Global mapping of CARM1 substrates defines enzyme specificity and substrate recognition

Evgenia Shishkova1,*, Hao Zeng2,†, Fabao Liu2, Nicholas W. Kwiecien3, Alexander S. Hebert3, Joshua J. Coon1,4 & Wei Xu2

Protein arginine methyltransferases (PRMTs) introduce arginine methylation, a post-translational modification with the increasingly eminent role in normal physiology and disease. PRMT4 or coactivator-associated arginine methyltransferase 1 (CARM1) is a propitious target for cancer therapy; however, few CARM1 substrates are known, and its mechanism of substrate recognition is poorly understood. Here we employed a quantitative mass spectrometry approach to globally profile CARM1 substrates in breast cancer cell lines. We identified >130 CARM1 protein substrates and validated in vitro >90% of sites they encompass. Bioinformatics analyses reveal enrichment of proline-containing motifs, in which both methylation sites and their proximal sequences are frequently targeted by somatic mutations in cancer. Finally, we demonstrate that the N-terminus of CARM1 is involved in substrate recognition and nearly indispensable for substrate methylation. We propose that development of CARM1-specific inhibitors should focus on its N-terminus and predict that other PRMTs may employ similar mechanism for substrate recognition.
Protein arginine methylation is an abundant post-translational modification (PTM), catalysed by nine mammalian arginine methyltransferases (PRMTs). A recent study reveals that ~7% of all arginines are methylated. This is comparable to 9% of serine residues being phosphorylated and 7% of lysine residues being ubiquitinated. Via transferring a methyl group from S-adenosyl-l-methionine (SAM) to the side chain of arginine residues, PRMTs catalyse formation of three final product types: o-N^0-atomethylarginine by PRMT7, o-N^0,2-N^0-asymmetric dimethylarginine (ADMA) by Type I PRMTs, and o-N^0,N^0-symmetric dimethylarginine by Type II PRMTs. While arginine methylation does not alter amino-acid charge, it does increase its bulkiness and hydrophobicity. This could affect protein–protein and protein–nucleic acid interactions, as a consequence, impacting myriad biological pathways, including transcription, RNA metabolism, DNA repair, and among others. Aberrant expression and/or enzymatic activity of multiple PRMTs are associated with human cancers; however, the functional significance of arginine methylation in oncogenic processes is poorly understood. This problem is further confounded by a lack of known cancer-relevant PRMT substrates. Similarly, the absence of the full-length enzyme structure co-crystallized with known substrates hampers the effective design of PRMT-specific inhibitors.

Here we focus on PRMT4, also known as coactivator-associated arginine methyltransferase 1 (CARM1), an indispensable enzyme in mammals. Both CARM1 knockout (KO) and enzyme-inactive CARM1 knock-in mice die at birth and display identical developmental defects; that is, defects in T-cell development and adipocyte differentiation, underscoring the essentiality of CARM1’s enzymatic function in physiology. Emerging evidence supports the pathological role of CARM1 in human disease, particularly in cancer. CARM1 is overexpressed in a variety of cancer types and its higher expression correlates with poor prognosis. Like several other PRMTs, CARM1 functions as both a coactivator and a methyltransferase. CARM1 interacts with a plethora of transcription factors (for example, p53, oestrogen receptor and E2F1), modulating gene expression and contributing to cancer progression. Besides histone H3 (ref. 18), CARM1 methylates non-histone substrates, likewise driving key oncogenic processes. For example, CARM1-mediated methylation of NCOA3 (ref. 19) and BAFl55 (ref. 20) promotes cancer progression and metastasis. Dissimilarly to PRMT1 and other PRMTs, CARM1 methylates p-histone substrates, where a well-characterized glycine and arginine-rich (GAR) methylation motif facilitates prediction of hundreds of protein substrates, in vivo substrate recognition motif(s) of CARM1 remain to be defined with fewer than 20 substrates known to date.

Given the prominance of CARM1 in oncogenesis, tremendous effort has been expended to design CARM1-specific inhibitors. However, no inhibitor has been shown to be effective in animal models. This is in large due to the lack of structural insight into the physical basis of substrate recognition by CARM1, like by all PRMTs in general. Members of the PRMT family share highly conserved central catalytic domain, but the primary sequences of N- and C-termini vary drastically. The catalytic core of CARM1, co-crystallized with several substrates, exhibits folding similar to that of other Type I PRMTs—the dimer form that accommodates SAM and peptide sequences. The N- and C-termini of CARM1, however, appear disordered, and the structure of the full-length CARM1 has not been solved to date. Intriguingly, when expressed alone, the N-terminal domain of CARM1 can be crystallized and displays a fold highly similar to a family of Drosophila-enabled/vasodilator-stimulated phosphoprotein homology 1 (EVH1) domain, a member of pleckstrin homology (PH) domain superfamily. Typically, EVH1 domains recognize and bind proline-rich sequences, however, the precise function of the N-terminal domain in CARM1 remains enigmatic.

In this study we employed high-resolution mass spectrometry (MS) and the newly developed ADMA antibody to globally profile CARM1 substrates in two human breast cancer cell lines. We distinguished a decrease in substrate methylation from the total protein changes induced by CARM1 loss, leading to the precise mapping of arginine methylation events regulated by CARM1 in vivo. Altogether, we identified over 300 CARM1-dependent ADMA sites, encompassed by ~130 novel bona fide CARM1 protein substrates. Many of these substrates have cancer-relevant functions and thus are possible mediators of CARM1’s oncogenic potential. In vitro methylation array confirmed the ability of CARM1 to methylate over 90% of the tested sequences. Further, informatic analysis revealed the presence of proline-rich motifs nearby CARM1 methylation sites. Both CARM1-methylated arginines and the surrounding recognition sequences were frequently targeted by somatic mutations in cancer, likely inducing reduction or complete abolishment of methylation by CARM1. Finally, we discovered that the N-terminal EVH1 domain of CARM1 is necessary and sufficient for substrate recognition and is required for methylation of most CARM1 substrates. This finding opens new routes in the design of CARM1-specific inhibitors and warrants functional investigation of the N-terminal domains of other PRMT family members in substrate recognition.

Results

Global profiling of CARM1 substrates in breast cancer cells. Our previous studies identified two CARM1 substrates, BAF155 (ref. 20) and MED12 (ref. 30), by individually selecting the differentially precipitated proteins by ADMA pan-specific antibody in CARM1 KO MCF7 cells and the parental cells with endogenous CARM1 expression. Here we aimed to globally profile CARM1 substrates in vivo. To do this, we immunoprecipitated ADMA-containing peptides using ADMA antibodies in two parental and CARM1 KO paired cell lines, MCF7 and MDA-MB-231, and mapped the sites of arginine methylation using nano-liquid chromatography tandem mass spectrometry (nanoLC-MS/MS) in biological triplicates (Fig. 1a). Our experimental design capitalized on the multiplexing capabilities of tandem mass tags (TMT), a technique that allowed us to simultaneously enrich and analyse ADMA-modified peptides from wild type and KO samples, assuring a complete overlap of peptide identifications between the two groups. This approach evades the common pitfall of PTM analysis—poor reproducibility. Note the process of correlating tandem mass spectra to peptide sequences can be confounded in cases where the spectra originate from methylated peptides. To counter this potential issue, we sequenced all MS/MS spectra corresponding to ADMA-containing peptides (deposited in PRIDE accession #PXD005871), confirming their accurate identification and ensuring correct localization of methylation sites. Neutral loss of dimethylamine, characteristic of ADMA peptides and not their isobaric symmetrically methylated counterparts, was systematically observed in the annotated spectra (Fig. 1b). Consideration of this loss improved achieved sequence coverage by ~4% (71.5 ± 21.7% and 75.2 ± 21.2% excluding and including the neutral loss, respectively), further fostering our confidence in identified peptides. Additionally, good reproducibility was observed among biological replicates across all experiments with median R^2 of 0.98 for both modified and unmodified peptides (Supplementary Fig. 1a).
Figure 1 | Global profiling of CARM1 substrates using quantitative mass spectrometry. (a) Experimental workflow for global detection of CARM1 substrates. Enrichment of ADMA—containing tryptic peptides from three biological replicates of parental and CARM1 KO cell lines—was followed by quantitative mass spectrometry using TMT. (b) Identification and quantification of ADMA-containing peptides and sites. The heat map displays hierarchical clustering using Pearson correlation of mean normalized log2 transformed intensities of ADMA containing peptides in three biological replicas. (c) Identification of putative CARM1 substrates from two breast cancer cell lines using three biological replicas of each sample. Volcano plots illustrate rates of somatic mutations at and nearby CARM1 methylation sites in human cancers. Frequencies of non-synonymous single nucleotide variants at or in the proximity of CARM1-regulated ADMA sites (± 5 nucleotides) were over twofold higher than those of non-modified arginines or a randomly selected residue (Fisher’s exact test P values of 5.6e-31 and 7.8e-24, respectively), according to the COSMIC. (e) Substrate interaction diagram (STRING 10.0) featuring four biological pathways (Reactome 2016) strongly enriched for the presence of putative CARM1 substrates (combined score > 5). Thickness of the lines radiating from CARM1 correlates to the pathway enrichment score (Supplementary Table 3). A black frame around a substrate indicates its causal implementation in cancer (COSMIC).
In both MCF7 and MDA-MB-231 cells, we observed pronounced reduction in the levels of ADMA-containing peptides on deletion of CARM1 (Fig. 1b and Supplementary Fig. 1b, respectively; Supplementary Data 1). Specifically, in CARM1 KO MCF7 cells, over 50% of the detected modified peptides, encompassing nearly 300 unique ADMA sites, decreased in abundance by greater than twofold, as compared to the parental cells. In both cell lines such drastic reduction in abundance uniquely affected modified peptides, as less than 1% of unmodified peptides exhibited changes comparable in magnitude (Supplementary Fig. 1c).

Because ADMA could be deposited by all Type I PRMTs, we segregated putative CARM1 substrates from substrates of other PRMTs by evaluating fold changes in response to CARM1 loss (Fig. 1c). The abundance of an ADMA site on PABP1, a known CARM1 substrate in MCF7 cells20,26, decreased by slightly over twofold. Thus, to stratify putative CARM1 substrates from those of other PRMTs, we selected ADMA sites that were reduced twofold or more in the CARM1 KO cells (Fig. 1c). To correctly categorize three MED12 peptides, harbouring known proteins that may be modified by CARM1.

Overall, we identified over 300 unique ADMA sites on 138 different substrates in parental and CARM1 KO cells (Supplementary Fig. 1c). Across both cell lines, only 4 substrates exhibited decreased abundance (~ twofold) upon CARM1 loss. We also noted that the levels of two other substrates exhibited moderate increases in abundance (less than threefold). By and large, however, nearly all the substrates showed no detectable modulation. We conclude that the observed reduction in ADMA peptides abundances were due to the loss of CARM1 catalysed methylation on arginine, rather than changes in protein abundance.

Finally, we examined levels of other PRMTs (PRMT1, 2, 5, 6 and 8) to ensure that changes in their abundance did not produce the observed reduction in ADMA-containing peptides. Together, mass spectrometric measurements and western blot analyses (Supplementary Data 2 and Supplementary Fig. 1d, respectively) confirmed that in both cell lines, levels of all PRMTs, but PRMT1 in MCF7 cells, were unchanged on CARM1 deletion. Note, PRMT8 was not detected by both MS and western blotting (data not shown), likely due to its brain-specific expression. The abundance of PRMT1 in MCF7 cells was slightly reduced by ~2.5-fold, and the implications of this change are discussed later.

Many CARM1 substrates have known oncogenic functions. Overall, we identified over 300 unique ADMA sites on 138 different proteins that may be modified by CARM1 in vivo; over 90% of them have not been previously reported as substrates of the enzyme. According to the Catalogue of Somatic Mutations in Cancer (COSMIC28), frequencies of non-synonymous single nucleotide variants at, and in the proximity of CARM1-regulated ADMA sites (+/-5 nucleotides), were over twofold higher as compared to the whole human proteome (Fig. 1d). This significant enrichment (Fisher’s exact test P values of 5.6e–31 and 7.8e–24) indicates that CARM1 methylation sites and recognition sequences are vulnerable to mutation in cancers and may provide a functional advantage. COSMIC analysis also reveals that 17 of the putative CARM1 substrates are causally implemented in cancer pathogenesis (Fig. 1e and Supplementary Table 1). Further, pathway enrichment analysis39 and protein–protein interaction40 analyses illuminated the critical role of these proteins in diverse biological pathways that are commonly deregulated in cancers (Fig. 1e and Supplementary Table 2). Unlike PRMT1 with most of its substrates involved in RNA processing and surveillance2, pronouncedly, 34 putative CARM1 substrates are involved in control of gene expression. Ten substrates possess functions in chromatin organization and enzymatic modifications. Various mRNA-processing pathways, such as mRNA stability, decay and splicing, also appeared over-represented. Interestingly, lipid metabolism via PPAR-alpha pathway and transcription regulation of white adipocyte differentiation were also identified, consistently with the reported functions of CARM1 in adipogenesis41. The functions of CARM1 substrates in HOX and TP53 signalling pathways were also mildly enriched. Together, identification of many cancer-relevant substrates across diverse biological pathways and the elevated mutation rate at, or near, CARM1-methylated arginines underscore the significance of substrate methylation by CARM1 in oncogenic processes.

Detection of proline-rich motifs among CARM1 substrates. With this extensive collection of CARM1 substrates, we sought to identify a consensus substrate recognition motif. Note no such motif confirmed in living cells presently exists1,5,6. A comparison of amino-acid frequencies in the human proteome and in the 11 amino-acid sequences centring on the putative CARM1 methylation sites revealed an over threefold enrichment of proline residues in the vicinity of the ADMA-bearing arginines (Fig. 2a). These results confirm the enrichment of proline among CARM1 methylation substrates previously inferred by Cheng et al.42; however, they dispute the significance of methionine and glycine residues in the so-called proline-, glycine- and methionine-rich (PGM) motif as the hallmark of CARM1 substrate recognition. Specifically, methionine was detected at frequencies typical for the human proteome. The abundance of glycine residues appeared slightly elevated, but the difference was not significant (P value of 0.13).

Using the MotifX algorithm43, we extracted an assortment of proline-containing motifs from the CARM1 methylation sequences (Fig. 2b and Supplementary Table 3). The resultant motifs feature proline residues in various positions up to five amino acids away from the central arginine. Several of these motifs are in good agreement with the termed ‘proline-rich arginine methylation’ motifs, previously reported in a global arginine methylation study44. As expected, the GAR motif and highly similar sequences, typically methylated by other PRMTs but not CARM1 (ref. 1), were detected in the vicinity of the methylation sites whose abundance was unaffected by the loss of CARM1 (Supplementary Fig. 2a and Supplementary Table 4). As our methodology did not permit direct measurements of abundance changes on individual ADMA sites in the context of multiply dimethylated peptides, we performed control analyses by separately extracting motifs from singly modified peptides. The detected motifs (Supplementary Fig. 2b,c) were nearly identical to a subset of motifs on Fig. 2b and Supplementary Fig. 2a, therefore further corroborating their enrichment.
High-throughput validation of CARM1 substrates. TET2, methylcytosine dioxygenase with a prominent role in DNA demethylation frequently mutated in various human cancers, was identified among novel CARM1 substrates. To assure its correct identification, we probed FLAG-tagged TET2 with ADMA antibody in wild type and CARM1 KO HEK293T cells (Fig. 2c). The western blot revealed complete disappearance of ADMA modification on TET2 in the absence of CARM1, confirming TET2 as a bona fide CARM1 substrate.

To validate the putative CARM1 substrates in a high-throughput fashion, we designed a peptide array containing nearly 200 peptide sequences of 15 amino acids in length (Supplementary Data 3); three-quarters of the included peptides were centred on substrate arginines. Designed by the blue bracket are substrates of CARM1 identified by IP-MS; the red bracket indicates substrates predicted based on the extracted motifs. Positive control (BAF155 peptide) is denoted with a blue circle; negative control (no peptide) is indicated with a black circle. 

Figure 2 | Discovery of proline-rich CARM1 methylation motifs and high-throughput validation of the detected substrates using peptide arrays. (a) Amino-acid abundance analysis of CARM1 methylation sites. Frequencies of amino acids in the vicinity of CARM1-regulated methylation sites (± 5 nucleotides) were compared to those in the human proteome, revealing pronounced enrichment of proline residues. Glycine, methionine and phenylalanine residues were detected at frequencies typical of the human proteome (P value > 0.01; two-tailed Student’s t-test). (b) Sequence logos of CARM1 methylation motifs. Various proline-containing motifs were enriched in the proximity of the detected CARM1 methylation sites. (c) Western blot with ADMA antibodies of FLAG-tagged TET2 in wild type and CARM1 KO HEK293T cells showing the reduction of TET2 dimethylation in CARM1 KO. (d) Autoradiography of 192-spot peptide arrays following the in vitro methylation assay using 3H-labelled SAM and indicated purified mammalian PRMTs. The peptide arrays consisted of ~15 amino-acid sequences centred on substrate arginines. Designated by the blue bracket are substrates of CARM1 identified by IP-MS; the red bracket indicates substrates predicted based on the extracted motifs. Positive control (BAF155 peptide) is denoted with a blue circle; negative control (no peptide) is indicated with a black circle. (e) Autoradiography of 96-spot peptide arrays following the in vitro methylation assay using 3H-labelled SAM and indicated purified mammalian PRMTs. The design of peptide array is depicted at the bottom. Light blue colour designates the location of novel, cancer-related substrates of CARM1 identified by IP-MS; blue—additional substrates and their naturally occurring mutations in human cancer; dark-blue—point mutations of the recognition sequence adjacent to the CARM1 methylation site on BAF155; grey—known substrates of PRMT1; purple—known substrates of PRMT5; pink—known substrates of PRMT6; white—negative control (no peptide).
were also methylated, showcasing the extrapolative power of the identified substrate recognition motifs. In contrast to CARM1, PRMT1 and 5 failed to methylate these peptides, although PRMT6 could weakly methylate a few of the tested sequences.

Having corroborated the ability of CARM1 to methylate most of the detected substrates, we designed an additional 96-spot peptide array (Fig. 2e and Supplementary Data 3) with two purposes in mind: (1) to examine the significance of surrounding amino acids on CARM1-directed arginine dimethylation, particularly those mutated in human cancers; and (2) to measure PRMT specificity towards respective substrates. The selected sequences encompassed a mixture herein identified and predicted CARM1 substrates (in light blue) and representative cancer-relevant substrates, accompanied by the sequences harbouring mutations detected in human patients\(^8\) (in blue). We also arbitrarily mutated the recognition sequence surrounding R1064 on BAF155 (ref. 20)—the known CARM1 methylation site—to assess the significance of surrounding amino-acid residues to CARM1 activity (four sequence in dark blue). The array also featured 17 known substrates of PRMT1 (in grey), 10 substrates of PRMT5 (in purple) and 4 substrates of PRMT6 (in pink) as positive controls for these enzymes.

The results of the in vitro methylation assay using purified CARM1 visualized via autoradiography (Fig. 2e, top left array) confirmed its ability to methylate over 90% of the tested putative substrates. The comparison between the wild-type sequences and their mutated counterparts revealed pronounced reduction to complete abrogation of CARM1’s ability to methylate the substrates. For example, the naturally occurring mutations on TET2 (R680) G678D, ARID1A (R429) Y430C and P431Q, and MED12 (R1862) P1864Q significantly reduced methylation by CARM1, underscoring their probable functional significance in tumorigenesis. Similarly, mutagenesis of the BAF155 (R1064) surrounding sequence demonstrated that mutations of the proximal residues (P1063A, G1062A and P1066A) nearly abolished methylation by CARM1. Together, these findings clearly demonstrate that the residues nearby the methylated arginine play a critical role in substrate recognition/methylation.

Assays using purified PRMT1, 5 and 6 (Fig. 2e, as indicated) revealed that a few substrates could be methylated by both CARM1 and other PRMTs. Specifically, ~8, 5 and 19% of the tested CARM1 substrates were also methylated by PRMT1, PRMT5 or PRMT6, respectively. Concomitant control experiments with respective known substrates of these PRMTs validated their normal enzymatic activity; PRMT1 methylated 9 out of 17 tested substrates, PRMT5–2 out 10, and PRMT6–3 out 4. Overall, these data demonstrate that although some overlap in substrate preference occurred among PRMTs, CARM1 was prevalently selective towards the identified peptide sequences designated as CARM1 substrates by the LC-MS/MS experiments.

The N-terminal domain of CARM1 in substrate recognition. The structure of the N-terminal domain of CARM1 (residues 28–140) resembles those of the EVH1 domains, a subfamily of PH structure of the N-terminal domain of CARM1 (residues 28–140). The N-terminus of CARM1 in substrate recognition. To test this hypothesis, we generated three CARM1 constructs: one full-length (CARM1 FL) and two N-terminal truncations CARM1 28–608 and CARM1 140–608 (Fig. 3a). CARM1 140–608 eliminates the entire EVH1 domain, while CARM 28–608 was included as a control, since the first 28 residues at the N-terminus of CARM1 do not participate in the formation of the EVH1 domain\(^26\). We transiently transfected the three FLAG-tagged CARM1 fusion constructs in HEK293T CARM1 KO cells. CARM1 was immunoprecipitated from all three cell lines using anti-FLAG M2 resin, and four known CARM1 substrates/interactors (that is, BAF155 (ref. 20), MED12 (ref. 30), PABP1 (ref. 36) and NCOA3 (ref. 19)) and newly identified TET2 were detected by western blotting. Figure 3b displays that all five proteins were co-precipitated with CARM1 FL and CARM1 28–608; however, substantially lower amounts of them were detected with CARM1 140–608. To globally assess the magnitude of reduction in CARM1–substrate interactions induced by the loss of the EVH1 domain, we visualized overall protein changes in CARM1 immunoprecipitates by western blotting using the ADMA antibody (Fig. 3c). The experiment revealed global decrease in abundance of CARM1-interacting proteins with CARM1 140–608, as compared with the CARM1 FL and CARM1 28–608.

These results support our hypothesis that CARM1 may employ its N-terminus for substrate recognition. To further test this, we performed fluorescence polarization (FP) assays using purified, recombinant His-tagged CARM1 EVH1 domain and fluorescein-labelled BAF155 peptide harboring the CARM1 methylation site (Fig. 3d,e). The assay detected dose-dependent increase in polarization signals as concentration of the CARM1 EVH1 domain increased, whereas the control experiment using bovine serum albumin (BSA) generated no change in signal. A similar result was obtained when the His-tag was replaced with glutathione S-transferase (GST), excluding possible effects of the fusion tag binding (Supplementary Fig. 3). Together, these findings strongly indicate that the N-terminal EVH1 domain of CARM1 can directly interact with CARM1 substrates so that its deletion broadly impacts substrate recognition by the enzyme.

The N-terminal domain of CARM1 in substrate methylation. Having established that the N-terminal domain of CARM1 mediates substrate recognition, we next examined its prerequisite for substrate methylation. First, we expressed and purified recombinant CARM1 proteins, including CARM1 FL, CARM1 28–608 and CARM1 140–608, for in vitro methylation assays. Autoradiography results showed that CARM1 FL and CARM1 28–608, but not CARM1 140–608, methylated recombinant full-length BAF155 protein (Fig. 4a). Similar observations were made when GST fused MED12, BAF155 and PABP1 peptides encompassing CARM1 methylation sites were used as substrates in in vitro methylation assays (Supplementary Fig. 4a,b,c). The greatly reduced ability of CARM1 140–608 to methylate protein substrates, as compared with CARM1 FL and CARM1 28–608, were not restricted to a few substrates. When total cell lysates derived from MCF7 CARM1 KO cells were in vitro methylated, we observed substantial reduction of total protein methylation in assays employing CARM1 140–608 (Fig. 4b). These results demonstrate that the N-terminal EVH1 domain of CARM1 is required for substrate methylation in vitro.

Next, we transiently transfected FL and the two aforementioned CARM1 truncations in HEK293T CARM1 KO cells and measured substrate methylation by western blotting with ADMA antibodies. Not only methylation of endogenous BAF155 and MED12 was restored in the presence of CARM1 FL and CARM1 28–608 (Supplementary Fig. 4d), but the total levels of ADMA modification were recovered to the level of parental cells (Supplementary Fig. 4e). In contrast, transient expression of CARM1 140–608 failed to restore endogenous protein methylation in HEK293T CARM1 KO cells. These data suggest that the N-terminal EVH1 domain is indispensable for CARM1 substrate recognition and methylation both in vitro and in cells transiently expressing CARM1.
We next sought to assess the impact of CARM1 N-terminal deletion in cells stably expressing mutant enzymes. We stably expressed GFP (control) and the previously described CARM1 constructs in MCF7 CARM1 KO cells and observed impaired in vivo methylation of BAF155 in CARM1 140–608 expressing, MCF7 CARM1 KO cells (Fig. 4c). Western blots of total cell lysates using ADMA antibody similarly revealed that GFP and CARM1 140–608 expressing cells exhibited drastically decreased levels of cellular protein methylation, as compared with CARM1 FL and CARM1 28–608 expressing MCF7 cells (Fig. 4d). To quantify the affected ADMA-containing peptides in MCF7 CARM1 KO and CARM1 140–608 expressing cells, we employed our quantitative MS method (Fig. 1a) using CARM1 28–608 as a reference for normalization (Supplementary Data 4). This experiment revealed a massive reduction in abundances of ADMA-containing peptides in the CARM1 140–608 expressing and complete CARM1 KO cells (Fig. 4e). The magnitude of reduction in CARM1 140–608 cells, however, was slightly less than that in the CARM1 KO cells. Figure 4f directly compares log2 transformed intensities of ADMA-containing peptides that met the criteria for CARM1 substrates in both cell types, that is, exhibited greater than twofold reduction with maximum P value of 0.01. The line of best fit suggests the overall good agreement between measured peptide intensities ($R^2$ of 0.82), but the
Figure 4 | In vitro and in vivo requirement of the N-terminal domain for substrate methylation by CARM1. (a) In vitro methylation assays using recombinant CARM1 proteins and $^3$H-SAM to test methylation of FLAG-BAF155 protein. Coomassie Brilliant Blue staining of purified BSA (control), FL CARM1, the N-terminally truncated CARM1 28–608 and 140–608, and FLAG-BAF155 are shown in the top two panels. Radioactive labelling of BAF155 is visualized using autoradiography after in vitro methylation assays. (b) In vitro methylation assay using recombinant CARM1 proteins, $^3$H-SAM and total cell lysates derived from MCF7 CARM1 KO cells. Coomassie Brilliant Blue staining (left panel) and autoradiograph (right panel) of total cell lysates after in vitro methylation assays are shown. (c) Western blot analyses of in vivo methylated BAF155 in parental MCF7 or MCF7 CARM1 KO cells stably expressing GFP control, CARM1 FL, CARM1 28–608 or CARM1 140–608. Expression of CARM1 in total cell lysates was detected by anti-FLAG antibody (bottom panel) and the levels of me-BAF155 (top panel) and total BAF155 (middle panel) were detected using corresponding antibodies. (d) Western blot analyses of ADMA-containing proteins in the total cell lysates as described in c. While Ponceau S staining (left panel) confirms equal loading, Western blot (right panel) demonstrates the reduced levels of ADMA in GFP control and CARM1 140–608 expressing cells, as compared with those in parental cells and cells expressing CARM1 FL and CARM1 28–608. (e) Heat map displaying hierarchical clustering of log$_2$ transformed ADMA peptide intensities in complete CARM1 KO MCF7 cells and CARM1 KO MCF7 cells expressing CARM1 140–608 cells, using three biological replicas of each. The intensities were normalized to their average respective levels in CARM1 KO, CARM1 28–608 expressing MCF7 cells. The ADMA peptides were prepared and quantified via nanoLC-MS/MS analysis, as described in Fig. 1a. (f) Comparison of the abundance of ADMA-containing peptides in CARM1 KO MCF7 cells and CARM1 KO MCF7 cells expressing CARM1 140–608. A close Pearson correlation in the levels of ADMA-containing peptides was detected in two cell lines ($R^2$ of 0.82).
reduced slope of the line indicated that the abundance of methylated peptides in cells expressing CARM1 140–608 was slightly, but systematically, higher than in cells completely lacking the enzyme. As indicated by $R^2$ of 0.65, noticeable differences existed between the levels of ADMA-containing peptides in cells expressing the N-terminal truncation of CARM1 and CARM1 28–608 control (Supplementary Fig. 4f). These data verify that the N-terminal EVH1 domain of CARM1 is necessary for recognition and methylation of its many endogenous substrates, even though a small fraction of them could bypass this requirement.

Surprisingly, the abundance of a few ADMA-containing peptides (yellow cluster, Fig. 4e) increased in CARM1 KO and CARM1 140–608 cells, as compared with the reference CARM1 28–608 cells. This subpopulation of sites experienced increased levels of ADMA methylation when CARM1 was absent. We surmise that these ADMA-containing peptides are not bona fide CARM1 substrates but are substrates of other Type I PRMTs that are affected by CARM1. As our MS data and western blot results (Supplementary Fig. 1d,e) previously suggested PRMT1 was reduced by ~2.5-fold in CARM1 KO MCF7 cells. Substrate scavenging by other PRMTs on decrease in the levels of PRMT1 was recently reported by Dhar et al.36, possibly accounting for the unexpected spike in abundance of some ADMA-containing peptides. Consistently with this suggestion, we detected several previously described PRMT1 substrates among the sites upregulated on CARM1 deletion and consequent reduction in PRMT1 (refs 1,5,6) (Supplementary Data 5). The change in PRMT1 level induced by CARM1 deletion was not observed in CARM1 140–680 cells (Supplementary Fig. 5a,b) and in MDA-MB-231 cells (Supplementary Fig. 1c), suggesting that CARM1 regulation of PRMT1 may be cell line specific and is a property of the FL enzyme.

**Discussion**

Using quantitative mass spectrometry along with pan-specific ADMA antibodies, we comprehensively profiled substrates of CARM1, the prototype PRMT with the strongest link to oncogenesis5,6,12, in two breast cancer cell lines. Specifically, we discovered in vivo over 300 CARM1-dependent arginine methylation events, corresponding to >130 novel CARM1 protein substrates, many of which have documented cancer-relevant functions. The large overlap in the identified substrates between two cell lines combined with high-throughput in vitro validation engenders a high confidence compendium of CARM1 substrates. Though our study expands on the known methylation sites of CARM1 by over 10-fold, doubtlessly additional substrates remain elusive. Several factors likely impact the efficiency of substrate discovery using the herein employed approach. First, extremely low-abundance substrates may fall below our limit of detection52, and substrate abundances almost certainly fluctuate across cell lines (Fig. 1c). Second, some ADMA sites likely exist within sequences that are impenetrable on trypsin digestion. For example, the known CARM1 methylation site on BAF155, R1064 (ref. 20), is located in a region whose nearby 100 amino acids are devoid of lysine and arginine residues necessary for trypsin cleavage. As the result, BAF155 was not identified by the IP-MS experiments. We envision that future works using other cell lines and possibly different digestion proteases will likely expand on this foundational data set.

Building on our compendium of in vivo methylated substrates, we uncovered new CARM1 substrate recognition motifs (Fig. 2a,b). An earlier study suggested the importance of proline, glycine and methionine residues in the proximity of CARM1 methylation sites (that is, the PGM motif)42. Our findings confirm the enrichment of proline; however, they do not validate the proposed significance of either glycine or methionine residues in CARM1 recognition sequences. Further, several amino acids identified as components of the CARM1 methylation motif by Gayatri et al.37 did not appear enriched or were selected against according to our results. For example, large aromatic amino acids, therein reported as the essential components of the CARM1 substrate recognition motif, were rarely found within our substrate sequences.

A unique combination of three facets distinguishes our work from these prior studies: (1) measurement of in vivo methylated substrates, (2) high-throughput validation of putative substrates in vitro and (3) successful in silico prediction of novel substrates. As such, we are confident that the revisions we offer to the existing knowledge of CARM1 recognition motifs are bona fide and will expedite further expansion of its substrate repertoire. For example, histone H3R17 and H3R26 are known CARM1 modification sites18 whose flanking sequences match the ‘PR’ and ‘RxxxP’ motifs (where x designates any amino acid; Fig. 2b), respectively. On the contrary, H3R15, which does not contain the detected motifs, is a putative CARM1 methylation site that is, however, called into question by in vivo evidence37,48. Still, we acknowledge that consensus recognition motifs likely do not universally apply to all substrates, as our findings indicate that ~6% of the detected CARM1 methylation sites do not contain proline residues in their vicinity (±5 amino acids).

Building on our improved CARM1 methylation motifs, we established the function of the enzyme’s N-terminal domain in substrate recognition. Toffer-Charlier et al.26 had previously detailed structural similarity between the N-terminals of CARM1 and EVH1 domains that recognize and bind proline-rich sequences; however, based on this observation, the authors hypothesized that the domain may participate in protein–protein interactions, enabling the cofactor function of CARM1. On the contrary, our results unambiguously demonstrate that the N-terminal EVH1 domain of CARM1 is necessary for substrate recognition and methylation, aiding the enzymatic, not cofactor, activity of CARM1.

Interestingly, the N-terminal of CARM1 lacks conserved aromatic residues that are typically located at the binding interface of EVH1 domains and engender their ligand specificity. Suitably, the absence of the conserved phenylalanine residues is the key difference between the classical consensus-binding motifs of the EVH1 domains28 and the CARM1 recognition motifs (Fig. 2b). The structural divergence and dissimilarities in recognition motifs clearly distinguish CARM1 from other proteins harbouring EVH1 domains, likely ensuring that CARM1 could outcompete the latter in specifically recognizing and subsequently methylaing its substrates.

In the study by Toffer-Charlier et al.,26 the N-terminal domain of the nearly FL CARM1 (residues 28–507) appeared disordered and, therefore, was dubbed the wobbly PH domain. We speculate that substrate binding may be a prerequisite for formation of the N-terminal PH fold, and the resultant, structurally defined N-terminal domain may assist in capturing substrates and sandwiching them against the catalytic core of the enzyme. A crystal structure of the FL CARM1 bound to a substrate peptide is needed to test this hypothesis. We also note that functions of the N-terminal domains of the entire PRMT family remain uncharacterized24, and in the light of the high sequence diversity of these regions, we speculate that other PRMT family members may similarly employ their N-terminal domains in substrate recognition.

Given the strong oncogenic potential of CARM1, screening for and designing CARM1 inhibitors is an active area of research, although one with the limited success7,23. Thus far, the structure-
activity relationship—based design of its inhibitors has focused on the catalytic domain of CARM1. However, due to the considerable degree of structural conservation among all PRMTs, SAM or simefugin derivatives that bind this region often lack specificity. Our discovery that the N-terminus of CARM1 enables substrate recognition opens new exciting routes in the design of CARM1-specific inhibitors and warrants functional investigation of the N-terminal domains of other PRMT members, as this approach may constitute a universal path towards specific inhibitors of all PRMTs.

Methods

Cell lines and cell growth. HEK293T, MCF7 and MDA-MB-231 cell lines were purchased from ATCC and tested for mycoplasma contamination. MCF7 and MDA-MB-231 have been authenticated by short tandem repeat profiling in the Translational Research Initiatives in Pathology (TRIP) laboratory at the University of Wisconsin – Madison.

HEK293T CARM1 KO, MCF7 CARM1 KO and MDA-MB-231 CARM1 KO cell lines were generated by transiently transfection of plasmids encoding the indicated sequence into the cells of the indicated origin. For all MS experiments, cells were collected via gentle centrifugation, rinsed twice with ice-cold 1 × phosphate-buffered saline (PBS; Sigma, St Louis, MO) and briefly stored at −80 °C before analysis. Cells for all MS experiments were grown in biological triplicates as three individual cell cultures.

Proteomics sample preparation and antibody enrichment. Cell pellets were thawed on ice and resuspended in lysis buffer (8 M urea, 100 mM Tris, 10 mM TCEP, 40 mM 2-mercaptoethanol and protease inhibitor cocktail; Roche, Indianapolis, IN), vigorously vortexed and sonicated in a water bath for 15–20 min. After centrifugation (15,000 r.p.m., 4 °C, 20 min), the cleared lysate was mixed with an equal volume of ice-cold 1× PBS and briefly sonicated. After centrifugation (15,000 r.p.m., 4 °C, 20 min), the cleared lysate was mixed with an equal volume of ice-cold 1× PBS and briefly sonicated. After centrifugation (15,000 r.p.m., 4 °C, 20 min), the cleared lysate was mixed with an equal volume of ice-cold 1× PBS and briefly sonicated. After centrifugation (15,000 r.p.m., 4 °C, 20 min), the cleared lysate was mixed with an equal volume of ice-cold 1× PBS and briefly sonicated.

For each experiment, a total of 10–12 mg of the labelled peptide mixture was used to enrich in tandem for modified peptides using PTMScan Asymmetric Di-Methyl Arginine Motif [adme-R] antibodies (Cell Signaling Technology, Danvers, MA), according to the manufacturer’s instruction29. The enrichment success, that is, the number of modified peptides among all peptides detected in the precipitate, varied from 12 to 26% across all experiments. Both flow through and the eluted modified peptides were desalted and resuspended in 0.2% formic acid for LC-MS/MS analysis.

Unmodified peptides were fractionated into 20 or 32 fractions using high pH reverse phase chromatography, concatenated into 10 or 16 combined fractions, respectively, lyophilized to dryness, and resuspended in 0.2% formic acid for LC-MS/MS analysis.

LC-MS/MS. All capillary columns were prepared in house. A laser puller (Sutter Instruments Co., Novato, CA) was used to generate 75–350 µm inner–outer diameter bare-fused silica capillary columns with electrospray emitter tips. The tip of each column was plugged with ~5 mm of 5 µm, 130 Å pore size, bridged ethylene hydrophobic C18 particles (Waters, Milford, MA). Columns were then packed with 1.7 µm diameter of bridged ethylene hydrophobic particles to a final length of ~20 cm and installed on a Dionex Ultimate 3000 nano HPLC system (Thermo Fisher, Waltham, MA). Peptides were reconstituted in 0.2% formic acid, and quantitative colorimetric peptide assay kit (Pierce, Rockford, IL) was used to determine peptide concentration. Peptides were consequently lyophilized and labelled with TMT (Pierce TMT, Rockford, IL), according to the manufacturer’s instruction; details on TMT channels used in each experiment are listed in Supplementary Table 5. Labelled peptides were then mixed in 1:1 ratio and the resulting mixture was desalted by affinity chromatography. An amount of 10–12 mg of the labelled peptide mixture was used to enrich in tandem for modified peptides using PTMScan Asymmetric Di-Methyl Arginine Motif [adme-R] antibodies (Cell Signaling Technology, Danvers, MA), according to the manufacturer’s instruction. The enrichment success, that is, the number of modified peptides among all peptides detected in the precipitate, varied from 12 to 26% across all experiments. Both flow through and the eluted modified peptides were desalted and resuspended in 0.2% formic acid for LC-MS/MS analysis.

LC-MS/MS analysis. All capillary columns were prepared in house. A laser puller (Sutter Instruments Co., Novato, CA) was used to generate 75–350 µm inner–outer diameter bare-fused silica capillary columns with electrospray emitter tips. The tip of each column was plugged with ~5 mm of 5 µm, 130 Å pore size, bridged ethylene hydrophobic C18 particles (Waters, Milford, MA). Columns were then packed with 1.7 µm diameter of bridged ethylene hydrophobic particles to a final length of ~20 cm and installed on a Dionex Ultimate 3000 nano HPLC system (Thermo Fisher, Waltham, MA), using a stainless steel ultra-high pressure union (IDEX, Oak Harbor, WA). Mobile phase buffer A consisted of water and 0.2% formic acid. Mobile phase B consisted of water, 70% acetonitrile, 0.2% formic acid and 5% DMSEO. Columns were heated to 60 °C inside an in-house made heater. Peptides were loaded onto a column in 0% B and separated at a flow rate of 300–400 nl min⁻¹ over a 90 min gradient. Eluting peptides were analysed on a quadrupole–ion trap–Orbitrap hybrid Fusion or Fusion Lumos mass spectrometer (Thermo Scientific, San Jose, CA). Orbitrap survey scans were performed at a resolution of 60,000 at 100 m/z, with 30 min acquisition time set to 100 ms. The instrument was operated in the top speed mode with 2× cycles and monoisotopic precursor selection turned on.

Tandem MS scans were collected in the Orbitrap at a resolving power of 60,000 at 200 m/z on precursors with 2–8 charge states, using higher-energy collisional dissociation (HCD) fragmentation with normalized collision energy of 35 and dynamic exclusion of 100 s. The MS2 ion count target was set to 5 × 10⁴ and the maximum injection time was 350 ms and 150 ms during analyses of modified and unmodified peptides, respectively.

Data search and bioinformatics analyses. Generated spectra were searched against the reviewed Uniprot database of human protein isoforms (downloaded 1.19.2015) and processed using the COMPASS software suite69. Carbamidomethylation (+57.0513 Da) of cysteine residues and TMT 10plex (+229.1629 Da) on N-termini of proteins and lysine residues were included as fixed modifications. Oxidation of methionine (+15.999 Da), TMT 10plex on proline (+229.1629 Da) and dimethylation of arginine (+28.0313 Da) were included as variable modifications. Average mass tolerances of 125 ppm and 0.015 Da were allowed for MS1 precursor searches and MS2 fragment searches, respectively. Up to three missed cleavages on tryptic peptides with proline rule were allowed. One per cent false discovery rate (FDR) correction was performed on all identified peptides and proteins. TMT and ADMA-containing peptides, separate 1% FDR was performed on search results that corresponded to target and decoy peptides containing the modification. TMT reporter region quantification was performed using an in-house software TagQuant, as previously described20. Briefly, the raw reporter ion intensity in each TMT channel was corrected for isotope impurities, as specified by the manufacturer for the used product lot, and normalized for mixing differences by equalizing the total signal in each channel. In cases where no signal was detected in a channel, the missing value was assigned with the noise level of the original spectrum (that is, noise-band capturing of missing channels), and the resultant intensity was not corrected for impurities or normalized for uneven mixing.

All spectra corresponding to ADMA-containing peptides were computationally annotated using in-house software Annotated Spectrum Generator. Briefly, each peptide was fragmented in silico into its b- and y-type product ions. Using data mass and an allowed +10 ppm tolerance, each fragment was searched for in the associated MS² spectrum, and matched peaks were annotated with product ion type and number. All m/z peaks having an intensity <1% of the MS² base peak were eliminated from consideration. All charge states less than the parent precursor charge state were considered when matching fragments to m/z peaks. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%. STRING database was searched for interactions among identified sites at the significance of 5%.
Purification of 6xHis-tagged proteins from bacteria

Desired concentration. Protein concentration was then determined using Bradford day, purified protein was retrieved from the dialysis tubing and concentrated to 150 mM NaCl, 0.005% NP-40, 0.5 mM EDTA) at 4°C. The crude lysate was cleared by centrifugation (15,000 r.p.m., 4°C, 20 min). The cleared lysate was then mixed with pre-washed ANTI-FLAG M2 Affinity Gel (60 μl resin for a 15 cm dish) and incubated with rotation at room temperature for 2 h. The resin was then washed 5–5 times with lysis buffer and then the recombinant protein was eluted using lysis buffer supplemented with 150 μg/ml 3×FLAG peptide with rotation at room temperature for 1 h. Protein concentration was determined using Bradford assay and protein purity was assessed by SDS-PAGE and Coomassie Blue staining.

Purification of Flag-tagged proteins. CMX-Flag-BAF155 construct was generated by subcloning a PCR-amplified Flag-BAF155 fragment into the CMX-expression vector. DNA construct was transiently transfected into HEK293T CARM1 KO cells in 15 cm dishes using transIT-LT1 reagent (Mirus Bio, Madison, WI) according to the manufacturer’s protocol. The protein was allowed to express for 48 h. Cell pellets were then collected and lysed in lysis buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 10% glycerol, 0.05% Triton X-100, supplemented with 1× protease inhibitor cocktail before use) with 30 min rotation at 4°C. The crude lysate was cleared by centrifugation (15,000 r.p.m., 4°C, 20 min). The cleared lysate was then mixed with pre-washed ANTI-FLAG M2 Affinity Gel (60 μl resin for a 15 cm dish) and incubated with rotation at room temperature for 2 h. The resin was then washed 3–5 times with lysis buffer and then the recombinant protein was eluted using lysis buffer supplemented with 150 μg/ml 3×FLAG peptide with rotation at room temperature for 1 h. Protein concentration was determined using Bradford assay and protein purity was assessed by SDS-PAGE and Coomassie Blue staining.

Purification of GST-tagged proteins from bacteria. The DNA fragments were cloned into pGEX-2T GST expression vector (GE Healthcare, Waukesha, WI). DNA constructs were transformed into BL21 (DE3) E. coli expression strain, Luria–Bertani media (1–2L) was seeded with 10 ml of overnight starter culture and grown at 37°C until OD60 mm reached 0.6. Protein expression was then induced with the addition of 0.2 mM isopropyl β-D-thiogalactoside for 4 h at ambient temperature. Bacterial pellets were then lysed by sonication on ice in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.05% NP-40, supplemented with 1× protease inhibitor cocktail before use) with 70% amplitude, 30 s, 30 s, 10 cycles. The crude lysate was cleared by centrifugation at 13,000 r.p.m., 4°C for 30 min. The cleared lysate was then mixed with pre-washed glutathione resin (Thermo Scientific, Rockford, IL) and incubated with rotation at ambient temperature for 2 h. The resin was then washed three times with lysis buffer and once with 100 mM Tris (pH 8.0). The recombinant protein was then eluted with elution buffer (100 mM Tris, pH 8.0, 15 mg/ml 1-glutathione) with rotation at ambient temperature for 1 h. Eluted protein was dialysed in snake skin dialysis tubing (Thermo Scientific, Rockford, IL) in dialysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 0.005% NP-40, 0.5 mM EDTA) at 4°C for overnight. The following day, purified protein was retrieved from the dialysis tubing and concentrated to desired concentration. Protein concentration was then determined using Bradford assay and protein purity was assessed by SDS-PAGE and Coomassie Blue staining.

Purification of 6xHis-tagged proteins from bacteria. The DNA fragment encoding CARM1 residues 28–140 was cloned into pET21a expression vector (Novagen, Madison, WI). DNA construct was transformed into BL21 (DE3) E. coli expression strain, Luria–Bertani media (1–2L) was seeded with 10 ml of overnight starter culture and grown at 37°C until OD60 mm reached 0.6. Protein expression was then induced with the addition of 0.2 mM isopropyl β-D-thiogalactoside for 4 h at ambient temperature. Bacterial pellets were then lysed by sonication on ice in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.005% NP-40, 0.5 mM EDTA) at 4°C for overnight. The following day, purified protein was retrieved from the dialysis tubing and concentrated to desired concentration. Protein concentration was then determined using Bradford assay and protein purity was assessed by SDS-PAGE and Coomassie Blue staining.

Western blotting. Cells were collected by trypsinization, washed with Dulbecco’s phosphate buffer saline (Life Technologies, Carlsbad, CA) and lysed by suspension in lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% glycerol, 0.5% Triton X-100 and 1× protease inhibitor cocktail (Sigma-Aldrich)). After brief sonication before use) with 70% amplitude, 30 s, 30 s, 10 cycles. The crude lysate was cleared by centrifugation at 13,000 r.p.m., 4°C for 30 min. The cleared lysate was then mixed with pre-washed TALON metal affinity resins (Clontech, Mountain View, CA) and incubated with rotation at ambient temperature for 2 h. The resin was then washed three times with lysis buffer and once with 100 mM Tris (pH 8.0). The recombinant protein was then eluted with elution buffer (100 mM Tris, pH 8.0, 15 mg/ml 1-glutathione) with rotation at ambient temperature for 1 h. Eluted protein was dialysed in snake skin dialysis tubing (Thermo Scientific, Rockford, IL) in dialysis buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 0.005% NP-40, 0.5 mM EDTA) at 4°C for overnight. The following day, purified protein was retrieved from the dialysis tubing and concentrated to desired concentration. Protein concentration was then determined using Bradford assay and protein purity was assessed by SDS-PAGE and Coomassie Blue staining.

In vitro methyltransferase assays. Purified recombinant enzyme and substrates were incubated in methylation buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 0.5 mM EDTA) for 1–4 h at 30°C in the presence of 1 μM radioactive S-adenosyl-L-[methyl-3H] methionine. The reaction was stopped by adding 5 × SDS sample buffer and boiling at 99°C for 5 min. Proteins were resolved by SDS-PAGE. The gel was stained with Coomassie Brilliant Blue (Coomassie Brilliant Blue R-250 0.05% w/v: methanol 50% acetic acid 10%: H2O 40%) followed by de-staining with dye-free buffer. Fixed gel was then soaked with amplify (GE Healthcare, Waukesha, WI) with gentle shaking at room temperature for 30 min before drying on a Whatman paper at 80°C and exposing to an X-ray film in dark at –80°C. Protein methylation was detected by autoradiography.

For methylation assay using the peptide array, 16 or 20 μg purified recombinant CARM1 or same molar concentration of the other PRMTs (for 96-spot and 192-spot arrays, respectively) was mixed well with 40 μl radioactive S-adenosyl-L-[methyl-3H] methionine, 80 μl 5 μl methylation buffer and ddH2O to achieve a final volume of 200 or 400 μl reaction for (96-spot and 192-spot arrays, respectively). The Celluspot peptide arrays were customized and produced by Intavis Bioanalytical Instruments, Germany. Before the methylation assay, the peptide array on the slide was circled using a hydrophobic PAP pen. The mixed methylation reaction was then applied onto the array slide to cover the peptide array area, and then the slide was placed in a humidified chamber and incubated at 30°C for 4 h. The reaction was then gently flicked off and the slide was rinsed with 1× methylation buffer. The array slide was washed in methylation buffer for three times and then soaked in amplification solution for 30 min with rotation at room temperature. The slide was then washed again with methylation buffer for three times and air dried before exposure to X-ray film at –80°C for 3 days. A well that contained no peptide sequence was used as the negative control and the quantification reference.

Virus packaging and stable cell line generation. Virus packaging was performed by transfection of three plasmids (PHT160, pLNOX-PRMT expressing plasmid, VSVG) into HEK293T cells. Supernatant containing the virus was collected for cell line infection after 48 h (ref. 20). To generate the stable cell lines, 2 × 10⁶ MCF7 CARM1 KO cells were seeded into six-well plate 24 h before infection. For infection, 1 ml virus was mixed with 1 ml fresh cell culture medium; polybrene was added at a final concentration of 8 μg/ml to increase infection efficiency. Cells were infected overnight before cell culture medium was changed. Cells were selected with 400 μg/ml G418 for at least 4 weeks to obtain the stable cell lines.

Data availability. All raw mass spectrometric data files and annotated spectra supporting the findings of this study are available through Chorus (ID 1174) and the ProteomeXchange Consortium via the PRIDE partner repository (accession #PXD003871). Original, uncropped images of blots and gels are provided as Supplementary Fig. 6. All other data are available within the paper or included as its Supplementary Materials, and available from the corresponding author on request.

References
1. Bedford, M. T. & Clarke, S. G. Protein arginine methylation in mammals: who, what, and why. Mol. Cell 33, 1–13 (2009).

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms15571 | www.nature.com/naturecommunications

NATURE COMMUNICATIONS | 8:15571 | DOI: 10.1038/ncomms15571 | www.nature.com/naturecommunications

11
2. Larsen, S. C. et al. Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. Sci. Signal. 9, 1–5 (2016).

3. Evian, M., Stelzl, F., Georg, Z. M. & Schuler, W. M. Effect of methylation on the side chain pKa value of arginine. Protein Sci. 6, 1–8 (2015).

4. Gary, J. D. & Clarke, S. RNA and protein interactions modulated by protein arginine methylation. Proc. Natl Acad. Sci. USA 100, 6464–6468 (2003).

5. Yang, Y. & Bedford, M. T. Protein arginine methyltransferases and cancer. Nat. Rev. Cancer 13, 37–50 (2013).

6. Pouillard, C., Corbo, L. & Le Romancer, M. Protein arginine methylation/demethylation and cancer. Oncotarget 7, 67532–67550 (2016).

7. Hu, H., Qian, K., Ho, M. C. & Zheng, Y. G. Small molecule inhibitors of protein arginine methyltransferases. Expert Opin. Investig. Drugs 25, 335–358 (2016).

8. Yadav, N. et al. Specific protein methylation defects and gene expression perturbations in coactivator-associated arginine methyltransferase 1-deficient mice. Proc. Natl Acad. Sci. USA 100, 1877–881 (2003).

9. Kim, Y.-R. et al. Differential CARM1 expression in prostate and colorectal cancers. BMC Cancer 10, 197 (2010).

10. Elakoum, R. et al. CARM1 and PRMT1 are dysregulated in lung cancer without hierarchical features. Biochem. 97, 210–218 (2014).

11. Morettn, A., Mitchell Baldwin, R. & Cote, J. Arginine methyltransferases as novel therapeutic targets for breast cancer. Mutagenesis 30, 177–189 (2015).

12. Ou, C.-Y. et al. The reactome pathway knowledgebase. Nucleic Acids Res. 42, 481–487 (2014).

13. Szklarczyk, D. et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nat. Biotechnol. 33, 1391–1398 (2015).

14. Uhmann, T. et al. A method for large-scale identification of protein arginine methyltransferases. Mol. Cell Proteomics 11, 1489–1499 (2012).

15. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 30, 733–750 (2016).

16. Dhar, S. et al. Loss of the major Type I arginine methyltransferase PRMT1 causes substrate scavenging by other PRMTs. Sci. Rep. 3, 1311 (2013).

17. Guccione, E. et al. Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. Nature 449, 933–937 (2007).

18. Hyllus, D. et al. PRMT6-mediated methylation of H2 in histone H3 antagonizes H3 K4 trimethylation. Genes Dev. 21, 3369–3380 (2007).

19. Weng, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. COMPASS: a suite of pre- and post-search proteomics software tools for OMSA. Proteomics 11, 1064–1071 (2011).

20. Lee, M. V. et al. A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. Mol. Syst. Biol. 7, 514 (2011).

21. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proc. Natl Acad. Sci. USA 100, 9440–9445 (2003).

22. Tyanova, S. et al. The perseus computational platform for comprehensive analysis of (pro)teomics data. Nat. Methods 13, 731–740 (2016).

23. Chen, E. Y. et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128 (2013).

24. Morgan, A. A. & Rubenstein, E. Proline: the distribution, frequency, and statistical significance of mass spectrometry experiments, in vitro methylation assays and generated mutant cell lines. F.L. assisted with construct design, stable cell line generation and performed in vitro methylation assays. N.W.K developed a computation approach to spectrum annotation and assisted with bioinformatics analyses. A.S.H. assisted with design, optimization and interpretation of mass spectrometry experiments. J.C. and W.X. directed and supervised all aspects of the study. All authors contributed to writing and critical review of the manuscript.

Author contributions
E.S. designed and conducted mass spectrometry experiments, interpreted results and performed bioinformatics analyses. H.Z. performed and interpreted biochemical experiments, in vitro methylation assays and generated mutant cell lines. F.L. assisted with construct design, stable cell line generation and performed in vitro methylation assays. N.W.K developed a computation approach to spectrum annotation and assisted with bioinformatics analyses. A.S.H. assisted with design, optimization and interpretation of mass spectrometry experiments. J.C. and W.X. directed and supervised all aspects of the study. All authors contributed to writing and critical review of the manuscript.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Acknowledgements
We thank Dr Jonathan Florentin for assistance with peptide array design and Dr Lu Wang for involvement in the project. We gratefully acknowledge support from the National Institutes of Health Grant P41 GM108538 (awarded to J.C.I.J.) and NIH R21CA196635 and NIH R01 CA213293 (awarded to W.X.). This work was also supported in part by NIH/NCI P30CA14520-1UW Cancer Center Support Grant and U54DK14031.
Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Shishkova, E. et al. Global mapping of CARM1 substrates defines enzyme specificity and substrate recognition. Nat. Commun. 8, 15571 doi: 10.1038/ncomms15571 (2017).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.