

Comprehensive Structural and Substrate Specificity Classification of the *Saccharomyces cerevisiae* Methyltransferome

Tomasz Wlodarski¹, Jan Kutner¹, Joanna Towpik¹, Lukasz Knizewski¹, Leszek Rychlewski², Andrzej Kudlicki³, Maga Rowicka^{3,4}, Andrzej Dziembowski⁵, Krzysztof Ginalski^{1*}

1 Laboratory of Bioinformatics and Systems Biology, Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland, **2** BioInfoBank Institute, Poznan, Poland, **3** Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, Texas, United States of America, **4** Institute for Translational Sciences, University of Texas Medical Branch, Galveston, Texas, United States of America, **5** Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

Abstract

Methylation is one of the most common chemical modifications of biologically active molecules and it occurs in all life forms. Its functional role is very diverse and involves many essential cellular processes, such as signal transduction, transcriptional control, biosynthesis, and metabolism. Here, we provide further insight into the enzymatic methylation in *S. cerevisiae* by conducting a comprehensive structural and functional survey of all the methyltransferases encoded in its genome. Using distant homology detection and fold recognition, we found that the *S. cerevisiae* methyltransferome comprises 86 MTases (53 well-known and 33 putative with unknown substrate specificity). Structural classification of their catalytic domains shows that these enzymes may adopt nine different folds, the most common being the Rossmann-like. We also analyzed the domain architecture of these proteins and identified several new domain contexts. Interestingly, we found that the majority of MTase genes are periodically expressed during yeast metabolic cycle. This finding, together with calculated isoelectric point, fold assignment and cellular localization, was used to develop a novel approach for predicting substrate specificity. Using this approach, we predicted the general substrates for 24 of 33 putative MTases and confirmed these predictions experimentally in both cases tested. Finally, we show that, in *S. cerevisiae*, methylation is carried out by 34 RNA MTases, 32 protein MTases, eight small molecule MTases, three lipid MTases, and nine MTases with still unknown substrate specificity.

Citation: Wlodarski T, Kutner J, Towpik J, Knizewski L, Rychlewski L, et al. (2011) Comprehensive Structural and Substrate Specificity Classification of the *Saccharomyces cerevisiae* Methyltransferome. PLoS ONE 6(8): e23168. doi:10.1371/journal.pone.0023168

Editor: Christos A. Ouzounis, The Centre for Research and Technology, Hellas, Greece

Received: March 3, 2011; **Accepted:** July 13, 2011; **Published:** August 9, 2011

Copyright: © 2011 Wlodarski et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the European Molecular Biology Organization, the Foundation for Polish Science, the Ministry of Science and Higher Education, the Innovative Economy 2007–2013 Programme, the European Social Fund and March of Dimes grant 5-FY10-136. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kinal@icm.edu.pl

Introduction

Methyltransferases (MTases) comprise a highly important class of enzymes that is present in all living organisms. MTases are involved in various cellular processes, including chromatin remodeling, DNA repair, development and signalling [1,2]. They act by catalyzing the transfer of a methyl group from mainly S-Adenosyl-L-methionine (known as AdoMet or SAM), to a nucleophilic acceptor, usually a nitrogen or oxygen atom within proteins, nucleic acids, small molecules and lipids.

Protein methylation is the second most common posttranslational modification after phosphorylation, mainly affecting the ϵ -amine group of lysine and the ω or δ guanidine nitrogen of arginine [3]. The primary targets of protein methylation in *S. cerevisiae* are histones, cytochrome C, ribosomal proteins and various translation factors. Addition of a methyl group to an amino acid residue shields the negative charge, increases the hydrophobicity and introduces steric clashes. Methylation changes both protein-protein and protein-nucleic acid interactions, influencing protein localization,

ribosome assembly, RNA processing, protein translation, protein metabolism and cell signalling [4]. Methylation of N-terminal histone tails plays an important role in transcriptional activation or repression, and is directly involved in differentiation, imprinting and X chromosome inactivation [5,6].

In *S. cerevisiae*, only RNA (not DNA) is methylated enzymatically, both base (mN) and ribose (2'-O-methylation, Nm) groups. Various RNA methylation sites in *S. cerevisiae* have been identified, including those present in rRNA (55 Nm and 10 mN), tRNA (6 Nm, 17 mN and 1 yW (wybutosine)), mRNA (2 mN), sno/snRNA, and telomerase RNA (1 mN) [7]. Importantly, rRNA ribose methylation is undertaken by MTases complexed with snoRNA (snoRNPs) [8]. Despite the fact that all of the guide snoRNPs, and most of the remaining rRNA MTases, in yeast have already been described [9], little is known about the exact role played by rRNA methylation. The functions of induced, single structural changes in rRNA are unclear; however, when they are combined, they are crucial for ribosomal processing and maturation [10]. Methylation of tRNA is essential for proper

folding, stabilization and codon recognition [11], whereas cap methylation (m^7G) is an essential step in mRNA synthesis [12]. The function of additional m^6A methylations within mRNA is not clear [13]. Telomerase RNA and sno/snRNA are hypermethylated at the cap; however, the biological significance of this modification is also unknown [14,15].

In *S. cerevisiae*, small molecules and lipids are usually methylated within metabolic pathways, such as ergosterol [16], pyrimidine deoxyribonucleotides [17], phospholipid [18] and siroheme biosynthesis [19]. Phospholipid methylation is critical for the fluidity of cell membranes, which enables receptor mobility. The modification of lipids and small molecules mainly involves C-methylation, although N- and O-methylation have also been reported.

Various experimental and theoretical studies have identified several dozen MTases in *S. cerevisiae*, the majority of which are AdoMet-dependent. A recent bioinformatic study classified AdoMet-dependent MTases into five main structural groups according to their adopted fold, the most common being the Rossmann-like [20]. Interestingly, the genes coding MTases with Rossmann-like fold make up 0.6%–1.6% of the whole genomes [21]. The remaining groups include methionine synthase fold MTases [22], tetrapyrrole methylases [23], SPOUT MTases [24] and SET domain MTases [25]. Although *S. cerevisiae* is a well-studied eukaryotic model organism, we still lack comprehensive knowledge of the MTases encoded by its genome (methyltransferome). For instance, there are a number of known methylation sites within *S. cerevisiae* proteins (e.g., eEF1A), 25S rRNA (7 sites) and tRNAs (2 sites), that have no associated MTase. Moreover, many potential yeast MTases still lack an experimental verification and substrate specificity assignment. In this study, we present a detailed picture of the *S. cerevisiae* methyltransferome, including its comprehensive structural and substrate specificity classification. We also identify previously unknown members of this class of enzymes and present a novel approach to predicting MTase substrate specificity based mainly on the Yeast Metabolic Cycle (YMC) gene expression data [26]. Finally, we show for the first time the correlation between substrate specificity of *S. cerevisiae* MTases and their time within the YMC.

Results and Discussion

Identification of the *S. cerevisiae* methyltransferome

We started with an initial set of known MTase families and structures and performed exhaustive transitive Meta-BASIC [27] searches against various protein databases, including the whole *S. cerevisiae* proteome. Meta-BASIC is a highly sensitive method for remote homology detection, capable of finding distant similarities between related proteins that are usually undetectable with standard bioinformatic tools such as PSI-BLAST [28] or RPS-BLAST [29]. This approach enabled us to identify the *S. cerevisiae* methyltransferome (Table 1), which consists of 86 proteins. Fifty three of these proteins were already described in the literature as MTases with biochemically verified MTase activity. Another 32 proteins are putative MTases that were identified in previous bioinformatic studies [21,30,31], or were automatically assigned by servers connected to Saccharomyces Genome Database (SGD) [32]. Although many of these proteins are already listed in yeast databases as putative MTases, their enzymatic activity has not been confirmed. We also identified one completely novel MTase (YIL096C) among the hypothetical proteins with unknown function.

Structural classification

Using 3D-Jury method of consensus fold recognition [33], we confidently predicted the 3D structure of the catalytic domains in

all of the identified *S. cerevisiae* MTases, for which the structure had not been solved experimentally (83 MTases in total). As a result, we structurally classified all 86 MTases and found that their catalytic domains adopt up to 9 different folds (Figure 1A), significantly more than previously reported [20]. The most common scaffold for methyl transfer is the Rossmann-like fold, found in 56 MTases. This is many more than described by Petrossian and Clarke [34], who used HMM profiles and combinatorial motif scanning for Rossmann-like MTases to identify 32 known and potentially several putative MTases in the yeast proteome.

Less commonly observed folds for methyl transfer in *S. cerevisiae* include SET domain (12 MTases), SPOUT (7 MTases), TIM beta/alpha-barrel (3 MTases), transmembrane (3 MTases), tetrapyrrole methylase (2 MTases), DNA/RNA-binding 3-helical bundle (1 MTase), SSo0622-like (1 MTase), and thymidylate synthetase (1 MTase). Consequently, the majority of *S. cerevisiae* MTases are alpha/beta proteins: Rossmann-like, SPOUT, TIM beta/alpha-barrel and tetrapyrrole methylase. However, other architectures were also found, such as alpha+beta (SSo0622-like and thymidylate synthetase), all beta (SET domain), all alpha (DNA/RNA-binding 3-helical bundle) and transmembrane proteins.

Our results suggest that methylation in *S. cerevisiae* almost exclusively uses AdoMet as a methyl group donor; only three MTases use other compounds (Table 1). The protein MTase, MGT1 (DNA/RNA-binding 3-helical bundle fold), uses DNA containing 6-O-methylguanine as a methyl group donor, while two small molecule MTases, MET6 (TIM beta/alpha-barrel fold), and CDC21 (thymidylate synthetase fold) use 5-methyltetrahydropteroyltri-L-glutamate and 5,10-methylenetetrahydrofolate, respectively. Interestingly, the different AdoMet-dependent MTase folds do not show any common features within the local architecture around the AdoMet binding site. In addition, AdoMet can be bound in significantly different conformations such as “extended” (Rossmann-like) or “bend” (SPOUT, SET domain and tetrapyrrole methylase).

At least half of the *S. cerevisiae* MTases possess additional domains that are mainly used for protein or RNA binding (Figure 2). Using distant homology detection and fold recognition, we structurally and functionally annotated the additional domains and identified several new domain architectures. The two most interesting predictions are described below.

SET5 (YPL165C). The Smyd protein family comprises genes found in Metazoa, Fungi and plants, but has not yet been described in *S. cerevisiae*. Smyd proteins are involved in the transcriptional regulation of cellular proliferation and differentiation, as well as in cancer development [35,36]. We found that SET5 contains an MYND (MYeloid, Nery and DEAF-1) domain (Figure 2) composed of two zinc fingers that are involved in the recruitment of histone deacetylase-containing complexes. This domain, together with the SET and post-SET domains that are also present in SET5, is characteristic feature of Smyd proteins. This suggests that the *S. cerevisiae* SET5 MTase, similarly to other Smyds, may be involved in histone methylation (gene silencing coupled with histone deacetylase activity).

CHO2 (YGR157W). Surprisingly, we identified the N-terminal-like domain of the coiled-coil coactivator (CoCoA) within the C-terminus of the transmembrane MTase, CHO2 (Figure 2). This domain mediates transcriptional activation by the β -catenin coactivator in the Wnt signalling pathway [37]. CHO2 is localized to the endoplasmic reticulum (ER), with the N-terminal-like CoCoA domain predicted to be positioned inside the ER lumen. In addition, CHO2 contains a highly conserved sequence motif (DWIGLYKV) that, in CoCoA, interacts with

Table 1. The *S. cerevisiae* methyltransferome.

Fold	Substrate		MTase
Rossmann-like	Protein		YBR034C (HMT1), YBR133C (HSL7), <u>YDR140W (MTQ2)</u> , YDR435C (PPM1), YDR440W (DOT1), YDR465C (RMT2), <u>YNL063W (MTQ1)</u> , <u>YBR261C (TAE1)</u> , <u>YBR271W</u> , <u>YDR316W (OMS1)</u> , <u>YIL064W (SEE1)</u> , <u>YIL110W (HPM1)</u> , <u>YJR129C</u> , <u>YKL155C (RSM22)</u> , <u>YLR285W (NNT1)</u> , <u>YNL024C</u> , <u>YOR239W (ABP140)</u>
	RNA	tRNA	YBL024W (TRM4), YDL201W (TRM8), YBR061C (TRM7), YDR120C (TRM1), YHR070W (TRM5), YJL125C (GCD14), YKR056W (TRM2), YML014W (TRM9), YOL124C (TRM11), YOL125W (TRM13), YOL141W (PPM2), YPL030W (TRM44)
		rRNA	YCL054W (SPB1), YCR047C (BUD23), YDL014W (NOP1), <u>YGL136C (MRM2)</u> , YPL266W (DIM1)
		tRNA/rRNA	<u>YDR083W (RRP8)</u> , <u>YNL022C (RCM1)</u> , <u>YNL061W (NOP2)</u> , <u>YBR141C</u> , <u>YIL096C</u> , <u>YLR063W</u>
		mRNA	<u>YBR236C (ABD1)</u> , YGL192W (IME4)
		telomerase/snoRNA	YPL157W (TGS1)
		Small molecule	YER175C (TMT1), YML110C (COQ5), YOL096C (COQ3)
		Lipid	YML008C (ERG6)
		Unknown	<u>YCL055W (KAR4)</u> , <u>YGR001C (AML1)</u> , <u>YHR209W (CRG1)</u> , <u>YMR209C</u> , <u>YMR228W (MTF1)</u> , <u>YNL092W</u> , <u>YBR225W</u> , <u>YKL162C</u> , <u>YLR137W</u>
	SPOUT	RNA	tRNA
		rRNA	YLR186W (EMG1), YOR201C (MRM1)
		tRNA/rRNA	<u>YGR283C</u> , <u>YMR310C</u> , <u>YOR021C</u>
SET domain	Protein		YBR030W (RKM3), <u>YDR198C (RKM2)</u> , <u>YDR257C (RKM4)</u> , YHR109W (CTM1), <u>YHR119W (SET1)</u> , <u>YJL168C (SET2)</u> , YPL208W (RKM1), <u>YHR207C (SET5)</u> , <u>YPL165C (SET6)</u> , <u>YHL039W (EFM1)</u> , <u>YJL105W (SET4)</u> , <u>YKR029C (SET3)</u>
SSo0622-like	RNA	tRNA	<u>YGL050W (TYW3)</u>
Transmembrane	Protein		YDR410C (STE14)
	Lipid		YGR157W (CHO2), YJR073C (OPI3)
Tetrapyrrole methylase	Protein		YLR172C (DPH5)
	Small molecule		YKR069W (MET1)
DNA/RNA-binding 3-helical bundle	Protein		<u>YDL200C^a (MGT1)</u>
TIM beta/alpha-barrel	Small molecule		YER091C ^b (MET6), YLL062C (MHT1), YPL273W (SAM4)
Thymidylate synthetase	Small molecule		YOR074C ^c (CDC21)

MTases are grouped according to the structural similarity of their catalytic domains (Fold) and substrate specificity (Substrate). Known MTases with experimentally determined substrate specificity are shown in regular font, putative MTases are italicized, and the newly detected MTase is highlighted in bold. Non-periodic MTases are underlined. With the exception of ^aDNA, containing 6-O-methylguanine, ^b5-methyltetrahydropteroyltri-L-glutamate, and ^c5,10-methylenetetrahydrofolate, AdoMet is the methyl group donor.

doi:10.1371/journal.pone.0023168.t001

p300, a cofactor that participates in transcriptional activation [37]. Although the Wnt signalling pathway is not present in yeast, detection of an N-terminal-like CoCoA domain in the CHO2 MTase suggests the existence of a previously unknown regulation mechanism that may be involved in chromatin-mediated reshaping of the ER during nuclear membrane formation [38].

Substrate specificity classification

We also classified *S. cerevisiae* MTases according to their substrate specificity by dividing them into four general groups: protein MTases, RNA MTases, lipid MTases and small molecule MTases. For the 53 MTases with known substrate specificity, these groups comprise 17, 25, three and eight proteins, respectively (Figure 1B). To predict the substrate specificity of the 33 putative *S. cerevisiae* MTases embracing 25 Rossmann-like, five SET domain and three SPOUT fold proteins, we developed a novel theoretical approach. The method exploits the characteristic gene expression patterns observed for the different classes of MTases in the YMC. The list of 51 MTases, reported as

periodic by Tu *et al.* [26], was extended based on a visual examination of temporal expression profiles at SCEPTRANS web server [39]. This procedure yielded 72 MTases (44 with known substrate specificity and 28 putative MTases), to which we are referring as “periodic MTases” in this paper (Table 1). We used the expression profiles seen during the YMC, since this system is known for “compartmentalization in time” [26]. This means that genes with similar functions tend to be expressed within a specific temporal window during the YMC. Indeed, hierarchical clustering of MTases based on the correlation between their gene expression profiles during the YMC, results in clusters enriched in MTases with certain substrate specificities (Figure 3). Particularly, cluster III comprises mainly of tRNA MTases, rRNA MTases, and a few protein MTases that methylate ribosomal proteins. The statistically significant grouping of tRNA and rRNA MTases within this cluster is supported by enrichment p-value (7×10^{-6}) calculated from hypergeometric distribution. We also found that these periodic MTases are expressed in a very short time window (35 min) during the YMC (300 min), what is not seen for any other yeast MTases (Figure 4).

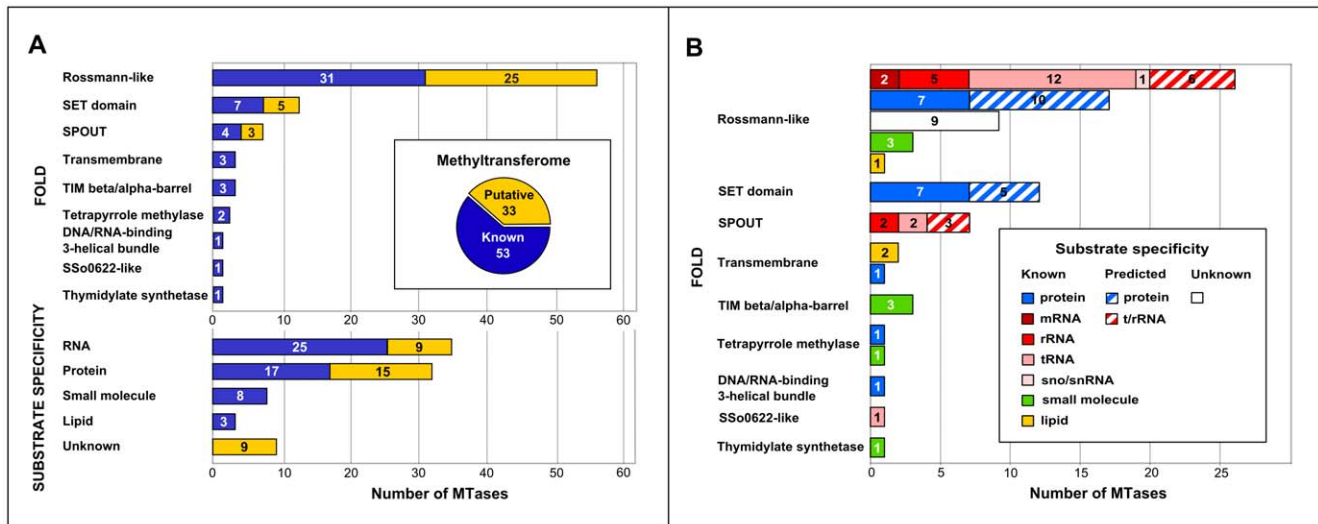


Figure 1. Comprehensive picture of the *S. cerevisiae* methyltransferome. (A) Structural (fold) and substrate specificity classifications. Eighty-six MTases (known MTases with experimentally verified activity and putative MTases identified in previous bioinformatic studies, including one newly detected here) were divided into several groups based on similarity of their structure (within catalytic domain) and substrate. (B) Detailed structural vs. substrate specificity classification.

doi:10.1371/journal.pone.0023168.g001

RNA MTases have statistically very significant tendency to have high maximum isoelectric point (max pI) (see Materials and Methods): 96% RNA MTases have max (i.e. local) pI higher than 8 (p-value 2×10^{-7} , hypergeometric). This is not unexpected, as protein domains involved in binding nucleic acids are characterized by the presence of extensive number of positively charged residues, resulting in their high pI. On the contrary, protein MTases exhibit statistically significant preference for low max pI: 65% protein MTases have max pI lower than 8 (p-value 0.01, hypergeometric). Therefore, the calculated max pI values can aid in distinguishing between RNA and protein MTases. Cellular localization is also informative, for the same reason that it is being used to filter protein-protein interaction data: it can help determine if two molecules potentially able to interact have a chance to be in proximity (the same organelle) in vivo. For example, all three lipid MTases are localized in the ER (p-value 2×10^{-4} , hypergeometric). Fold is bringing even more information about MTase substrate specificity and in some cases it is the only information needed. While Rossmann-like fold MTases can methylate all types of substrates (proteins, RNA, small molecules and lipids), SPOUT [24] and SET domain [25] MTases seem to methylate only RNA and proteins, respectively.

We combined all described above relationships between the YMC expression profiles, max pI, subcellular localization, fold assignment and the type of MTase substrate, to propose a novel approach for predicting substrate specificity. This heuristic method was implemented using decision tree (Table 2), validated on 44 known periodic MTases. We applied this approach to 28 putative periodic MTases, and predicted the substrate specificity for 19 (16 Rossmann-like, two SPOUT and one SET domain) of them, leading to the prediction of 11 novel protein MTases (YHR207C, YBR271W, YJR129C, YNL024C, YLR285W, YIL064W, YIL110W, YOR239W, YKL155C, YDR316W and YBR261C) and eight novel RNA (tRNA or rRNA) MTases (YLR063W, YBR141C, YIL096C, YNL022C, YNL061W, YDR083W, YMR310C and YGR283C). Our prediction regarding the RNA substrate specificity for YIL096C (a newly identified MTase in this study) is consistent with experimental data indicating that this

protein associates with 60S ribosomal subunit precursors, and is potentially involved in ribosome biogenesis [40,41]. Taken together, these results suggest that YIL096C is a novel rRNA MTase.

For the five non-periodic and nine periodic MTases lacking predicted substrate based on decision tree, we based our predictions on fold assignment only. Consequently, using fold assignment for the catalytic domains, we predicted the substrate specificity for all the remaining putative SPOUT and SET domain fold MTases, including three non-periodic (YHL039W, YJL105W and YKR029C, predicted to be protein MTases) and two periodic (YPL165C and YOR021C, predicted to be protein and RNA MTases, respectively).

Altogether, we predicted the substrate specificity of 24/33 putative MTases, and identified 15 new protein MTases and nine new RNA MTases; thus, significantly increasing our knowledge regarding methylation in yeast (Table 1). Finally, our classification shows that the *S. cerevisiae* methyltransferome embraces 32 protein MTases, 34 RNA MTases, eight small molecule MTases, three lipid MTases and nine Rossmann-like fold MTases with, as yet, unknown substrate specificity (Figure 1B).

Experimental verification

Firstly, we analyzed whether YIL096C, identified here as novel MTase, is able to bind a methyl group donor, using UV crosslinking [31,42]. Recombinant YIL096C protein was thus exposed to UV light in the presence of [3 H] AdoMet. The crosslink product was detected on a tritium screen (Figure 5) providing that AdoMet can bind both to YIL096C and HMT1 (known MTase used as positive control) but not TEV protease (negative control).

To validate the approach for substrate specificity prediction, we performed MTase activity assays for two *S. cerevisiae* proteins predicted to be Rossmann-like protein MTases: YBR271W and YLR285W (NNT1). Briefly, purified recombinant proteins were incubated with native total cell extracts from the wild-type and respective knockout strains in the presence of tritium-labeled AdoMet. The reaction products were then analyzed by SDS-

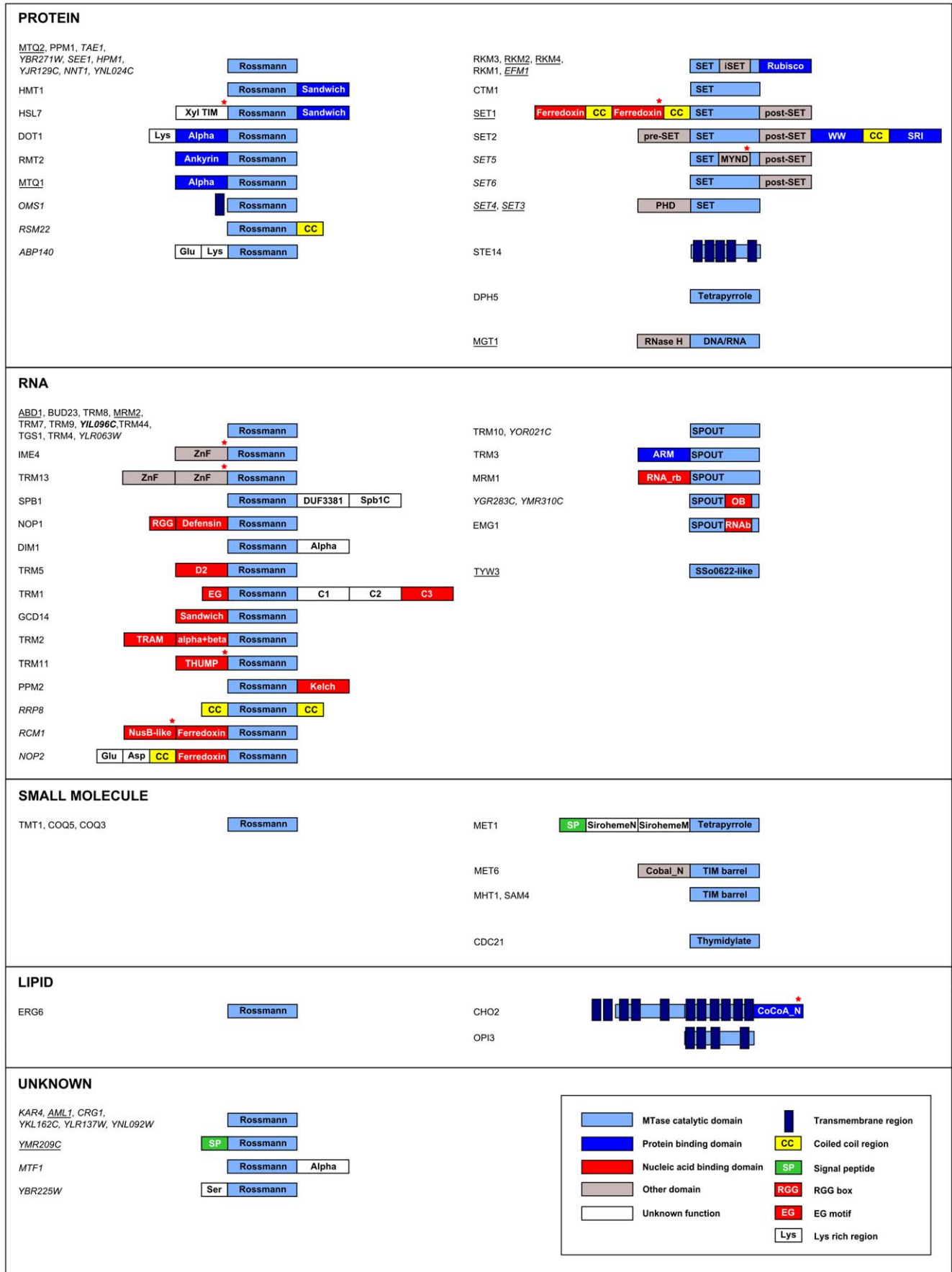


Figure 2. Domain architecture of *S. cerevisiae* MTases. The MTases were grouped according to their common substrate specificities (e.g., protein, RNA, small molecule or lipid) and the fold of catalytic domain. Known MTases with experimentally determined substrate specificity are shown in a regular font, putative MTases in italics, and newly detected MTase in bold. Non-periodic MTases are underlined. The new domains identified in this study are marked with a red asterisk. Sandwich, beta sandwich; Xyl TIM, TIM beta/alpha-barrel belonging to the Xylose isomerase-like superfamily; Alpha, α -helical domain; Ankyrin, Ankyrin repeats; ZnF, zinc finger; Spb1C, Spb1 C-terminal domain; Defensin, defensin-like fold; iSET, SET-inserted domain; Rubisco, Rubisco LSMT C-terminal-like domain; SRI, SET2 Rpb1 interacting domain; PHD, PHD zinc finger; DNA/RNA, DNA/RNA-binding 3-helical bundle; RNase H, RNase H-like domain; ARM, ARM repeat; RNA_rb, RNA ribose binding domain; SirohemeN, Siroheme synthase N-terminal domain-like; SirohemeM, Siroheme synthase middle domain-like; CobaL_N, Cobalamin-independent synthase N-terminal domain; CoCoA_N, Calcium binding and coiled-coil domain-like (N-terminal); OB, OB-fold domain; RNAb, RNA binding domain.
doi:10.1371/journal.pone.0023168.g002

PAGE followed by autoradiography (Figure 6). The presence of protein methylation products confirmed that YBR271W and YLR285W are protein MTases. Interestingly, YLR285W was previously assigned as a putative nicotinamide N-methyltransferase based on distant similarity to human NNMT [43]. This suggests that a simple sequence comparison approach cannot correctly predict the substrate specificity of Rossmann-like fold MTases. Our data indicate that YBR271W modifies several proteins, while YLR285W appears to have only one specific protein substrate. In both cases, methylated proteins were detected only when the deletion strains were used (Figure 6, lane 1), which strongly suggests that these modifications are stable. Interestingly, YBR271W is predicted to be a part of rRNA and ribosome biosynthesis (RRB) regulon [40] and is also expected, from a genome-wide in vivo screen (PCA) [44], to interact with RPC34, an RNA polymerase III subunit C34, a key determinant in Pol III recruitment. Methylation of RPC34 (Figure 6, one of the observed bands corresponds to the molecular weight of RPC34, which is 36 kDa) would be in agreement with the predicted role of

YBR271W in rRNA biosynthesis, because Pol III is responsible for 5S rRNA synthesis. As expected, we also found protein methylation patterns matching known substrates for HMT1, but not for RNA MTase TRM4 (Figure 6, the smear at the bottom of the gel represents tRNA). These results strongly support the theoretical model used in this study to predict the substrate specificity of putative MTases.

Conclusion

Identification of the *S. cerevisiae* methyltransferome, together with its structural and substrate specificity classification, has enabled us to shed new light on enzymatic methylation in yeast. In *S. cerevisiae*, methylation is carried out by 86 MTases among which are 34 RNA MTases, 32 protein MTases, eight small molecule MTases, three lipid MTases and nine MTases with unknown substrate specificity. These MTases may adopt up to nine different folds within the catalytic domain; however, as may be expected, the Rossmann-like fold is the most commonly used scaffold for the methyl transfer reaction. In addition, genes encoding *S. cerevisiae*

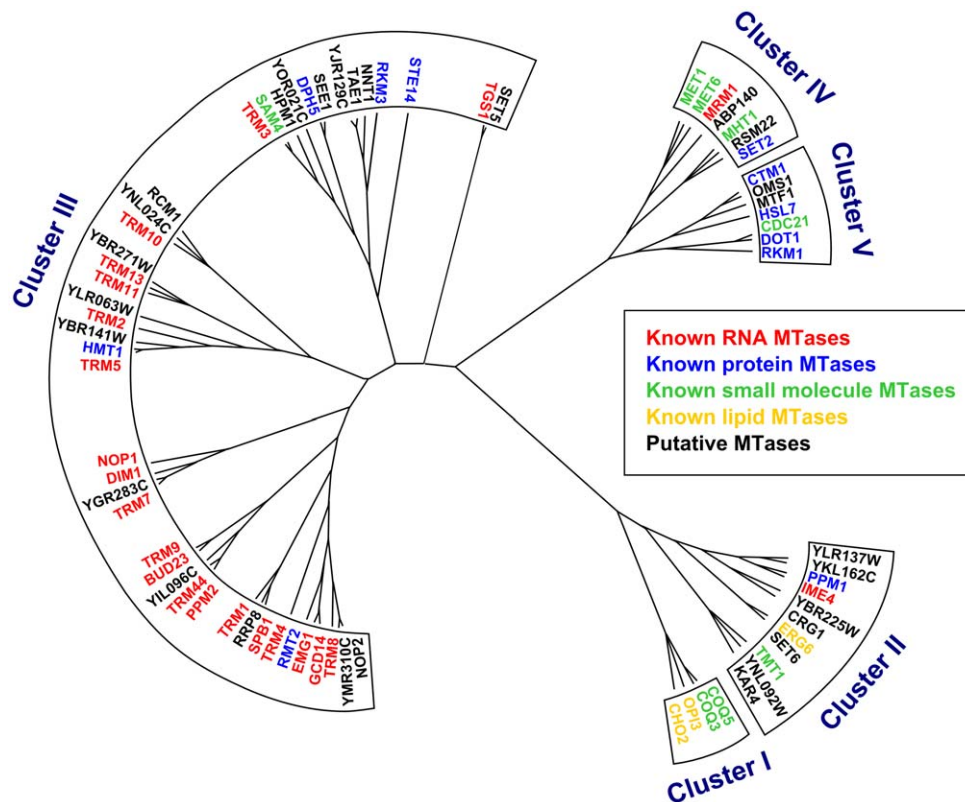


Figure 3. Hierarchical clustering tree for all *S. cerevisiae* periodic MTases. Seventy-two periodic MTases were divided into five clusters, each containing MTases with similar expression profiles during the Yeast Metabolic Cycle (YMC). Branch lengths correspond to correlation coefficients of gene expression profiles during the YMC obtained from SCEPTRANS.
doi:10.1371/journal.pone.0023168.g003

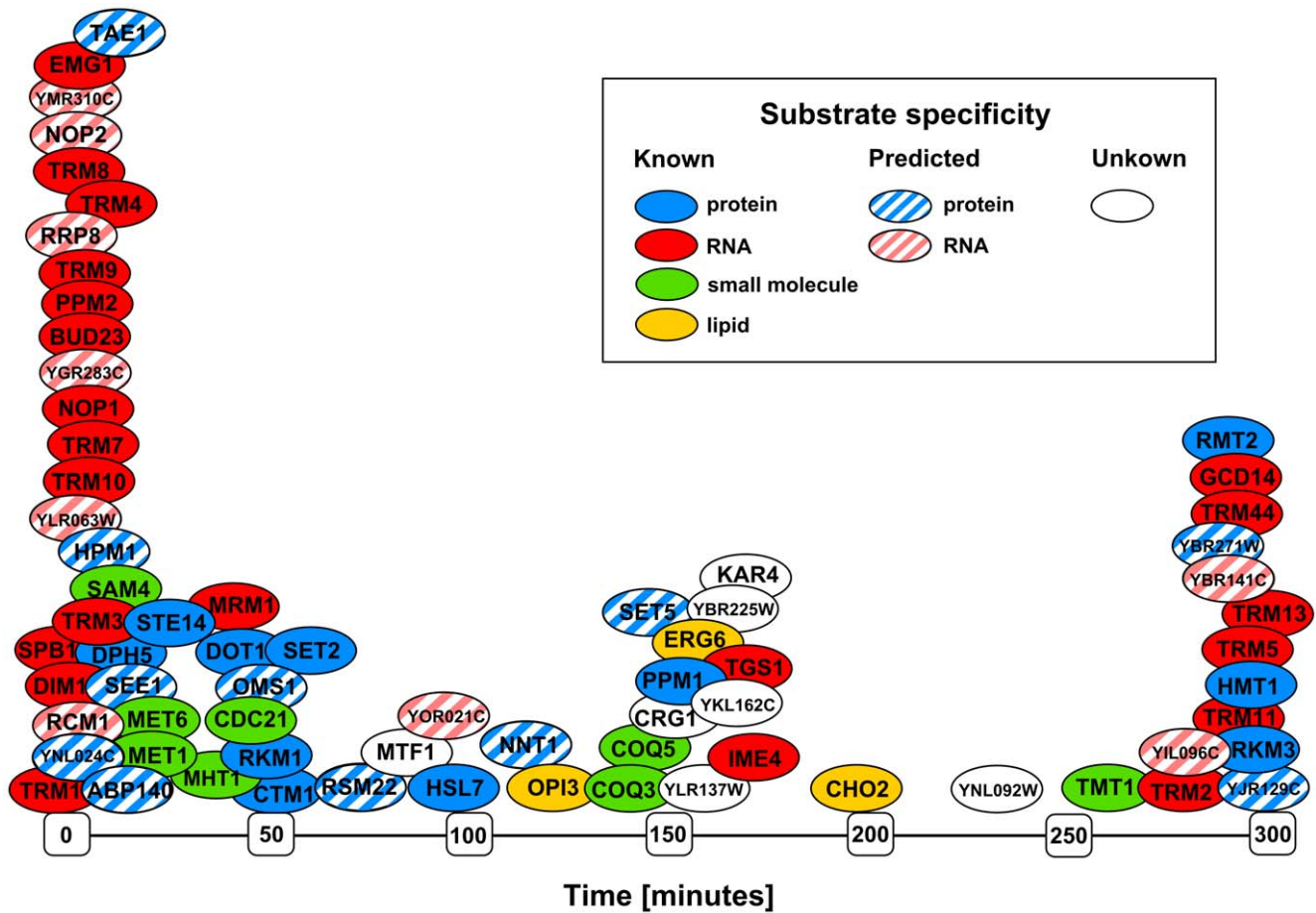


Figure 4. Metabolic cycle-dependent expression of *S. cerevisiae* periodic MTases. Each MTase is positioned at its gene expression peak within the YMC (which lasts 300 min). doi:10.1371/journal.pone.0023168.g004

MTases are almost uniformly distributed across all chromosomes, with the exception of chromosomes I and VI, where they are not present (Figure 7).

It should be noted that prediction of substrate specificity for Rossmann-like fold MTases (25 of 33 putative MTases in *S. cerevisiae*) based on simple sequence similarity is not feasible. Specifically, while SPOUT or SET domain folds seem to have

clearly defined substrate specificity (RNA and protein, respectively), Rossmann-like fold MTases perform all types of methylation. For example, PPM1 and PPM2, which share more than 30% sequence identity within their catalytic domains, are protein and tRNA Rossmann-like fold MTases, respectively. We discovered that gene expression profile during the YMC, fold assignment, max pI and protein localization all statistically significantly

Table 2. Decision tree rules used for predicting MTase substrate specificity.

Substrate	Max pI	Cluster	Fold	Localization	R	P	F	N
Protein	<8	III or V	Rossmann-like, SET domain, Tetrapyrrole methylase or Transmembrane	not mitochondrial	91%	100%	0.95	11
	≥8	IV or V	Rossmann-like or SET domain	not ER				
tRNA or rRNA	≥8	III			90%	95%	0.92	8
Small molecule	<8	I or IV			62%	100%	0.77	0

Decision tree rules were based on the maximum isoelectric point (Max pI), similarity of expression profiles (Cluster; the numbering of clusters that group MTases with similar gene expression profiles is given in Figure 3), fold assignment (Fold) and cellular localization (Localization). Columns R, P and F show the Recall, Precision and F-measure values calculated for each substrate specificity class with non-zero R. The number of MTases with the substrate specificity predicted by each of the rules is shown in column N.

doi:10.1371/journal.pone.0023168.t002

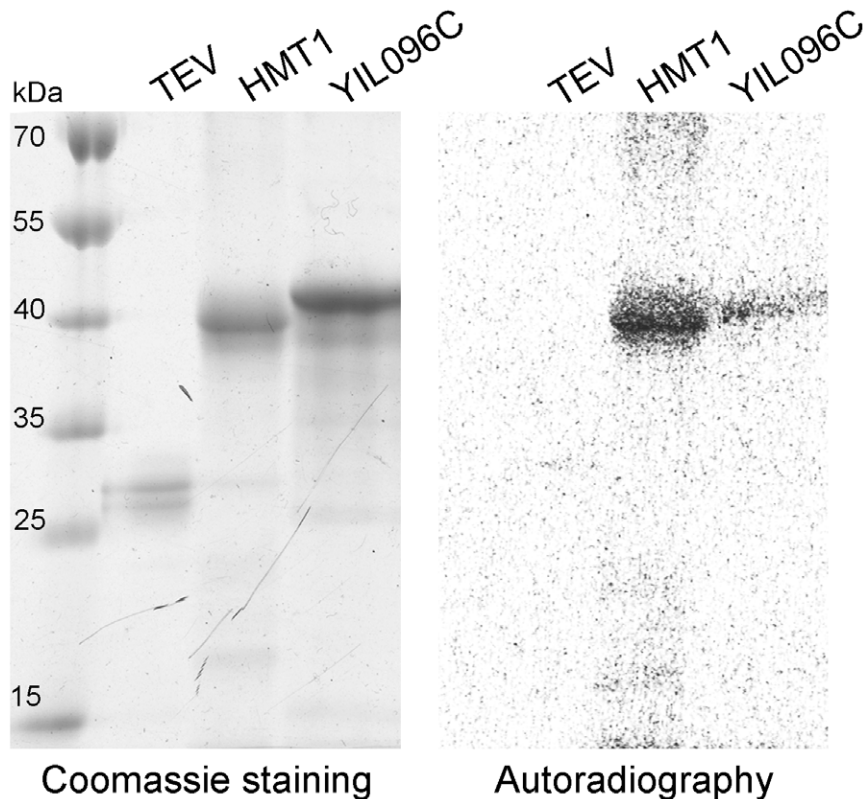


Figure 5. Putative MTase YIL096C binds AdoMet. Purified YIL096C (with H1StagSUMO), HMT1 and TEV protease were exposed to UV light in the presence of [3 H] AdoMet. Both Coomassie stained proteins (left panel) and the autoradiography of crosslink products (right panel) are shown. HMT1 (known MTase) and TEV protease were used as positive and negative controls, respectively.
doi:10.1371/journal.pone.0023168.g005

correlate with the type of MTase substrate and thus can be used to predict MTase substrate specificity. We proposed a simple set of decision rules which allowed to assign general substrates to 24 putative MTases, including 16 of the Rossmann-like fold. We confirmed our predictions experimentally in both cases tested, while in some other cases there is a strong supporting evidence in the literature (e.g. for newly detected RNA MTase, YIL096C). Our results provide the basis for more detailed biochemical analyses of individual MTases and the identification of their specific protein and RNA substrates.

Materials and Methods

Identification of the methyltransferome

Initially, both known and putative MTases were selected from various databases, including the Saccharomyces Genome Database (SGD) [32] and the catalogued protein families (PFAM [45], COG and KOG [46]) and structures (PDB [47] and SCOP [48]). The databases were searched using the term “methyl” as a text query and the list of hits was screened manually. The PDB database was also searched using E.C. number 2.1.1., which corresponds to MTase function. The obtained data set was used for further comprehensive searches against the whole *S. cerevisiae* proteome using Meta-BASIC [27], a highly sensitive method for distant homology detection based on the comparison of meta-profiles (sequence profiles enriched with predicted secondary structures). Consequently, novel MTases, which were not present in the initial set, were identified using the Gene Relational DataBase (GRDB) system. GRDB includes pre-calculated Meta-

BASIC connections between 11,127 PFAM, 10,361 KOG and COG families, 20,877 proteins of known structure (PDB90; representatives from the PDB database, filtered at 90% sequence identity) and 6719 *S. cerevisiae* proteins (the complete *S. cerevisiae* proteome). Each family, structure and *S. cerevisiae* protein in the system was represented by: (i) its sequence (for *S. cerevisiae* proteins and PDB90) or consensus sequence (for PFAM, COG and KOG families), (ii) its sequence profile generated with PSI-BLAST [28] (3 iterations, inclusion threshold 0.001) using the NCBI non-redundant protein sequence database derivative (NR70), and (iii) its secondary structure, predicted using PSI-PRED [49].

The search strategy was based on the concept of transitivity, where each newly identified PFAM, KOG and COG family, PDB structure, or *S. cerevisiae* protein was used in further Meta-BASIC searches until no new additional MTase hits were found. In addition to the highly reliable Meta-BASIC predictions with scores above 40 (corresponds to E-value <0.05), all hits with scores between 30 and 40 were also considered to identify any potentially correct predictions that may have been placed among the unreliable or incorrect ones. These potentially correct predictions were selected based on the manual assessment of conservation of the core secondary structure elements and sequence motifs deemed critical for the given fold and MTase function.

To confirm any non-trivial predictions that met the above criteria, the corresponding sequences were submitted to the Protein Structure Prediction Meta Server (<http://meta.bioinfo.pl>), which integrates various top-of-the-line fold recognition methods. The models generated by these methods were analyzed with 3D-Jury [33], a meta-predictor that uses a consensus approach to

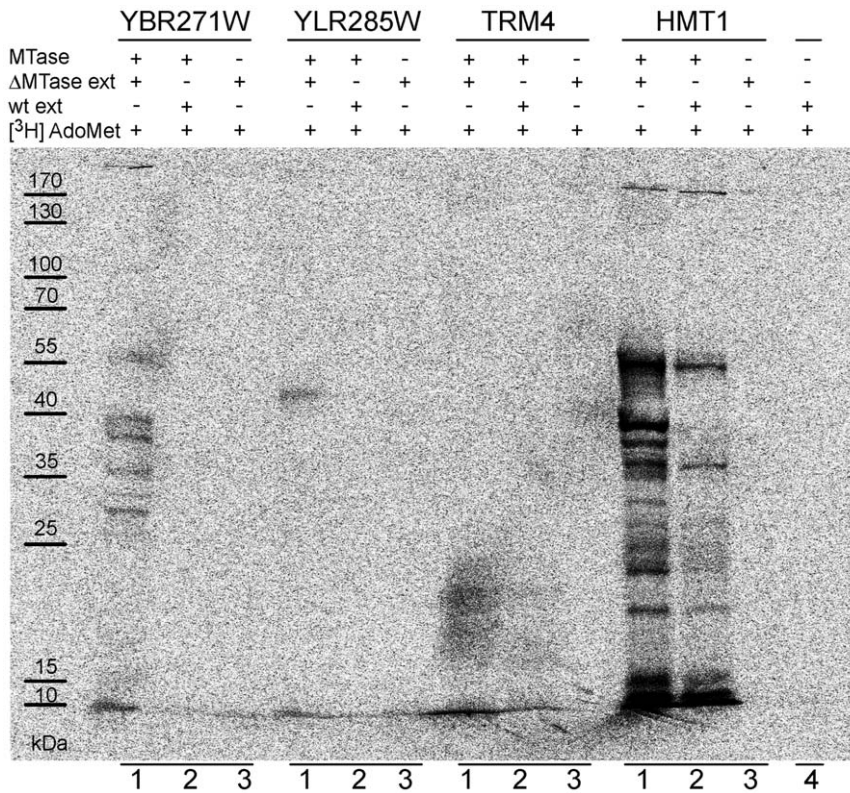


Figure 6. YBR271W and YLR285W (NNT1) are protein MTases. Recombinant proteins (MTases) were incubated with native yeast extracts from the respective knockout strains (Δ MTase ext) and [³H] AdoMet (lane 1). Reaction products were resolved on SDS-PAGE gel and exposed to tritium screen. To test the specificity of these reactions, analyzed proteins were also incubated with yeast extract from the wild-type strain (wt ext) and [³H] AdoMet (lane 2). As a control, yeast extracts from knockout and wild-type strains were incubated with [³H] AdoMet only (lanes 3 and 4). HMT1 (a protein MTase) and TRM4 (an RNA MTase) were used as positive and negative controls, respectively. doi:10.1371/journal.pone.0023168.g006

select the most abundant models. Predictions were deemed reliable when the assigned scores were above a confidence threshold of 50 [50].

Fold assignment and domain architecture analysis

The *S. cerevisiae* MTases were divided into distinct structural groups based on the structural similarity (fold) of their catalytic domains. The catalytic domain of each MTase was first identified with Meta-BASIC, followed by fold assignment using 3D-Jury and SCOP classification. To detect additional protein domains, sequence fragments (after removal of the catalytic MTase domain) were also analyzed using Meta-BASIC coupled, for non-trivial predictions, with 3D-Jury. Confident assignments were selected based on reliable Meta-BASIC and 3D-Jury scores and conservation of critical secondary structure elements and sequence motifs. Transmembrane regions were predicted using ConPred II [51], Phobius [52] and TOPCONS [53], and only those regions identified by all these methods as “highly probable” were accepted. Coiled coils were detected with Marcoil 1.0 [54], signal peptides with SignalP 3.0 [55] and functionally important motifs with PROSITE [56].

Prediction of substrate specificity

Information regarding the substrate specificity of the known MTases was obtained from SGD and literature searches. Predictions of general substrate specificity (protein, RNA, lipid and small molecule) of the putative MTases were based on similarities in their gene expression patterns during the YMC,

calculated pI, fold assignment and cellular localization. The periodic MTases were selected based on a visual examination of temporal expression profiles at SCEPTRANS [39] (Table 1). Then, a matrix of correlation coefficients of expression profiles for periodic MTases was obtained from the SCEPTRANS and was used to group these MTases into 5 clusters, each consisting of MTases with similar expression profiles during the YMC. The β -values of enrichment of MTases with a given substrate specificity was computed using a hypergeometric probability distribution. The results confirmed potential usefulness of this clustering for predicting substrate specificity of periodic MTases. Also, three pI values were calculated for each MTase (for the whole protein, the catalytic domain and the remaining regions) using the ExpASY tool (ProtParam) [57]. The maximum of these three values was defined as max pI. The cellular localization data obtained from SGD was also considered. For MTases with unknown cellular localization, this data was predicted using the consensus of the results obtained from three localization prediction servers (WoLF PSORT [58], BaCellLo [59] and MultiLoc [60]).

The groupings derived from hierarchical clustering for known periodic MTases, together with the maximum pI value, the fold assignment and the cellular localization were used to propose a decision tree. Removing each MTase from the training set (all periodic known MTases) and trying to predict its substrate specificity based on the newly generated decision tree allowed us to assess how dependent the results were upon the training set. Using “leave-one-out” cross-validation allowed for the correct prediction of the general substrate specificity for 75% of the whole

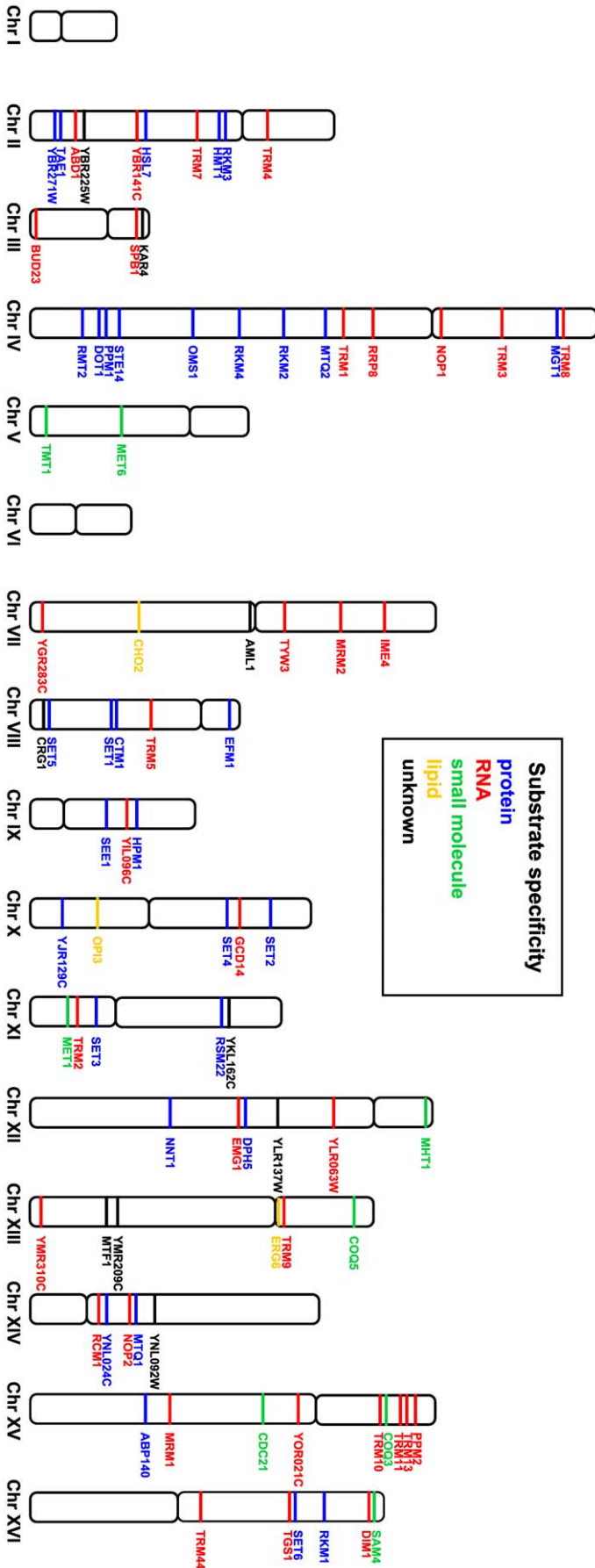


Figure 7. Localization of MTase genes within the *S. cerevisiae* genome. MTases are colored according to their substrate specificity. doi:10.1371/journal.pone.0023168.g007

training set. The rules used for the prediction of substrate specificity for the periodic putative MTases are listed in Table 2. For each decision rule, the Recall, Precision and F-measures were calculated as follows: Recall = TP/(TP + FN), Precision = TP/(TP + FP), F-measure = 2(Recall*Precision)/(Recall + Precision), where TP, FN and FP represent true-positives, false-negatives and false-positives, respectively. Predicting the substrate specificity of non-periodic putative MTases was based on fold assignment (SPOUT: RNA MTase, SET domain: protein MTase) only.

Strains and media

The following yeast strains (Euroscarf) were used in this study: BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*), BY4741 ΔYBR271W, BY4741 ΔYLR285W, BY4741 ΔYBR034C (ΔHMT1), and BY4741 ΔYBL024W (ΔTRM4). The standard yeast genetic methods and selective growth media have been previously described [61].

Protein expression and purification

The putative MTases, YIL096C, YBR271W and YLR285W, along with the HMT1 and TRM4 MTases (controls) were produced in *E. coli* (BL21-CodonPlus-RIL strain) as N-terminal HISTagSUMO tag fusions using LB medium and overnight IPTG inductions at 18°C. The bacterial pellets were lysed by sonication in buffer A (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM imidazole, 10 mM 2-mercaptoethanol) and purified on His-Trap FF Crude columns (GE Healthcare). The proteins were further purified by size-exclusion chromatography on a Superdex 75 10/300 GL column (GE Healthcare) in buffer containing 10 mM Tris-HCl pH 8.0 and 150 mM NaCl. Finally, glycerol was added to the protein aliquotes (30% final concentration), which were then stored at -80°C. The purity and quantity of the proteins was assessed by SDS-PAGE.

References

- Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, et al. (1996) S-Adenosylmethionine and methylation. *Faseb J* 10: 471–480.
- Cheng X, Blumenthal RM (1999) S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions. Singapore: World Scientific Publishing Company.
- Walsh CT (2005) Posttranslational Modification of Proteins: Expanding Nature's Inventory. Englewood, Colorado: Roberts and Co. Publishers.
- Paik WK, Paik DC, Kim S (2007) Historical review: the field of protein methylation. *Trends Biochem Sci* 32: 146–152.
- Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293: 1074–1080.
- Martin C, Zhang Y (2005) The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 6: 838–849.
- Dunin-Horkawicz S, Czerwoniec A, Gajda MJ, Feder M, Grosjean H, et al. (2006) MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res* 34: D145–149.
- Kiss T (2001) Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *Embo J* 20: 3617–3622.
- Piekna-Przybylska D, Decatur WA, Fournier MJ (2008) The 3D rRNA modification maps database: with interactive tools for ribosome analysis. *Nucleic Acids Res* 36: D178–183.
- Chow CS, Lamichhane TN, Mahto SK (2007) Expanding the nucleotide repertoire of the ribosome with post-transcriptional modifications. *ACS Chem Biol* 2: 610–619.
- Gustilo EM, Vendeix FA, Agris PF (2008) tRNA's modifications bring order to gene expression. *Curr Opin Microbiol* 11: 134–140.
- Shuman S (2002) What messenger RNA capping tells us about eukaryotic evolution. *Nat Rev Mol Cell Biol* 3: 619–625.
- Clancy MJ, Shambaugh ME, Timpste CS, Bokar JA (2002) Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res* 30: 4509–4518.
- Mouaikel J, Bujnicki JM, Tazi J, Bordonne R (2003) Sequence-structure-function relationships of Tgs1, the yeast snRNA/snoRNA cap hypermethylase. *Nucleic Acids Res* 31: 4899–4909.
- Franke J, Gehlen J, Ehrenhofer-Murray AE (2008) Hypermethylation of yeast telomerase RNA by the snRNA and snoRNA methyltransferase Tgs1. *J Cell Sci* 121: 3553–3560.
- McCammon MT, Hartmann MA, Bottema CD, Parks LW (1984) Sterol methylation in *Saccharomyces cerevisiae*. *J Bacteriol* 157: 475–483.
- Carreras CW, Santi DV (1995) The catalytic mechanism and structure of thymidylate synthase. *Annu Rev Biochem* 64: 721–762.
- Kodaki T, Yamashita S (1987) Yeast phosphatidylethanolamine methylation pathway. Cloning and characterization of two distinct methyltransferase genes. *J Biol Chem* 262: 15428–15435.
- Hansen J, Muldbjerg M, Cherest H, Surdin-Kerjan Y (1997) Siroheme biosynthesis in *Saccharomyces cerevisiae* requires the products of both the MET1 and MET8 genes. *FEBS Lett* 401: 20–24.
- Schubert HL, Blumenthal RM, Cheng X (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem Sci* 28: 329–335.
- Katz JE, Diakic M, Clarke S (2003) Automated identification of putative methyltransferases from genomic open reading frames. *Mol Cell Proteomics* 2: 525–540.
- Dixon MM, Huang S, Matthews RG, Ludwig M (1996) The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B12. *Structure* 4: 1263–1275.
- Schubert HL, Wilson KS, Raux E, Woodcock SC, Warren MJ (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nat Struct Biol* 5: 585–592.
- Anantharaman V, Koonin EV, Aravind L (2002) SPOUT: a class of methyltransferases that includes spoU and trmD RNA methylase superfamilies, and novel superfamilies of predicted prokaryotic RNA methylases. *J Mol Microbiol Biotechnol* 4: 71–75.
- Dillon SC, Zhang X, Trievel RC, Cheng X (2005) The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol* 6: 227.

UV crosslinking

Recombinant proteins (10–25 µg) were mixed with 2 µCi [³H] AdoMet (80 Ci/mmol, Hartmann Analytic GmbH) in a buffer containing 10 mM potassium phosphate pH 7.0, 100 mM NaCl, 2 mM EDTA, 1 mM dithiothreitol, 5% glycerol in PCR tube [42]. The reaction mixture was exposed to UV irradiation in a UVC 500 crosslinker (Amersham Bioscience) for 10 min on ice, 3 cm from the light source. The products were run on a 12% SDS-PAGE gel. After Coomassie blue staining, dried gel was exposed to tritium screen (GE Healthcare) for 72 hr at RT.

In vitro methylation assay

Yeast whole-cell extracts were prepared as previously described [62]. Recombinant proteins (5 µg) were incubated with 30 µg of native yeast extract (from a wild-type strain and strain from which the gene encoding the analyzed protein had been deleted) in the presence of [³H] AdoMet (0.5 µCi/reaction) in 15 µl of reaction buffer (10 mM HEPES pH 8.0, 2 mM EDTA, 50 mM KCl, 1 mM DTT). Protein extracts were incubated at RT for 1 hr before being diluted 2-fold in Laemmli buffer and resolved on a 12% SDS-PAGE gel. Gel was stained with Coomassie blue, dried and exposed overnight to tritium screen.

Acknowledgments

We are grateful to Dr Darek Plewczynski for helpful suggestions and discussion. During preparation of this manuscript, three putative *S. cerevisiae* MTases with the Rossmann-like fold: TAE1 [63], SEE1 [64], HPM1 [65], and one putative SET domain MTase, EFM1 [64], have been shown experimentally to methylate proteins. These data are in agreement with our assignments and confirm our theoretical approach for predicting substrate specificity.

Author Contributions

Conceived and designed the experiments: AD KG. Performed the experiments: TW JK JT. Analyzed the data: TW JK JT LK LR AK MR AD KG. Wrote the paper: TW MR AD KG.

26. Tu BP, Kudlicki A, Rowicka M, McKnight SL (2005) Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science* 310: 1152–1158.
27. Ginalski K, von Grotthuss M, Grishin NV, Rychlewski L (2004) Detecting distant homology with Meta-BASIC. *Nucleic Acids Res* 32: W576–581.
28. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–3402.
29. Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Res* 32: W327–331.
30. Niewmierzycka A, Clarke S (1999) S-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase. *J Biol Chem* 274: 814–824.
31. Petrossian T, Clarke S (2009) Bioinformatic Identification of Novel Methyltransferases. *Epigenomics* 1: 163–175.
32. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, et al. (1998) SGD: *Saccharomyces Genome Database*. *Nucleic Acids Res* 26: 73–79.
33. Ginalski K, Elofsson A, Fischer D, Rychlewski L (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* 19: 1015–1018.
34. Petrossian TC, Clarke SG (2009) Multiple Motif Scanning to identify methyltransferases from the yeast proteome. *Mol Cell Proteomics* 8: 1516–1526.
35. Brown MA, Sims RJ, III, Gottlieb PD, Tucker PW (2006) Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex. *Mol Cancer* 5: 26.
36. Thompson EC, Travers AA (2008) A *Drosophila* Smyd4 homologue is a muscle-specific transcriptional modulator involved in development. *PLoS One* 3: e3008.
37. Yang CK, Kim JH, Stallcup MR (2006) Role of the N-terminal activation domain of the coiled-coil coactivator in mediating transcriptional activation by beta-catenin. *Mol Endocrinol* 20: 3251–3262.
38. Anderson DJ, Hetzer MW (2008) Shaping the endoplasmic reticulum into the nuclear envelope. *J Cell Sci* 121: 137–142.
39. Kudlicki A, Rowicka M, Otwinowski Z (2007) SCEPTRANS: an online tool for analyzing periodic transcription in yeast. *Bioinformatics* 23: 1559–1561.
40. Wade CH, Umbarger MA, McAlear MA (2006) The budding yeast rRNA and ribosome biosynthesis (RRB) regulon contains over 200 genes. *Yeast* 23: 293–306.
41. Saveanu C, Namane A, Gleizes PE, Lebreton A, Rousselle JC, et al. (2003) Sequential protein association with nascent 60S ribosomal particles. *Mol Cell Biol* 23: 4449–4460.
42. Subbaramaiah K, Simms SA (1992) Photolabeling of CheR methyltransferase with S-adenosyl-L-methionine (AdoMet). Studies on the AdoMet binding site. *J Biol Chem* 267: 8636–8642.
43. Anderson RM, Bitterman KJ, Wood JG, Medvedik O, Sinclair DA (2003) Nicotinamide and PNC1 govern lifespan extension by caloric restriction in *Saccharomyces cerevisiae*. *Nature* 423: 181–185.
44. Tarassov K, Messier V, Landry CR, Radinovic S, Serna Molina MM, et al. (2008) An in vivo map of the yeast protein interactome. *Science* 320: 1465–1470.
45. Finn RD, Mistry J, Tate J, Coghill P, Heger A, et al. (2010) The Pfam protein families database. *Nucleic Acids Res* 38: D211–222.
46. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
47. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242.
48. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247: 536–540.
49. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292: 195–202.
50. Ginalski K, Rychlewski L (2003) Detection of reliable and unexpected protein fold predictions using 3D-Jury. *Nucleic Acids Res* 31: 3291–3292.
51. Arai M, Mitsuke H, Ikeda M, Xia JX, Kikuchi T, et al. (2004) ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. *Nucleic Acids Res* 32: W390–393.
52. Kall L, Krogh A, Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 338: 1027–1036.
53. Bernsel A, Viklund H, Hennerdal A, Elofsson A (2009) TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res* 37: W465–468.
54. Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* 18: 617–625.
55. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2: 953–971.
56. de Castro E, Sigrist CJ, Gattiker A, Bulliard V, Langendijk-Genevaux PS, et al. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res* 34: W362–365.
57. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server; Walker JM, ed. Totowa, New Jersey: Humana Press.
58. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, et al. (2007) WoLF PSORT: protein localization predictor. *Nucleic Acids Res* 35: W585–587.
59. Pierleoni A, Martelli PL, Fariselli P, Casadio R (2006) BaCelLo: a balanced subcellular localization predictor. *Bioinformatics* 22: e408–416.
60. Hoglund A, Donnes P, Blum T, Adolph HW, Kohlbacher O (2006) MultiLoc: prediction of protein subcellular localization using N-terminal targeting sequences, sequence motifs and amino acid composition. *Bioinformatics* 22: 1158–1165.
61. Rose MD, Winston F, Hieter P (1999) *Methods in Yeast Genetics: A Laboratory Course Manual*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
62. Seraphin B, Kretzner L, Rosbash M (1988) A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *Embo J* 7: 2533–2538.
63. Webb KJ, Lipson RS, Al-Hadid Q, Whitelegge JP, Clarke SG (2010) Identification of protein N-terminal methyltransferases in yeast and humans. *Biochemistry* 49: 5225–5235.
64. Lipson RS, Webb KJ, Clarke SG (2010) Two novel methyltransferases acting upon eukaryotic elongation factor 1A in *Saccharomyces cerevisiae*. *Arch Biochem Biophys* 500: 137–143.
65. Webb KJ, Zurita-Lopez CI, Al-Hadid Q, Laganowsky A, Young BD, et al. (2010) A novel 3-methylhistidine modification of yeast ribosomal protein Rpl3 is dependent upon the YIL110W methyltransferase. *J Biol Chem* 285: 37598–37606.