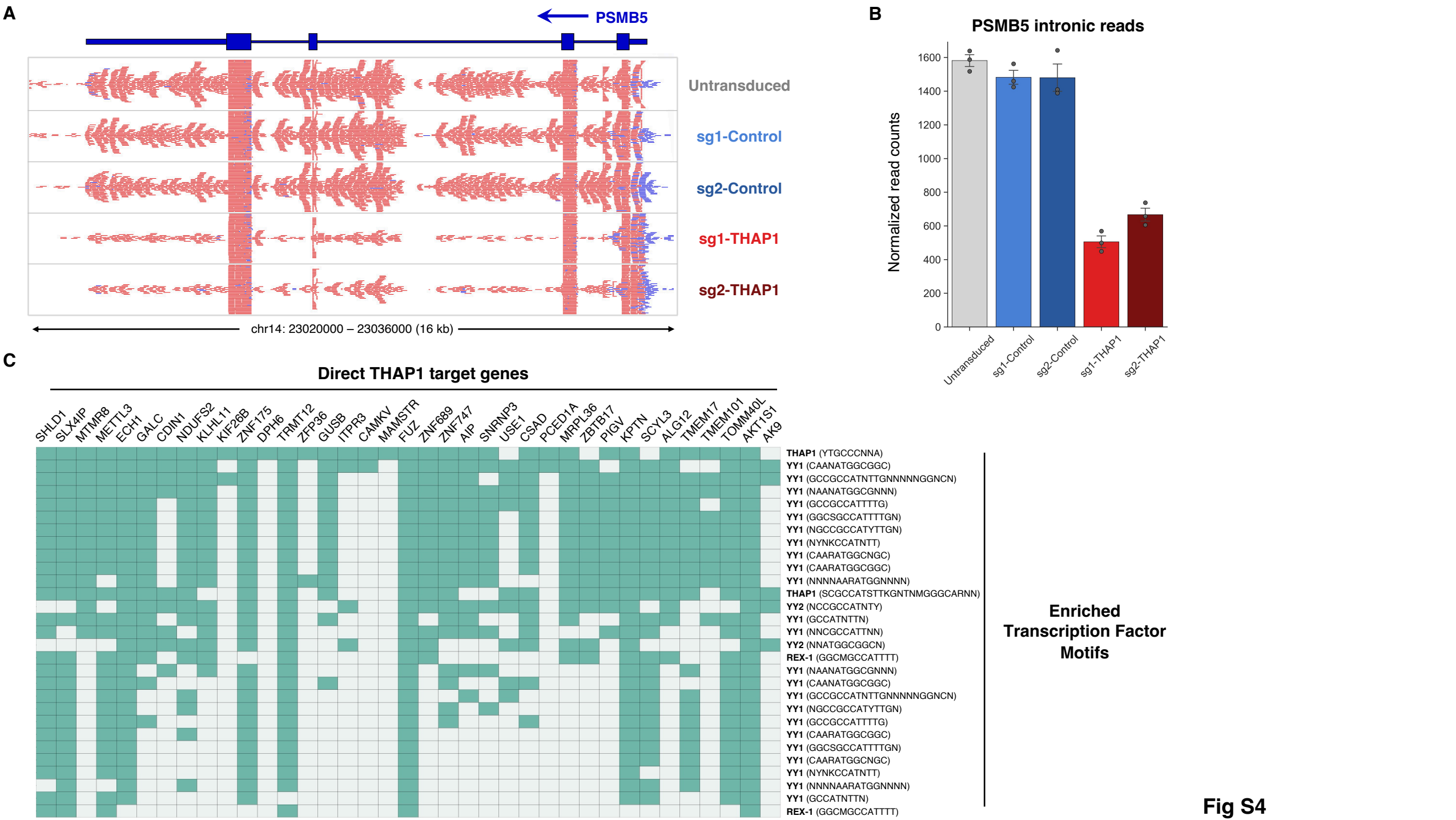

A
PSMB5
Untransduced
sg1-Control
sg2-Control
sg1-THAP1
sg2-THAP1
chr14: 23020000 – 23036000 (16 kb)
B
PSMB5 intronic reads
Normalized read counts
0
200
400
600
800
1000
1200
1400
1600
Untransduced
sg1-Control
sg2-Control
sg1-THAP1
sg2-THAP1
C
Direct THAP1 target genes
SHLD1
SLX4IP
MTMR8
METTL3
ECH1
GALC
CDIN1
NDUFS2
KLHL11
KIF26B
ZNF175
DPH6
TRMT12
ZFP36
GUSB
ITPR3
CAMKV
MAMSTR
FUZ
ZNF689
ZNF747
AIP
SNRNP3
USE1
CSAD
PCED1A
MRPL36
ZBTB17
PIGV
KPTN
SCYL3
ALG12
TMEM17
TMEM101
TOMM40L
AKT1S1
AK9
THAP1 (YTGCCCNNA)
YY1 (CAANATGGCGGC)
YY1 (GCCGCCATNTTGNNNNNGGNCN)
YY1 (NAANATGGCGNNN)
YY1 (GCCGCCATTTTG)
YY1 (GGCSGCCATTTTGN)
YY1 (NGCCGCCATYTTGN)
YY1 (NYNKCCATNTT)
YY1 (CAARATGGCNGC)
YY1 (CAARATGGCGGC)
YY1 (NNNNAARATGGNNNN)
THAP1 (SCGCCATSTTKGNTNMGGGCARNN)
YY2 (NCCGCCATNTY)
YY1 (GCCATNTTN)
YY1 (NNCGCCATTNN)
YY2 (NNATGGCGGCN)
REX-1 (GGCMGCCATTTT)
YY1 (NAANATGGCGNNN)
YY1 (CAANATGGCGGC)
YY1 (GCCGCCATNTTGNNNNNGGNCN)
YY1 (NGCCGCCATYTTGN)
YY1 (GCCGCCATTTTG)
YY1 (CAARATGGCGGC)
YY1 (GGCSGCCATTTTGN)
YY1 (CAARATGGCNGC)
YY1 (NYNKCCATNTT)
YY1 (NNNNAARATGGNNNN)
YY1 (GCCATNTTN)
REX-1 (GGCMGCCATTTT)
Enriched Transcription Factor Motifs

Fig S4

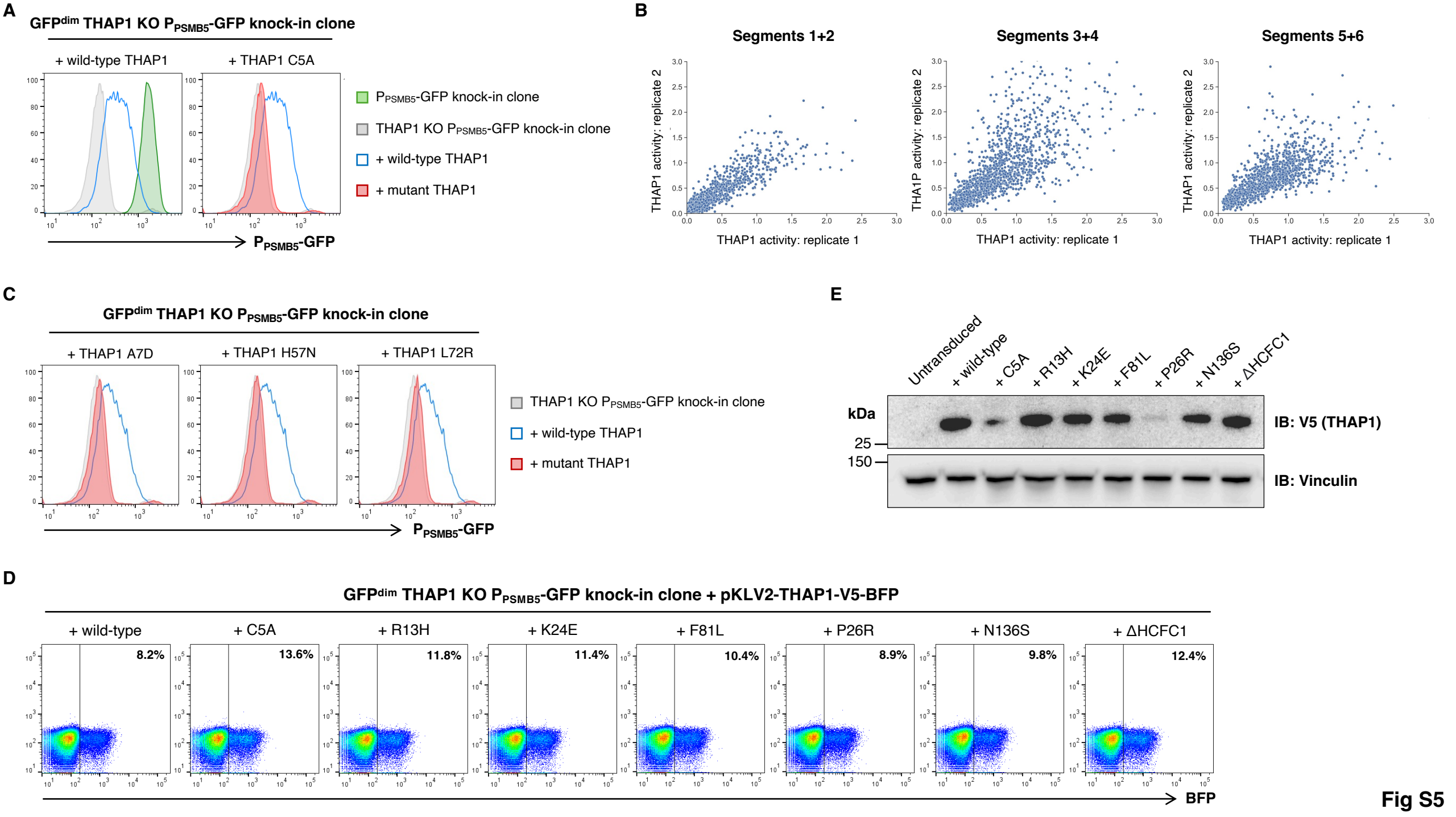

A
GFPdim THAP1 KO PPSMB5-GFP knock-in clone
+ wild-type THAP1
+ THAP1 C5A
PPSMB5-GFP knock-in clone
THAP1 KO PPSMB5-GFP knock-in clone
+ wild-type THAP1
+ mutant THAP1
PPSMB5-GFP
B
Segments 1+2
Segments 3+4
Segments 5+6
THAP1 activity: replicate 2
THA1P activity: replicate 2
THAP1 activity: replicate 1
C
GFPdim THAP1 KO PPSMB5-GFP knock-in clone
+ THAP1 A7D
+ THAP1 H57N
+ THAP1 L72R
THAP1 KO PPSMB5-GFP knock-in clone
+ wild-type THAP1
+ mutant THAP1
PPSMB5-GFP
E
Untransduced
+ wild-type
+ C5A
+ R13H
+ K24E
+ F81L
+ P26R
+ N136S
+ ΔHCFC1
kDa
25
150
IB: V5 (THAP1)
IB: Vinculin
D
GFPdim THAP1 KO PPSMB5-GFP knock-in clone + pKLV2-THAP1-V5-BFP
+ wild-type
+ C5A
+ R13H
+ K24E
+ F81L
+ P26R
+ N136S
+ ΔHCFC1
8.2%
13.6%
11.8%
11.4%
10.4%
8.9%
9.8%
12.4%
BFP

Fig S5