Supplementary File

An Integrative Proteomic Approach Identifies Novel Cellular SMYD2 Substrates

Hazem Ahmed[1], Shili Duan[2], Cheryl Arrowsmith[1,2], Dalia Barsyte-Lovejoy[1], Matthieu Schapira[1,3,*]

[1]Structural Genomics Consortium, University of Toronto, 101 College Street, MaRS Centre, South Tower, Toronto, ON M5G 1L7, Canada.
[2]Princess Margaret Cancer Centre and Department of Medical Biophysics, University of Toronto, Toronto, ON M5S 1A8, Canada.
[3]Department of Pharmacology and Toxicology, University of Toronto; 1 King's College Circle, Toronto, ON M5S 1A8, Canada.

Running Title: Novel Cellular SMYD2 Substrates

Keywords: SMYD2, non-histone substrates, lysine methylation, data integration, pan-methyl lysine antibody.

* Correspondence to: matthieu.schapira@utoronto.ca

Figure S-1: Mapping all 127 hits onto protein interaction networks. STRING database[1] was used to generate protein functional association and physical interaction networks for all 127 proteins passing our two hard filters.

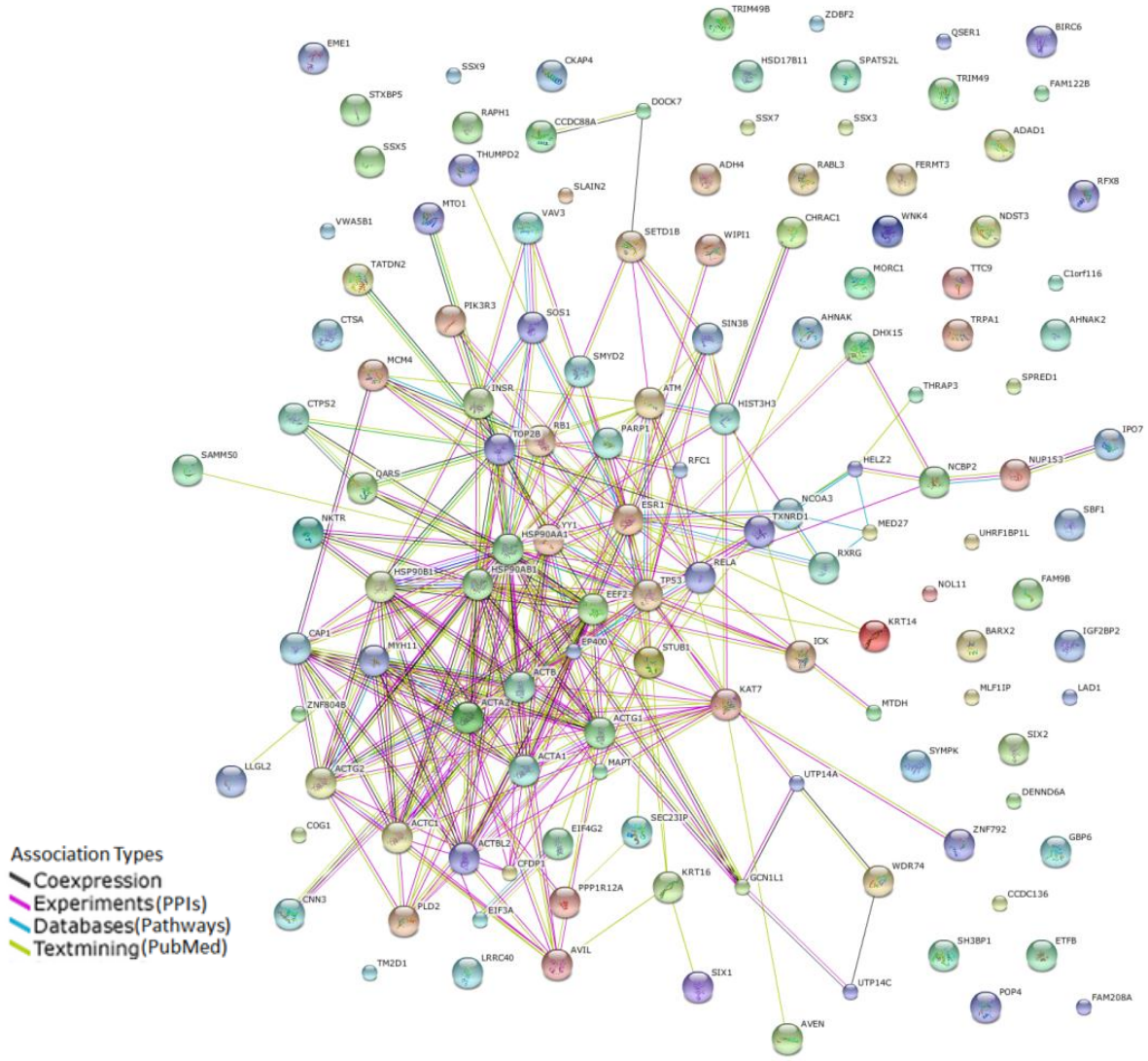

Figure S-2: Identified novel SMYD2 substrates can be nuclear or cytoplasmic. Cellular localization of confirmed SMYD2 substrate GFP fusions (green). DAPI images (blue) below indicate nuclei. The scale bar represents 15µm.

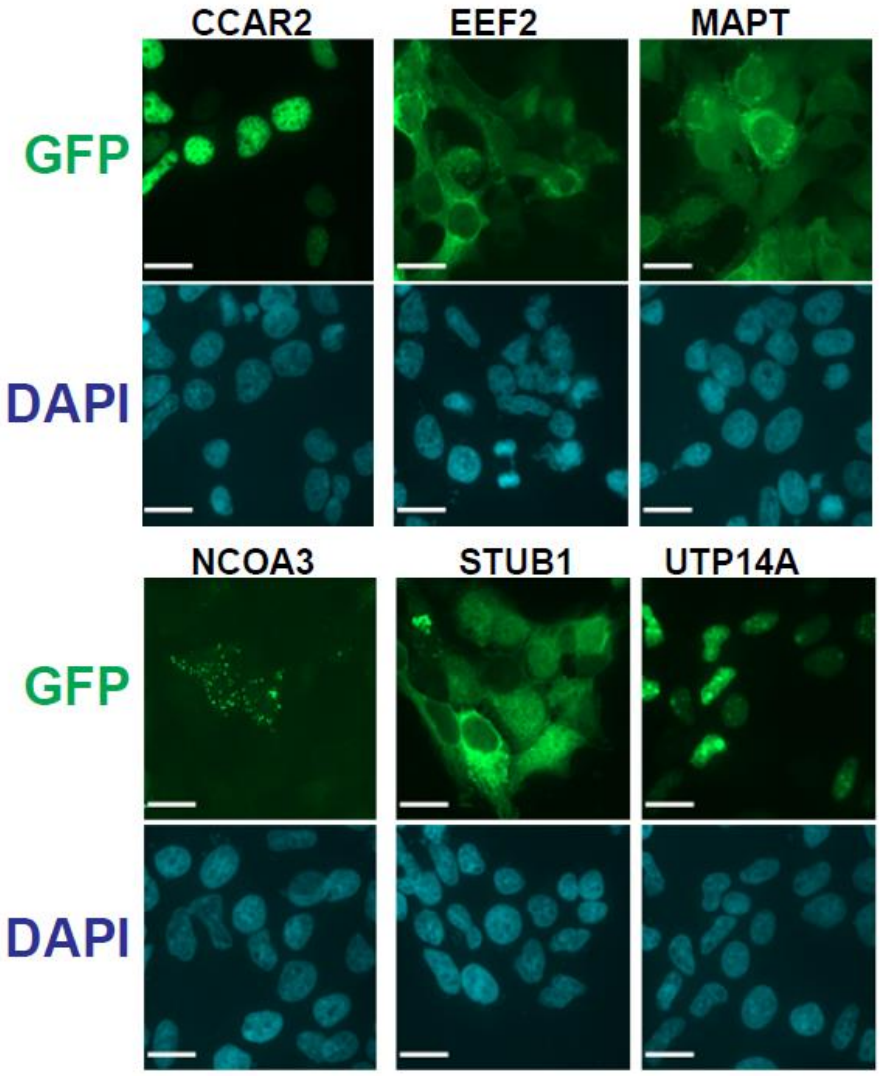

Figure S-3: A more permissive SMYD2 selectivity profile produces novel substrate candidates. A. Enriched SMYD2 specificity profile used in this work. B. Substrate candidates with major mismatches (bold) from the enriched SMYD2 specificity profile but that directly interact experimentally with SMYD2[2] are shown.

A

$[LFM]_{-1}$-$[K_{me1}]$-$[AFYMSHRK\underline{G}]_{+1}$-$[LYK\underline{QSPG}]_{+2}$

B

TP53	K370	L	K	S	K
BTF3	K2	M	K	**E**	**T**
NACA	K76	**R**	K	A	**M**
CBWD2	K204	**K**	K	A	**I**
CKAP5	K1093	**S**	K	**P**	**M**
		-1	*	1	2

Table S-1: Summary of filters implemented and associated data source. Hard filters must be met. Soft filters are selection criteria used as weighting factors.

Questions (in a descending order of importance to the selection process)	Data Source	Filter Type
Is the lysine under investigation reported to be methyalted, in the first place?	Phosphosite Plus DB, Mass Spec experiments	Hard
Is the flanking sequence around this lysine similar to SMYD2 substrate preference?	Published Specificity + Published Substrates	Hard
Does the KMT target having this lysine directly interact with SMYD2	Published interaction data + Curated PPI Databases	Soft
Does it share interaction partners with SMYD2, e.g., HSP90, or other SMYD2 published substrates?	STRING Interaction network + clustering analysis	Soft
Is the KMT target localized in the cytoplasm?	UniProt, GeneCards DB	Soft
Is the KMT target notably expressed in the heart and/or brain?	PubMed, Human Proteome Map + Human Protein Atlas	Soft
Is the KMT target functionally related to SMYD2? Similar cellular process, biological pathway?	Gene Ontology + Reactome, KEGG and STRING-DB	Soft
Is there a reliable literature support associating the KMT target with SMYD2?	Text Mining, STRING-DB	Soft

Table S-2: 160 sites selected from the methylome matching our permissive SMYD2 signature motif. A single methylation site may be present more that once. For example, TP53-K370 is present three times because it is reported to be mono- di- and tri-methylated in the methylome.

Supplementary Table 2: 160 sites in 127 proteins selected from the methylome matching our permissive SMYD2 recognition motif.

Uniprot_ID	Name	Meth Residue	Flanking Seq	Meth Degree	Domain	Source
A0JNW5	UHRF1BP1L	K1077	RTLKSQS	MONO-METHYLATION	N/A	Phosphosite
A4D1E1	ZNF804B	K113	KALKRLH	MONO-METHYLATION	N/A	Wu et al.
A6NDI0	TRIM49B	K202	EMLKKKG	DI-METHYLATION	N/A	Phosphosite
A7E2D7	EP400	K1423	AKLKASR	MONO-METHYLATION	N/A	Wu et al.
B4DXM8	PIK3R3	K328	QDLKKQA	DI-METHYLATION	N/A	Phosphosite
B7Z2S5	TXNRD1	K38	PTLKAYQ	MONO-METHYLATION	N/A	Wu et al.
E9PL73	FAM9B	K2	__MKAKK	MONO-METHYLATION	N/A	Wu et al.
F5H0L8	SEC23IP	K273	THFKKSL	MONO-METHYLATION	N/A	Wu et al.
F5H335	EIF3A	K36	DVMKSKK	MONO-METHYLATION	N/A	Wu et al.
F8W930	IGF2BP2	K493	AQFKAQG	MONO-METHYLATION	N/A	Wu et al.
F8WE41	NCBP2	K7	GLLKALR	MONO-METHYLATION	N/A	Wu et al.
H0Y901	LAD1	K18	SSLKRSR	MONO-METHYLATION	N/A	Wu et al.
H0Y9L2	SLAIN2	K50	TSLKAKQ	MONO-METHYLATION	N/A	Wu et al.
H3BS86	STUB1	K2	__MKGKE	MONO-METHYLATION	N/A	Wu et al.
H7C5D0	DENND6A	K82	HFLKSPN	MONO-METHYLATION	N/A	Wu et al.
I3L2C9	PLD2	K388	IMLKRKA	MONO-METHYLATION	N/A	Wu et al.
J3KPA2	TM2D1	K5	HILKGSP	MONO-METHYLATION	N/A	Wu et al.
J3QRV5	LLGL2	K647	KSLKKSL	MONO-METHYLATION	N/A	Wu et al.
O00515	LAD1	K367	SSLKRSS	MONO-METHYLATION	N/A	Phosphosite
O14974	PPP1R12A	K1002	EELKMLP	DI-METHYLATION	N/A	Phosphosite
O60225	SSX5	K52	VYMKRKY	MONO-METHYLATION	N/A	Phosphosite
O75366	AVIL	K319	IKMKSYP	MONO-METHYLATION	Gelsolin	Phosphosite

See excel file for full list

References

(1) Franceschini, A.; Szklarczyk, D.; Frankild, S.; Kuhn, M.; Simonovic, M.; Roth, A.; Lin, J.; Minguez, P.; Bork, P.; von Mering, C., STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **2013,** 41, (D1), D808-D815.

(2) Marcon, E.; Ni, Z.; Pu, S.; Turinsky, A. L.; Trimble, S. S.; Olsen, J. B.; Silverman-Gavrila, R.; Silverman-Gavrila, L.; Phanse, S.; Guo, H., Human-chromatin-related protein interactions identify a demethylase complex required for chromosome segregation. *Cell Rep.* **2014,** 8, (1), 297-310.