Cell. Mol. Life Sci. (2014) 71:4471–4486
DOI 10.1007/s00018-014-1651-y

Cellular and Molecular Life Sciences

RESEARCH ARTICLE

Deep sequencing reveals novel Set7 networks

Samuel T. Keating · Mark Ziemann · Jun Okabe · Abdul Waheed Khan · Aneta Balcerczyk · Assam El-Osta

Received: 8 November 2013 / Revised: 13 May 2014 / Accepted: 15 May 2014 / Published online: 30 May 2014

Abstract
Background Methyl-dependent regulation of transcription has expanded from a traditional focus on histones to encompass transcription factor modulation. While the Set7 lysine methyltransferase is associated with pro-inflammatory gene expression in vascular endothelial cells, genome-wide regulatory roles remain to be investigated. From initial characterization of Set7 as specific for methyl-lysine 4 of H3 histones (H3K4m1), biochemical activity toward non-histone substrates has revealed additional mechanisms of gene regulation.
Results mRNA-Seq revealed transcriptional deregulation of over 8,000 genes in an endothelial model of Set7 knockdown. Gene ontology identified up-regulated pathways involved in developmental processes and extracellular matrix remodeling, whereas pathways regulating the inflammatory response as well as nitric oxide signaling were down-regulated. Chromatin maps derived from ChIP-Seq profiling of H3K4m1 identified several hundred loci with loss of H3K4m1 at gene regulatory elements associated with an unexpectedly subtle effect on gene expression. Transcription factor network analysis implicated six previously described Set7 substrates in mRNA-Seq changes, and we predict that Set7 post-translationally regulates other transcription factors associated with vascular endothelial gene expression through the presence of Set7 amino acid methylation motifs.
Conclusion We describe a role for Set7 in regulating developmental pathways and response to stimuli (inflammation/immune response) in human endothelial cells of vascular origin. Set7-dependent gene expression changes that occurred independent of H3K4m1 may involve transcription factor lysine methylation events. The method of mapping measured transcriptional changes to transcription factors to identify putative substrates with strong associations to functional changes is applicable to substrate prediction for other broad-substrate histone modifiers.

Keywords Lysine methylation · Set7 · ChIP-Seq · Transcriptome · Substrate prediction · STAT1

Electronic supplementary material The online version of this article (doi:10.1007/s00018-014-1651-y) contains supplementary material, which is available to authorized users.

S. T. Keating · M. Ziemann · J. Okabe · A. W. Khan · A. Balcerczyk · A. El-Osta (✉)
Epigenetics in Human Health and Disease Laboratory, Baker IDI Heart and Diabetes Institute, The Alfred Medical Research and Education Precinct, Melbourne, VIC 3004, Australia
e-mail: assam.el-osta@bakeridi.edu.au

M. Ziemann · A. El-Osta
Epigenomics Profiling Facility, Baker IDI Heart and Diabetes Institute, The Alfred Medical Research and Education Precinct, Melbourne, VIC 3004, Australia

J. Okabe · A. El-Osta
Department of Medicine, Central Clinical School, Monash University, Melbourne, VIC 3800, Australia

A. El-Osta
Department of Pathology, The University of Melbourne, Melbourne, VIC 3010, Australia

Introduction

Site-specific chemical modification of lysine residues has emerged as a common regulatory mechanism of protein activity and gene expression. While the gene regulatory effects of post-translational modifications are historically associated with the N-terminal tails of chromatinized histones, the importance of lysine methyltransferase (KMT)

and lysine acetyltransferase (KAT) activity toward transcription factors is increasingly appreciated. As further insights are gained into methylation-dependent pathways, discovery-based approaches that accelerate substrate identification are required to complement conventional methods of characterization. For instance, in silico studies indicate proteome-wide canonical motifs at acetylated residues [1, 2]. Similar coupling of motif-based prediction methods with peptide arrays and conventional biochemical approaches has proven useful for identification of novel targets for various KMTs [3, 4]. Although these methods have provided a clear foundation for in vivo validation of several candidate substrates, unbiased reliance on the presence of commonly ubiquitous consensus motifs as potential sites of modification is perhaps limited for discovery of biologically relevant substrates. The increased use of high-throughput sequencing technologies that facilitate genome and transcriptome-wide investigations has driven rapid accumulation of large repositories of publicly accessible data. Leverage of these data sets permits bioinformatic analysis of transcription factor enrichment relevant to gene regulation associated with enzymes of interest [5], remarkably narrowing the focus for discovery to substrates implicated in functional changes.

The broad-substrate specificity of protein modifiers confounds the discrimination of transcriptional changes associated with histones from those generated by post-translational modification of transcription factors. The Set7 KMT (also described as Set9 or Set7/9) mono-methylates H3 histones at lysine 4 (H3K4m1) [6]. The necessity of this enzymatic function for transcriptional regulation of glioblastomic *MMP1* expression [7] was later extended to hyperglycemic vascular endothelial cells at the promoters of *RELA* [8, 9] and *IL8* [10]. Contrasting the largely invariant chromatin landscape of promoters observed across diverse human cell lines [11], distal enhancer elements exhibit remarkable cell lineage-specific patterns of histone tail modification in accordance with reports of cell-specific H3K4m1 distribution [12, 13]. As a regulator of H3K4m1 patterns, a potential role for Set7 exists in enhancer-associated tissue-specific transcription. While precise effects of H3K4m1 on transcription remain unclear, Set7 could maintain a state of transcriptionally competent chromatin through distinct mechanisms of epigenetic cross talk with the repressive NuRD histone deacetylase complex [6], Suv39h1-mediated H3K9 methylation [14, 15], as well as histone acetylation [14]. Exemplified by the Set7-Suv39h1 regulatory interaction, the increasing focus of studies toward Set7-mediated methylation of non-histone peptide sequences [3] highlights the prominent regulatory role of this modification in the stability and activity of many transcription factors including p53 [16–18], NFκB-p65 [19, 20] and FOXO3 [21] (for recent review, see [22]). Consistent with these observations, chromatin immunoprecipitation (ChIP) has identified distinct regulatory roles for Set7 in transcriptional events associated with hyperglycemic stimulation of human vascular endothelial cells by H3K4m1-dependent and H3K4m1-independent mechanisms [10].

Analysis of Set7-mediated transcriptional regulation of vascular endothelial gene expression is anticipated to significantly expand current understanding [8–10]. Comprehension of pathways and key genes regulated by Set7 under basal and disease conditions is critical to the development of targeted therapeutic interventions for endothelial dysfunction associated with vascular complications of diabetes. The current investigation endeavored to predict novel substrates from ontological relationships to deregulated gene expression observed by massive parallel sequencing in a vascular endothelial model of Set7 knockdown. By intersecting genome-wide H3K4m1 enrichment and gene expression data with Gene Set Enrichment Analysis (GSEA) [23] and known substrate interactions, we describe widespread changes to the endothelial transcriptome by mechanisms subtly associated with H3K4 and implicative of previously described and novel transcription factors. With the resolution for prediction narrowed to factors associated with functional mRNA changes, the novelty of our approach is apparent in the prediction of protein interactions from gene expression data. This method complements conventional methods of substrate prediction and will provide a particularly useful resource for studies of enzymes with specific substrate motif profiles. As the scope for discovery is increased by the accumulation of publicly available data and more substrates of the methyl- and acetyl-proteome are discovered, similar challenges facing interpretation of transcriptome profiling of other broad-substrate histone-modifying enzymes simultaneously offer means to identify novel factors that connect to gene expression.

Materials and methods

Set7 shRNA knockdown

Human microvascular endothelial cells (HMEC-1s) were cultured in MCDB 131 medium (Invitrogen). Knockdown (KD) of Set7 expression in HMEC-1s was achieved using MISSION shRNA-expressing lentivirus vectors (Sigma) as described previously [8]. Cells were incubated with the lentivirus for 2 days, followed by selection in puromycin (1 μg/ml; Sigma) for 7 days. Cells transduced with MISSION non-target shRNA control (NTC) vector were used as controls.

Western immunoblot analysis

Whole cell lysates were prepared using MPER (mammalian protein extraction reagent, Pierce) according to the manufacturer's instructions. 1×10^6 cells were resuspended in 100ul of MPER and incubated at 4 °C with agitation for 30 min. Supernatants were recovered by centrifugation at 15,000 x g at 4 °C for 15 min followed by SDS-PAGE and immunoblot analysis using antibodies against Set7 (#2813, Cell Signaling Technology, Inc.), STAT1 (#9172S, Cell Signaling Technology, Inc), BRG1 (gift from Dr. S. Sif, Ohio State University), NRSF (#07-579, Millipore) and GAPDH (#9484, Abcam). For global analysis of H3K4m1, approximately 3×10^6 cells were resuspended in 600 µl acid extraction buffer (10 mM KCl, 1.5 mM $MgCl_2$, 10 mM Hepes (pH 7.9), Protease Inhibitor Cocktail Mix (Roche)) to pre-swell the cells. The cells were then pelleted by centrifugation at 9,000 x g for 20 s, and the supernatant was discarded. To lyse the cells, pellets were resuspended in 100 µl of acid extraction buffer, to which 6.5 µl of 5 M sulfuric acid was added. The samples were incubated on ice for 1 h with intermittent vortexing every 15 min. Cellular debris was removed by centrifugation at 15,000 x g for 10 min at 4 °C, and the supernatant was retained. The acid-soluble proteins were precipitated with 900 µl of acetone at -20 °C for 1 h and washed with 70 % ethanol. Samples were then centrifuged at 15,000 x g for 1 min at 4 °C, and the supernatant was discarded. The pellets were air-dried, resuspended in 60 µl of sterile, nuclease-free H_2O and incubated on ice for 2 h with intermittent resuspension by careful pipetting. Undissolved pellet was removed by centrifugation at 13000RPM for 3 min. Samples were analyzed by SDS-PAGE and Western immunoblotting using antibodies specific to H3K4m1 (#8895, Abcam) and total H3 histone (#1791, Abcam).

Proliferation and cell viability

The proliferation rate of cells following Set7 knockdown was analyzed by growth assays. Equal numbers of NTC and Set7KD cells were seeded in 60-mm tissue culture plates and harvested every 2 days by trypsin digestion. The reaction was quenched with 1.5 ml of growth media, and the 2 ml cell suspension was transferred to a fresh Eppendorf tube. In total, 20 µl of cell suspension was loaded to a hemocytometer, and the average number of cells in a 1 mm × 1 mm area (counted 4 times individually) was determined. The following equation was used to calculate total cell numbers in the suspension: Average # cells counted x 10,000 × total volume of cell suspension (ml). Cell numbers were determined every 2 days for 8 days to compare proliferation rates between NTC and Set7KD cells. Viability of Set7KD HMEC-1s was determined by (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium; MTS) assay. 5×10^3 cells were seeded into wells of the 96-well plate and cultured for 24 h. The cells were incubated with 20 µl of the MTS reagent (Promega) for 2 h. Colorimetric signal was measured at 490 nm using a microplate reader (Bio-Rad).

Cell cycle analysis by FACS

To analyze the replication of HMEC-1s following Set7 knockdown, fluorescent labeling of the nuclei of cells in suspension followed by fluorescence-activated cell sorting (FACS) was conducted. NTC and Set7KD cells were harvested by trypsin digestion and centrifuged at 12,000 rpm for 5 min at room temperature. The cells were then washed twice with PBS and pelleted by centrifugation at 300 x g for 5 min at room temperature. 2×10^6 cells were resuspended in 1 ml of resuspension buffer (3.8 mM sodium citrate, 0.1 % NP40) and transferred to fresh FACS tubes. 14 µl of RNAse A stock solution (Sigma) was added, and each cell suspension was then incubated for 30 min at 37 °C. 8 µl of propidium iodide staining solution was added to each cell suspension. FACS analysis was conducted on Flow Cytometer/FACS calibur (Beckton Dickinson) according to standard protocols. For each sample, 10,000 events were acquired for analysis. Events corresponding to single cells were gated to exclude debris and cellular aggregates.

Analysis of caspase activity

The Apo-ONE® homogeneous caspase-3/7 assay kit (Promega) was utilized to assess caspase activity according to the manufacturer's protocol. NTC and Set7KD cells were seeded at a density of 2×10^4 per well with 100 µl of growth medium in a 96-well plate (NUNC) and incubated overnight at 37 °C in a 5 % CO_2 humidified incubator. Normal growth medium was used as a blank control. The following day, caspase substrate Z-DEVD-R110 (100x) was diluted 100-fold in Apo-ONE® homogeneous caspase-3/7 assay buffer, and 100 µl of the diluted substrate was added to each well. The contents of the wells were mixed on a plate shaker for 4 h at room temperature. Fluorescence was measured using a fluorescence plate reader at an excitation wavelength of 485 nm and an emission wavelength of 530 nm.

RNA profiling

Total RNA was isolated using the Qiagen RNeasy kit, and RNA integrity was verified using the MultiNA capillary electrophoresis instrument (Shimadzu). Five microgram of total RNA from Set7KD and NTC cells was used

to produce double-stranded cDNA-sequencing libraries according to the Illumina mRNA-Seq library preparation kit. Library quality control was performed with the MultiNA using the DNA500 kit, and these libraries were sequenced on the Illumina Genome Analyzer IIx with flow cell preparation undertaken on Illumina Cluster Station using a DNA concentration of 6 pM. Thirty-six-bp mRNA-Seq reads were aligned with Burrows Wheeler Aligner [24] to the human genome (hg19/GRCh37). Reads aligning to GENCODE v14 exons [52] with a mapping quality (mapQ) $\geq$ 10 were counted using BedTools [25]. Genes with an average of fewer than 10 reads per sample were omitted from downstream analysis. Differential analysis was performed by edgeR v3.0.8 [26] using pairing information for the three independent replicates. A significance threshold of 0.05 after false discovery rate (FDR) correction was employed.

H3K4m1 profiling

Chromatin Immunoprecipitation for H3K4m1 using cross-linked and sonicated chromatin input (X-ChIP) was performed as previously described [8] for Set7KD and NTC samples using H3K4m1 antibody supplied by Abcam (#8845). In addition, H3K4m1 ChIP was performed on S1 chromatin prepared by micrococcal nuclease digestion (MNase-ChIP). ChIP DNA was quantified by fluorometry using the Qubit instrument (Invitrogen). Ten nanogram of ChIP DNA was used in library preparation according to the Illumina ChIP-Seq sample preparation kit protocol. Library DNA was quantified with MultiNA (DNA500 kit) and sequenced as above. Reads were aligned to the human genome as above and filtered for reads with alignment mapQ $\geq$ 10. H3K4m1 peaks were identified with MACS software and peaks called with a p value 1×10^{-7} [27]. Peaks were annotated genes names if they occurred within 5 kbp of a transcriptional start site (TSS) using the GENCODE v14 annotation set. In addition to TSS analysis, we scanned for differential enhancer activity by overlapping H3K4m1 peaks with DNase hypersensitivity regions classified as putative proximal and distal *cis* elements connected to promoters by ENCODE [28]. Independent validation of H3K4m1 changes was conducted on fresh preparations of MNase-ChIP DNA using qRT-PCR with specific primers for *FOSB* promoter region (for-GCGTACTTTGAGGACTCGCT, rev-AAGTAAGGTGCAAGTCCGGG) and a DRE associated with *APLN* (for-CCATTTGGACAGGAGCATCT, rev-CTCTCCACCCTCTCCCTTTC).

Pathways and gene set analysis

Differential gene expression sets were analyzed with MetaCore version 6.8 build 30387 (GeneGo, Inc.). Gene sets of up-regulated and down-regulated transcripts were analyzed individually to identify cellular pathways significantly impacted by knockdown of Set7.

Integration of gene expression and H3K4m1 data sets

Transcripts detected by mRNA-Seq were ranked by p value, and direction of gene expression changes for GSEA according to Subramanian et al, [23]. Genes with altered H3K4m1 at TSSs and distal regulatory elements (DREs) were collected into custom gene sets for GSEA to assess the effect of altered H3K4m1 on gene expression. DREs were defined as DNase hypersensitivity sites that show correlation with promoter DNase hypersensitivity, many cross-validated with the chromosome conformation capture carbon copy (5C) technique [28]. Gene set associations were considered significant at a p value of 0.1 after FDR correction.

Transcription factor enrichment analysis and prediction of novel Set7 substrates

Gene sets of transcription factor targets were mined from ENCODE transcription factor-binding site data (http://genome.ucsc.edu/ENCODE/downloads.html) according to observed occupancy of these transcription factors at genes bodies and promoter regions up to 5 kbp upstream of TSSs. GSEA was used to determine the changes in transcription factor activity caused by Set7 knockdown as above. Peptide sequences of significantly connected transcription factors without prior associations to Set7 were searched using the MOTIF search function of GenomeNet (http://www.genome.jp/) for short amino acid sequences according to the formula: [G/R/H/K/P/S/T]-[K>R]-[S>K/Y/A/R/T/P/N]-K-[Q/N> other amino acids, but not FYWPL]-[A/Q/G/M/S/P/T/Y/N] [3], where K represents the lysine residue targeted for methylation. With K representing the target lysine at position 0, many previously identified Set7-binding sites closely follow this consensus profile of predominantly basic residues at the -2 and -3 positions. The -1 position favors serine, alternatively this position can be occupied by a number of basic and aromatic residues. The substrate-binding pocket of Set7 is flanked by several acidic amino acid residues, indicating a general preference for positively charged substrates most likely governed by the charge of residues that occupy the three sites immediately preceding the target lysine. While the amino acids at positions $+1$ and $+2$ are less important for Set7 binding, several verified substrates favor polar residues at $+1$ (Q/N) and exclude F, Y, W, P, and L as predicted by the optimized formula [22]. Surface accessibility of lysine residues was predicted using NetSurfP-1.1 [29].

Validation of mRNA-Seq

Total RNA from NTC and Set7KD cells was reverse transcribed with the SuperScript III First Strand Synthesis System (Invitrogen). Gene expression was measured by Sybr green (Invitrogen)-based reverse transcription PCR (qRT-PCR) on ABI Prism 7500 (Applied Biosystems). Reactions were incubated at 50 °C for 2 min and 95 °C for 10 min, followed by 50 cycles of 95 °C for 15 s and 60 °C for 1 min using the following gene-specific primers: *RELA* for-ACCGCTGCATCCACAGTTTCC, rev-AGTCCCCACGCTGCTCTTCT; *IL1B* for-ACGGACCCCAAAAGATGAAGGGCT, rev-ACTGCCTGCCTGAAGCTCTTGT; *JUN* for-GTCCTTCTTCTCTTGCGTGG, rev-GGAGACAAGTGGCAGAGTCC; *CCL2* for-AGCAAGTGTCCCAAAGAAGC, rev-TGGAATCCTGAACCCACTTC; *SLC7A11* for-GGCAGTGACCTTTTCTGAGCGGC, rev-ACACACCACCGTTCATGGAGCC; *ATF3* for-TGGGATCAGATGGGAAGATGTGACA, rev-TCATTTTGATTTTGGGGCAAGGTGC, *MX1* for-GATGATCAAAGGGATGTGGC, rev-AGCTCGGCAACAGACTCTTC; *BST2* for-CTTTTGTCCTTGGGCCTTCT, rev-AGAAGGGCTTTCAGGATGTG; *IRF7* for-AGGGTGACAGGTACGGCTCT, rev-CTCCTGGAGAGGGACAAGAA and *ACTB* for-GCCCTGAGGCACTCTTCCA, rev-GCGGATGTCCACGTCACAC or *HPRT1* for-TCCATTCCTATGACTGTAGATTT, rev-AACTTTTATGTCCCCCGTTGACT as the internal control.

In vitro methyltransferase activity

FLAG-tagged Set7 (FLAGSet7) protein was obtained, and the Set7 methyltransferase assay was performed as described previously [10]. The synthetic STAT1 (Y Y S R P K E A P E P M E L) and mutated STAT1 (Y Y S R P R E A P E P M E L) peptides were purchased from Auspep, Australia. In total, 200 pmol of these peptides, or H3K4 peptide or H3R4 peptide (Sigma), were incubated with recombinant FLAG-tagged Set7 (FLAGSet7) protein in a buffer containing 50 mM Tris–HCl pH 8.5, 0.25 mM PMSF, 0.5 mM DTT and [H3]-S-adenosyl-methione (0.5uCi, 373 nmol, PerkinElmer Life Sciences) for 1 h at 30 °C. The reaction mix was placed on P81 phosphocellulose squares (Millipore), and the incorporation of tritiated methyl group to peptide was measured by liquid scintillation counter (Beckman).

Statistics

mRNA-Seq

Identification of differentially expressed genes from RNA profiling data was performed using edgeR software [30]. This software employs an overdispersed Poisson model with negative binomial distribution to account for technical and biological variability. Overdispersion is limited by empirical Bayes shrinkage. A significance threshold of 0.05 after false discovery rate (FDR) correction was utilized.

ChIP-Seq

Differential histone 3 lysine 4 monomethylation changes were identified from ChIP-Seq data using model-based analysis of ChIP-Seq [26]. This software employs a Poisson model at intervals across the genome to identify candidate regions, which are then resolved in detail by summit detection and boundary definition. Peaks called with a p value 1×10^{-7} were considered significant.

Pathway and gene set analysis

Pathway analysis was performed using MetaCore software which implements a hypergeometric model to calculate p values to describe the over-representation of pathways (gene ontology sets) within a list of differentially regulated genes. To determine the effect of differential histone methylation on gene expression, GSEA was used [23]. After generating lists of genes with differential histone methylation, we used GSEA software to determine the enrichment of these differentially methylated genes at the extremes of the p value ranked gene expression data (most activated and most repressed genes). GSEA estimates significance of differential enrichment based upon empirical permutation test by randomisation of gene names in the rank and performs FDR correction.

qRT-PCR

Significance of differential gene expression measured by qRT-PCR was determined by one-sided Student's t test. The degree of histone modification measured by qRT-PCR was expressed as percent input corrected, and significance of differential modification was determined by one-sided Student's t test.

Results

Cellular phenotype of Set7 knockdown endothelial cells

To study the importance of Set7 in endothelial cell regulation, we performed shRNA knockdown in HMEC-1s (Set7KD). Western blot analysis of whole cells lysates prepared from NTC and Set7KD cells probed with antibodies against Set7 confirmed reduced protein expression (Fig. 1a). No significant difference in gross morphology was observed following knockdown of Set7

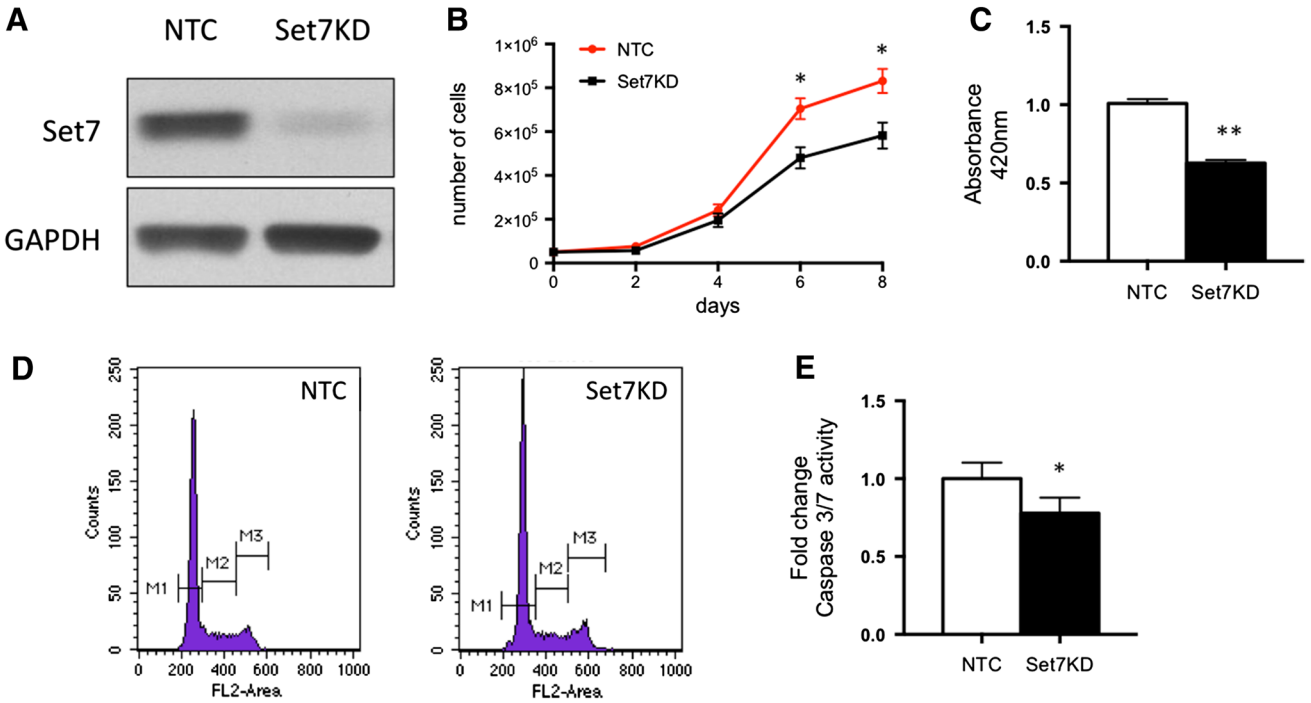

Fig. 1 shRNA knockdown of Set7 confers changes to growth and survival of HMEC-1s. **a** Reduced expression of Set7 protein following shRNA transfection was confirmed by Western blot analysis of whole cell lysates prepared from NTC and Set7KD cells using specific antibodies. Specific detection of GAPDH confirmed equal loading. These blots are representative of three independent experiments. **b** Growth curve analysis revealed decreased proliferation of Set7KD cells. **c** Cell viability measured by MTS assay was significantly decreased for Set7KD cells. **d** Cell cycle analysis by FACS revealed that the proportion of cells at the G1/S boundary (M1) of NTC cells (*left panel*) was greatly increased in Set7KD cells (*right panel*). **e** Significant reduction of caspase 3/7 activity was observed for Set7KD cells. $*p < 0.05$ $**p < 0.01$

(data not shown). Phenotypic characterization of Set7KD cells revealed large decreases in the rate of proliferative growth (Fig. 1b) compared with NTC control cells. MTS assay revealed a strong reduction in the viability of Set7KD cells (Fig. 1c). Cell cycle analysis by FACS indicated that a greater proportion of Set7KD cells were arrested at the G1/S boundary relative to NTC cells (Fig. 1d). To investigate the possibility that changes in apoptotic signaling could alter the proliferation of Set7KD cells, we measured markers of apoptosis and found significantly decreased caspase activation following Set7 knockdown (Fig. 1e).

Transcriptome profiling

In order to study the importance of Set7 in endothelial gene expression, we profiled the transcriptome of Set7KD cells with mRNA-Seq in triplicate in comparison with the NTC vector. We obtained approximately 23 million reads per sample, of which approximately 90 % of reads aligned to the human genome Hg19 and approximately 71 % of all reads aligning to GENCODE exons with a mapQ $\geq$ 10 (Online Resource and Table S1). We performed differential expression analysis and confirmed Set7 mRNA expression knockdown of between 76 and 85 % (Supplemental Figure S1a). From the 55,889 genes annotated in the GENCODE v14 database, 18,494 genes were detected (average of 10 reads per sample). The multi-dimensional scaling plot (Supplemental Figure S1b) indicates grouping of Set7KD and NTC samples by dimension 1, with the smaller dimension 2 consisting predominantly of genes whose expression is altered by biological replicate. Differential gene expression analysis identified 3,901 genes up-regulated and 4,247 genes down-regulated by Set7 depletion (FDR $p \leq 0.05$). The large number of gene expression differences is graphically depicted as a smear plot in Supplemental Figure S1c.

Pathway analysis

To gauge functional changes associated with transcriptional deregulation in endothelial cells depleted of Set7 expression, differentially expressed genes identified by mRNA-Seq were analyzed for canonical pathway enrichment. Gene ontology analysis using the MetaCore™ GeneGo Pathway Analysis tools revealed widespread changes to multiple gene networks in Set7KD samples. Moreover, separate

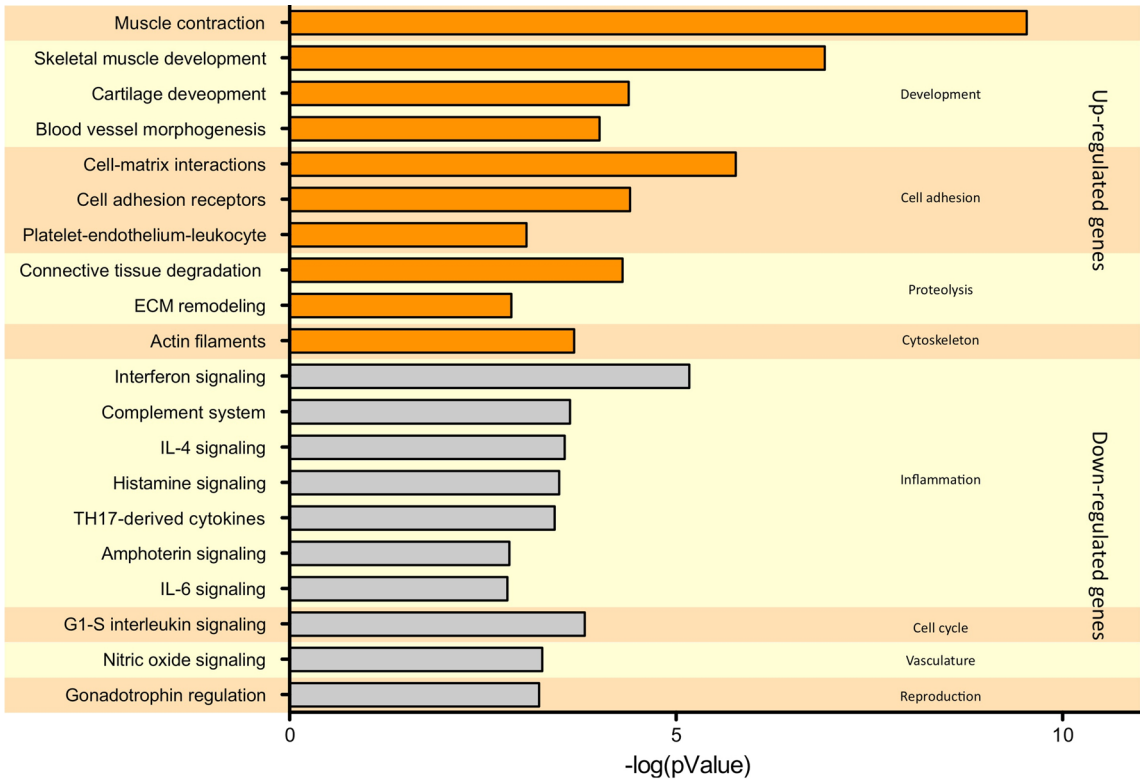

Fig. 2 Vascular pathways are differentially regulated by Set7 knockdown. Widespread transcriptional deregulation in Set7KD cells was associated with numerous gene networks with physiological relevance to endothelial dysfunction. Enriched pathways associated with genes up-regulated under conditions of Set7 depletion (*orange bars*) include development of blood vessels, cell adhesion pathways, proteolytic connective tissue degradation and extracellular matrix remodeling. Down-regulated genes were most strongly associated with inflammatory pathways, as well as vascular-specific nitric oxide signaling (*gray bars*)

analysis of genes that were up- and down-regulated by Set7 knockdown shows distinct pathway enrichment. The pathway most significantly affected by Set7 knockdown and strongly associated with up-regulated genes was muscle contraction (Fig. 2, orange bars). Other pathways enriched in the up-regulated gene set include skeletal, cartilage and blood vessel development. Cell adhesion pathways relevant to the endothelium including cell matrix interactions, adhesion receptor regulation and leukocyte recruitment were also enriched. Similarly, several genes that exhibited increased expression in the HMEC-1 model of Set7 knockdown were associated with proteolytic connective tissue degradation and extracellular matrix remodeling, as well as regulation of the cytoskeleton. By contrast, genes down-regulated in Set7KD cells were most strongly associated with inflammation paralleling previous reports [8, 9] (Fig. 2, gray bars). Significantly enriched inflammatory pathways include interferon, histamine, IL-4 and IL-6 signaling. Further, specificity to the vasculature was observed in the down-regulation of genes associated with nitric oxide regulation further implicating Set7 in vascular endothelial function.

H3K4m1 changes at promoters and DREs confer changes in gene expression

Despite previous associations of Set7 to gene regulation by H3K4m1 enrichment, changes to this chromatin mark were indistinguishable by Western immunoblot analysis of cell lysates prepared under acidic conditions from NTC and Set7KD cells (Fig. 3a). This method of analysis is limited for examination of subtle changes at specific loci necessary to establish correlations to gene expression. To investigate the effect of Set7 knockdown on genome-wide H3K4m1 profiles and the association of this modification to gene expression changes observed by mRNA-Seq, ChIP-Seq was performed with antibodies specific for this histone mark. Deep sequencing resulted in data sets containing 22–32 million reads, of which 97–98 % aligned to Hg19 reference and 78–81 % of reads with alignment mapQ $\geq$ 10 (Online Resource and Table S2). X-ChIP-Seq identified 27 peaks ($p \leq 10^{-7}$), and just eight of these were located within 5 kbp of a TSS. ChIP-Seq from MNase-digested material yielded 1,262 peaks altered for H3K4m1. Of these changes, more than 9 out of 10 regions exhibited

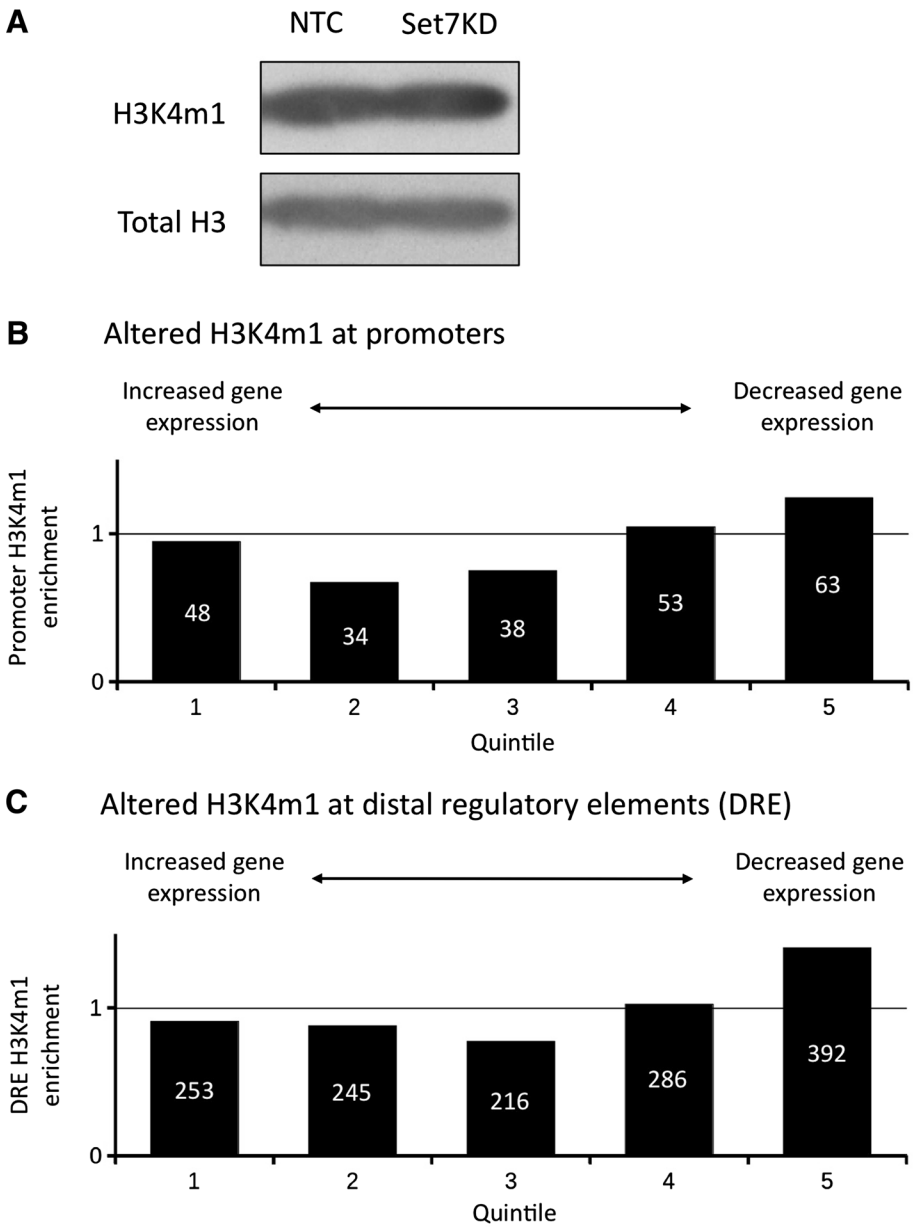

Fig. 3 H3K4m1 changes in S1 chromatin measured by MNase-ChIP-Seq at regulatory elements are not determinative for gene expression changes. **a** Total histone proteins purified by acidic extraction from NTC and Set7KD cells were analyzed by Western immunoblot using antibodies specific to H3K4m1 and total H3 histones. These blots are representative of three independent experiments. **b** Genes associated with loss of H3K4m1 peaks at promoters exhibit a tendency for down-regulation of mRNA expression shown graphically as quintiles (GSEA; $p = 0.027$). **c** Genes associated with loss of H3K4m1 peaks at DREs are more numerous and exhibit a similar tendency toward down-regulation of mRNA expression shown graphically as quintiles (GSEA; $p = 0.006$) The number of genes binned in each quintile is indicated

reduced H3K4m1. The loss of H3K4m1 events could be associated with the promoters of 388 genes with decreased mRNA expression (GSEA $p = 0.027$, 140 down-regulated, 95 up-regulated, 153 not detected; Fig. 3b). To investigate whether loss of Set7 expression conferred changes to distal enhancer H3K4m1 profiles, we filtered for peaks that occurred at DREs as previously described by [28] and analyzed the expression of genes assigned to these distal regions. The 1,262 peaks obtained by MNase-derived ChIP-Seq mapped to DREs implicated in the regulation of 2,198 genes. From the 2,192 genes associated with a loss of H3K4m1 due to Set7KD, expression of 1,392 genes was detected in our mRNA-Seq experiment, with 798 reduced and 594 increased in expression ($p = 0.006$ for association of loss of H3K4m1 at enhancers with reduced gene expression; Fig. 3c). These results are important for two reasons. First, our data suggest the loss of methylation at proximal and distal regulatory elements conferred by Set7KD correlated with reduced gene expression in the majority of cases (60 % at promoters and 57 % at DREs). Recent experimental data suggest that Set7 is not alone in writing H3K4m1 [31] and the contribution to altered activity of other histone-modifying enzymes to transcriptional changes cannot be discounted. To this end, mRNA expression of H3K4 KMTs MLL3 and MLL4 as well as lysine demethylase LSD1 was deregulated in Set7KD cells (Table S3). Second and more important to understanding Set7 in regulation, our data indicate gene expression changes independent of H3K4m1.

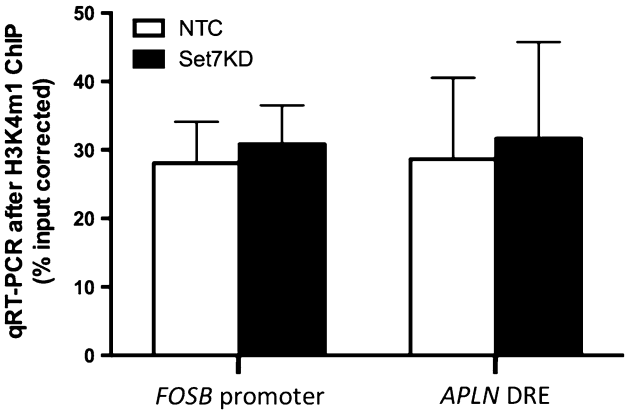

Fig. 4 H3K4m1-independent changes at the *FOSB* promoter and *APLN* DRE. Soluble chromatin was isolated from the NTC and Set7KD cells and immunopurified with anti-H3K4m1 antibody. Quantitative RT-PCR was used to measure the level of enrichment, % input-corrected normalized to NTC cells. Data represent mean values of three independent experiments. *Error bars* represent SEM

Indeed, our independent analysis of MNase-derived ChIP DNA by qRT-PCR confirmed stable H3K4m1 enrichment at *FOSB* promoter and a DRE associated with *APLN* in Set7KD cells, (Fig. 4) despite significantly reduced expression of these genes (data not shown).

Previously characterized Set7 substrates are associated with changes in gene expression

To explore alternative mechanisms of transcriptional regulation, we examined the expression profile of Set7KD cells for regulatory associations to known transcription factor targets. Drawing from publicly available gene sets of specific transcription factor knockdown and inhibition experiments across various human cell lines, GSEA of the mRNA-Seq gene set identified six previously characterized transcription factor substrates with strong connectivity to differential gene expression (Table 1). This method identified the overall number of transregulatory targets as well as the number of unique targets connected to individual factors. Remarkably, overall factor enrichment was significantly associated with gene repression and not genes up-regulated by Set7KD. Furthermore, a high degree of overlap was observed for transcription factors with proportionally few unique gene associations. NFκB was identified with the highest connectivity to the mRNA-Seq data set (886 genes). Previous studies associate Set7-mediated methylation at specific lysine residues of the NFκB-p65 subunit with transcriptional events [19, 20]. In addition, characterized Set7 substrates STAT3 [27], IRF1 [3] and ERα [32] were associated with 608, 70 and 184 down-regulated genes, respectively. Consistent with previous reports

Table 1 Factors associated with gene expression changes in Set7KD HMEC-1s

Factor	Connected genes		FDR *q* value
	Overall	Unique	
1. NFκB	*886 (20.86 %)*	*1*	*0.001*
2. STAT3	*608 (14.32 %)*	*8*	*0.021*
3. IRF1	*70 (1.65 %)*	*2*	*0.022*
4. p53	*112 (2.64 %)*	*2*	*0.02*
5. ERα	*184 (4.33 %)*	*0*	*0.033*
6. TAF7	*188 (4.43 %)*	*0*	*0.04*
1. FOXA1	845 (19.89 %)	14	0.01
2. STAT1	375 (8.83 %)	3	0.008
3. JUND	750 (17.66 %)	9	0.02
4. STAT2	125 (2.94 %)	5	0.021
5. POL2	677 (15.9 %)	4	0.019
6. p300	634 (14.92 %)	1	0.022
7. c-FOS	816 (19.21 %)	4	0.022
8. c-JUN	516 (12.15 %)	7	0.021
9. JUNB	229 (5.39 %)	1	0.021
10. TCF12	650 (15.3 %)	2	0.021
11. NELFE (RDBP)	71 (1.67 %)	6	0.021
12. GCRα	316 (7.44 %)	5	0.02
13. BCL3	235 (5.53 %)	2	0.02
14. FOSL2	521 (12.27 %)	3	0.02
15. SIRT6	70 (1.65 %)	0	0.019
16. BATF	261 (6.14 %)	3	0.02
17. GATA-1	532 (12.53 %)	1	0.02
18. BRG1	32 (7.53 %)	11	0.02
19. GATA-2	422 (9.94 %)	6	0.021
20. NR4A1	9 (0.21 %)	3	0.02
21. ERRA	30 (0.71 %)	3	0.019
22. C/EBPβ	1,109 (26.11 %)	3	0.021
23. RXRA	427 (10.05 %)	2	0.021
24. HNF4G	358 (8.43 %)	2	0.022
25. RFX5	731 (17.21 %)	1	0.022
26. HDAC2	521 (12.27 %)	2	0.022
27. SP1	1,140 (26.84 %)	3	0.023
28. IRF4	280 (6.59 %)	3	0.024
29. FOSL1	196 (4.62 %)	4	0.024
30. SREBP2	29 (0.68 %)	2	0.024
31. MEF2C	104 (2.45 %)	0	0.024
32. HDAC8	20 (0.47 %)	5	0.025
33. GTF2B	188 (4.43 %)	2	0.026
34. NRSF	210 (4.94 %)	0	0.026
35. USF2	682 (16.06 %)	1	0.027
36. ZBTB33	202 (4.76 %)	1	0.027
37. TAL1	333 (7.84 %)	1	0.027
38. HNF4A	461 (10.85 %)	1	0.028
39. TFIIIC	73 (1.72 %)	1	0.029
40. BHLHE40	233 (5.49 %)	2	0.03

Table 1 continued

Factor	Connected genes		FDR *q* value
	Overall	Unique	
41. GTF2F1	512 (12.06 %)	4	0.031
42. PAX5	813 (19.14 %)	1	0.031
43. PBX3	437 (10.29 %)	0	0.033
44. ELK4	432 (10.17 %)	1	0.034
45. ZNF263	162 (3.81 %)	8	0.033
46. USF1	1,088 (25.62 %)	1	0.035
47. AP-2G	682 (16.06 %)	0	0.036
48. EBF1	839 (19.76 %)	0	0.035
49. TCF4	256 (6.03 %)	0	0.036
50. PU.1	840 (19.78 %)	0	0.037
51. AP-2A	558 (13.14 %)	1	0.039
52. CHD2	512 (12.06 %)	0	0.039
53. BDP1	4 (0.09 %)	0	0.039
54. TBP	608 (14.32 %)	0	0.04
55. PRDM1	74 (1.74 %)	0	0.04
56. ZEB1	633 (14.90 %)	0	0.039
57. SREBP1	221 (5.20 %)	0	0.041
58. CTCFL	476 (11.21 %)	0	0.044
59. MEF2A	254 (5.98 %)	0	0.048
60. SMC3	519 (12.22 %)	0	0.049

Factors with previous associations to Set7 are italics. The number of gene associations and the percentage of overall repressive changes as well as the number of associations unique to each factor are described. FDR: false discovery rate

of post-translational regulation by Set7 [16], the classical tumor suppressor protein p53 was strongly implicated in the transcriptional deregulation of Set7KD HMEC-1 cells by connectivity to 122 genes. Similarly, the previously described in vitro Set7 methylation target TAF7 [33] was connected to 188 genes.

Validation of changes to gene expression associated with NFκB-p65

To test the validity of transcriptional changes identified by mRNA-Seq, we examined the expression of select gene targets by qRT-PCR. Because of previous descriptions of strong associations to Set7 [8, 9, 19, 20, 34] and the robust connectivity to gene expression changes in Set7KD HMEC-1 s, downstream transactivation targets of NFκB-p65 were selected for analysis. Under basal conditions, p65 (*RELA*) mRNA expression was unaffected by Set7KD (Fig. 5). The direction of mRNA-Seq transcriptional changes down-regulated in Set7KD samples was validated using this independent approach. Specifically, robust down-regulation of *JUN*, *IL1B, CCL2, SLC7A11* and *ATF3* was observed in mRNA samples prepared from Set7KD cells.

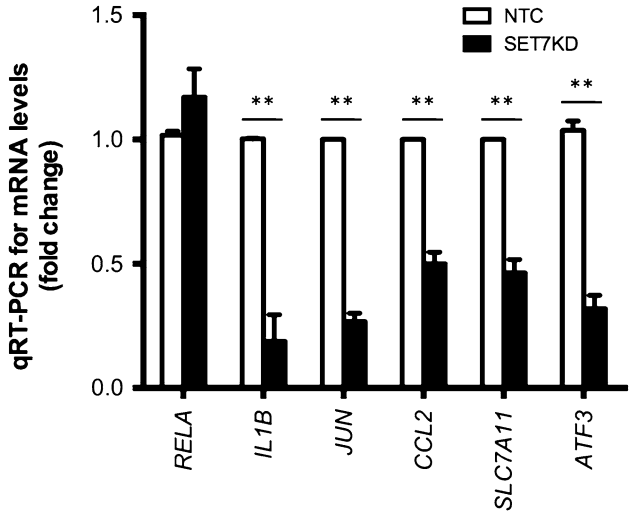

Fig. 5 Validation of changes to gene expression conferred by Set7KD and associated with p65 transactivity. Transcriptional expression of transregulatory targets of p65 in Set7KD vascular endothelial cells (*closed bars*) compared with NTC cells (*open bars*). *RELA* and downstream NFκB-p65 targets *IL1B*, *JUN, CCL2, SLC7A11* and *ATF3* mRNA levels were determined by quantitative qRT-PCR using *ACTB* as the internal control. Data represent mean values of three independent experiments. *Error bars* represent SEM $^{**}p < 0.01$

Additional factors associated with transcriptional changes in the Set7KD endothelial cell

In addition to the previously characterized Set7 substrates, we identified a further 60 transcription factors implicated in the deregulated transcription profile of Set7KD cells (Table 1). In combination, the 66 transcription factors identified by this analysis were associated with 53.31 % (2,264 genes) of the deregulated gene expression of Set7KD HMEC-1s. The combined deregulated activities of factors without previous associations to Set7 account for 53.07 % (2,254 genes) of the transcriptional deregulation observed by mRNA-Seq, contrasting 31.6 % (1,342 genes) associated with NFκB, STAT3, IRF1, p53, ERα and TAF7. Of the factors without previous associations to Set7, FOXA1 and STAT1 displayed the most significant connectivity to the mRNA-Seq data set. In fact, only NFκB-p65 displayed greater connectivity. FOXA1 was highest ranked by statistical significance with 845 gene associations, of which 14 were not associated with other factors. Similarly, 11 of 32 deregulated genes connected to the BRG1 network were unique associations. Other factors with high connectivity to deregulated genes include C/EBPβ, SP1 and USF1 with 1,109, 1,140 and 1,088 associations, respectively. Additional factors implicated by this study include general transcription factors such as BRG1, GTF2F1, GTF2B and TFIIIC, as well as components of the pleiotropic AP-1 transcription factor

complex: JUNB, JUND, c-FOS, c-JUN, FRA1 and FRA2. Furthermore, STAT2, the p300, HDAC8 and HDAC2 regulators of acetylation, as well as angiogenesis-associated GATA and endothelial-specific MEF2 transcription factors, were implicated in gene expression changes of the Set7KD cell.

Novel candidate substrate prediction

Since loss of Set7 was associated with subtle changes in H3K4m1 (Fig. 3), we explored the possibility that widespread transcriptional deregulation could occur downstream of changes to uncharacterized lysine methylation events and subsequent modulation of transcription factor transactivity. Rapid expansion of characterized non-histone substrates has derived a consensus recognition sequence to predict novel lysine substrates for Set7 [3]. To link Set7 function with transcription factor regulation in endothelial cells, the formula was applied to the amino acid sequences of factors identified through connectivity to gene deregulation identified by mRNA-Seq. This analysis revealed 29 potential sites for Set7 methylation across peptide sequences of 19 transcription factors not previously associated with Set7 function (Table 2). To determine physiological validity of predicted amino acid motifs, the surface accessibility for chemical modification was inferred from hydrophobicity of amino acids using the NetSurfP-1.1 software package [29]. The bioinformatic approach to site prediction is validated in part by the high relative surface accessibility predicted of the previously described modified lysines of characterized substrates (Table 2). Several candidate substrates contain multiple predicted sites for methylation by Set7. Pertinent examples are sites exposed at tertiary structures of BRG1, RFX5, GTF2F1, AP-2G and CTCFL. The most commonly observed Set7 methylation site within the peptide sequences centers around the KSK amino acid triplet [22]. STAT1 and GATA2 harbor this motif at K87 and K408, respectively, with the latter candidate site encompassing lysines with high relative residue surface accessibility. Several closely related variants of the recognition sequence were also detected surrounding lysines with high relative surface accessibility. The majority of predicted sites are marked by lysines at the −2 position followed by lysine, arginine, proline or serine at −1. Prominent examples include variations centered on the KKK and KPK triplets. Similar frequencies were observed for motifs surrounding lysines of RKK and RRK triplets. Several candidate methylation sites have been associated with post-translational modification, exemplified by acetylation at K316 of GATA1 [35]. Putative sites at K87 of STAT1 and K21 of BCL3 were predicted to be buried within tertiary protein structures.

Biochemical validation of novel substrate methylation

Previous descriptions of transcription factor methylation have revealed a role for Set7 in the regulation of protein stability. To investigate potential impacts of Set7 depletion on novel candidate factors implicated in the current study, we analyzed protein expression of STAT1, BRG1 and NRSF by Western immunoblot with specific antibodies (Fig. 6). Analyses of whole cell lysates indicate that Set7 depletion did not affect the stability of this subset of predicted substrates. However, changes to the activity of these factors by methylation as previously described for several confirmed Set7 substrates cannot be discounted. To test the validity of transcriptional changes identified by mRNA-Seq implicated in the prediction of novel methylation substrates for Set7, we examined the expression of key gene targets by qRT-PCR. Because of the highly significant connectivity of STAT1 to transcriptome changes, this analysis focused on STAT1 transregulatory gene targets. *MX1*, *BST2* and *IRF7* were significantly down-regulated in Set7KD HMEC-1s (Fig. 7). To investigate the potential for Set7 to methylate STAT1, we confirmed methyltransferase activity of FLAG-Set7 against a 14 amino acid peptide fragment, representing 680–693 of the STAT1 protein sequence (Fig. 8). Mutation of the predicted target lysine residue (representing K685 of the native protein) to arginine (STAT1mutR) reduced this activity, which was similar to the change in Set7 methyltransferase activity observed using histone H3K4 and H3R4 peptides. In accordance with our observations of robust connectivity to changes in gene expression (Table 1), these results provide strong evidence that Set7 could methylate this motif of STAT1, and thereby modulate widespread changes in gene expression in Set7KD cells.

Discussion

The role of Set7 in regulating chromatin structure and function

The post-translational modifying activities of KMTs are central to a variety of cellular processes, regulating substrates as diverse as proteins, DNA, RNA, lipids, metabolites and signaling molecules. Recent advances permitting integration of high-throughput profiling of multiple levels of biological information provides a platform to better understand these pathways. The ability for massive parallel sequencing to interrogate the transcriptome of vascular endothelial cells has revealed novel insight into diverse signaling relevant to transcriptional regulation. As for Set7-mediated lysine methylation, we identified strong enrichment of pathways regulating critical processes associated

Table 2 Methylation substrate prediction

Factor	Accession	Lysine	−3	−2	−1	K	+1	+2	Lys RSA	Full RSA	Class
NFκB-p65	*Q04206*	*K37*	*F*	*R*	*Y*	***K***	*C*	*E*	*0.429*	*0.309*	*E*
		K314	*S*	*I*	*M*	***K***	*K*	*S*	*0.541*		*E*
		K315	*I*	*M*	*K*	***K***	*S*	*P*	*0.546*		*E*
STAT3	*P40763*	*K140*	*V*	*T*	*E*	***K***	*Q*	*Q*	*0.514*	*0.299*	*E*
IRF1	*P10914*	*K126*	*S*	*K*	*S*	***K***	*S*	*S*	*0.586*	*0.329*	*E*
p53	*P04637*	*K372*	*L*	*K*	*S*	***K***	*K*	*G*	*0.597*	*0.331*	*E*
ERα	*P03372*	*K302*	*K*	*R*	*S*	***K***	*K*	*N*	*0.35*	*0.302*	*E*
TAF7	*Q15545*	*K5*	*S*	*K*	*S*	***K***	*D*	*D*	*0.570*	*0.375*	*E*
STAT1	P42224	K87	R	K	S	**K**	R	N	0.240	0.315	B
		K685	S	R	P	**K**	E	A	0.620		E
NELFE	P18615	K26	K	K	K	**K**	K	A	0.416	0.458	E
BCL3	P20749	K21	T	R	P	**K**	A	A	0.145	0.236	B
GATA1	P15976	K316	G	K	K	**K**	R	G	0.595	0.374	E
BRG1	P51532	K443	K	R	S	**K**	R	Q	0.304	0.344	E
		K587	K	K	K	**K**	K	A	0.411		E
GATA2	P23769	K408	K	K	S	**K**	K	G	0.555	0.320	E
NR4A1	Q6IBU8	K354	S	K	P	**K**	Q	P	0.420	0.311	E
C/EBPβ	P17676	K264	S	K	A	**K**	K	T	0.501	0.333	E
RFX5	P48382	K267	S	K	P	**K**	N	G	0.539	0.478	E
		K472	P	R	K	**K**	S	G	0.597		E
MEF2C	Q06413	K5	G	R	K	**K**	I	Q	0.402	0.298	E
NRSF (REST)	Q13127	K517	S	K	T	**K**	K	S	0.312	0.367	E
TFIIIC	Q12789	K1232	K	R	K	**K**	K	G	0.447	0.327	E
GTF2F1 (T2FA)	P35269	K201	G	R	R	**K**	A	S	0.604	0.451	E
		K250	K	K	K	**K**	K	G	0.563		E
		K251	K	K	K	**K**	G	S	0.504		E
		K290	S	K	A	**K**	A	P	0.454		E
		K339	K	R	R	**K**	D	S	0.573		E
PAX5	Q02548	K89	S	K	P	**K**	V	A	0.375	0.274	E
ELK4	P28324	K320	S	R	S	**K**	K	P	0.681	0.425	E
AP-2G (AP2C)	Q92754	K241	S	K	Y	**K**	V	T	0.428	0.302	E
		K295	G	R	R	**K**	A	A	0.369		E
OCT2	P09086	K300	R	R	K	**K**	R	T	0.478	0.282	E
CTCFL	Q8NI51	K251	R	K	T	**K**	G	A	0.380	0.381	E
		K335	R	R	Y	**K**	H	T	0.517		E
		K588	R	K	R	**K**	Q	T	0.396		E
MEF2A	Q02078	K5	G	R	K	**K**	I	Q	0.432	0.292	E

Amino acids flanking the target lysine (bold and bold italics) are depicted for factors previously identified as Set7 substrates (italics) and factors with novel associations to Set7. Relative surface accessibility (RSA) for target lysines relative to RSA for all amino acids of individual proteins (Full RSA) indicates the likelihood that a residue is exposed (E) or buried (B) within the tertiary structure of a protein

with endothelial dysfunction and vascular disease. Our pathway analysis indicates Set7 is a negative regulator of genes broadly involved in developmental processes encompassing ECM deposition, cytoskeleton, morphogenesis and remodeling. On the other hand, Set7 positively regulates a range of inflammatory processes specifically related to interleukin signaling, innate immune function and immediate early response. These results indicate that Set7 plays a critical role in endothelial cells positively regulating innate immune responses that extends beyond a previously described role for Set7 in activating *NFκB-p65* (*RELA*) gene expression in the vasculature [8]. A role for Set7 in NFκB-dependent activation of inflammatory gene promoters has been described for monocytes as well as macrophages from diabetic mice [34]. Although we did not observe specific enrichment of apoptotic and cell cycle regulatory pathways (Fig. 2), altered expression of individual genes encoding cyclins and *MOAP1* (data not shown) could account for some of these phenotypic changes in Set7KD cells.

As a KMT initially ascribed to writing permissive monomethyl marks to lysine 4 of H3 [6, 14], the importance

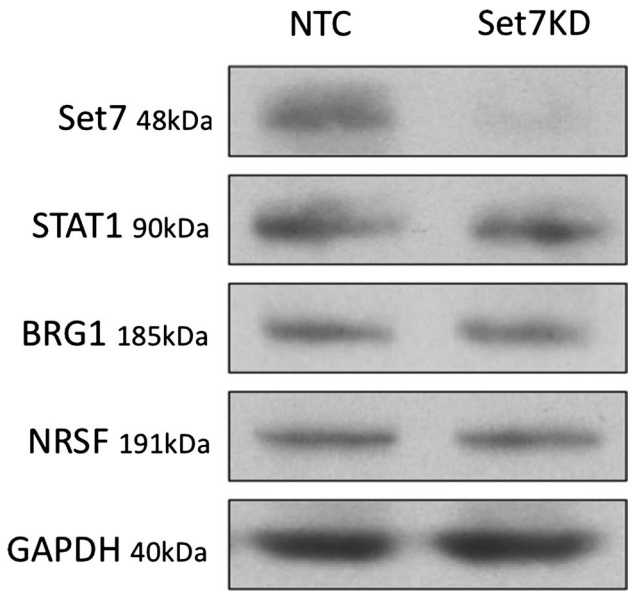

Fig. 6 Stable protein expression of the predicted substrates in Set7KD cells. Whole cell extracts were prepared from NTC and Set7KD endothelial cells and analyzed by immunoblot with antibodies specific to Set7, STAT1, BRG1, and NRSF. GAPDH was used as a loading control. These blots are representative of three independent experiments

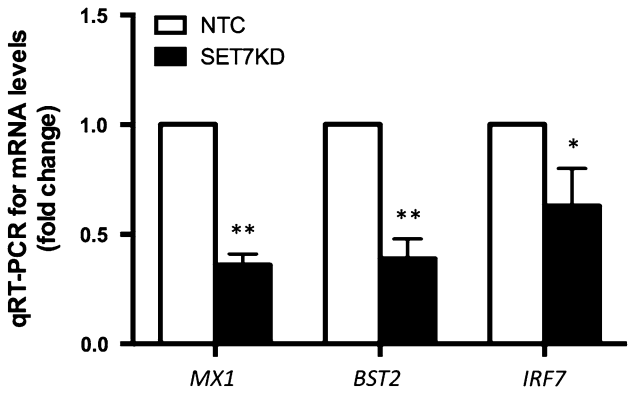

Fig. 7 Validation of changes to gene expression conferred by Set7KD and associated with STAT1 transactivity. Transcriptional expression of transregulatory targets of STAT1 in Set7KD vascular endothelial cells (*closed bars*) compared to NTC cells (*open bars*). Downstream STAT1 targets *MX1, BST2* and *IRF7* mRNA levels were determined by qRT-PCR using *HPRT1* as the internal control. Data represent mean values of three independent experiments. *Error bars* represent SEM $*p < 0.05$ $**p < 0.01$

of differentially localized histone methylation patterns to Set7-mediated transcriptional regulation may extend beyond gene promoters. For instance, genome-wide analysis revealed H3K4 methylation patterns at more than 3,500 distal DNase hypersensitive sites in human T cells including regulatory elements with functional associations to *IFNG* and *IL13* enriched for H3K4m1 [36]. A similar study

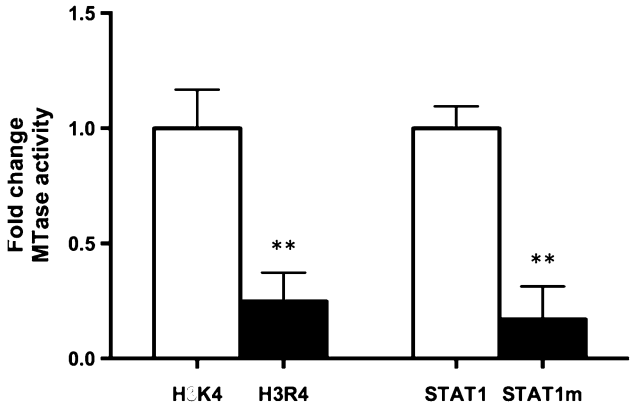

Fig. 8 Set7 methylates STAT1 peptide in vitro. FLAG-tagged Set7 protein was incubated with [H^3]-S-adenosyl-methione (373 nM) and 200 pmol of histone (H3K4 or H3R4) or short STAT1 (STAT1 or STAT1 m) peptides containing the predicted substrate sequences for 1 h at 30 °C. Tritiated histone peptide was measured by liquid scintillation. Activity is presented as fold-changes normalized to H3K4 or STAT1 peptide. Data represent mean values of three independent experiments. Error bars represent SEM $**p < 0.0035$

of 30 Mb of the HeLa genome reported strong H3K4m1 enrichment at enhancer elements with a common role for the H3K4m1 signature in enhancer-associated transcriptional regulation confirmed by the accurate prediction and validation of the location and activity of multiple independently identified DREs [13]. Integration of our mRNA-Seq and H3K4m1 ChIP-Seq data sets revealed a subtle association between the loss of this histone modification at DREs and transcriptional deregulation of nearby genes (Fig. 3). Combined with similar associations to promoters of deregulated genes, these data are indicative of a small yet significant proportion of Set7-mediated H3K4m1-dependent vascular endothelial gene expression. The underwhelming association of gene repression and H3K4m1 depletion as well as our results describing stable K4m1 for H3 histones prepared under acidic conditions is in accordance with reports of stable global histone methylation in cells depleted of Set7 expression [19, 37, 38]. Despite a strong level of Set7 mRNA and protein depletion in our Set7KD cells, we cannot discount the possibility that a small proportion of active Set7 could maintain levels of H3K4m1 enrichment. Furthermore, recent experimental evidence suggests that Set7 is not the only mammalian enzyme responsible for writing the mono-methyl modification [31]. Similarly, it is currently unclear if tri- or di-methylated H3K4 are reduced to the mono-methylated form through subsequent demethylase reactions. Several additional enzymes associated with methylation at H3K4 were altered for mRNA expression in Set7KD cells including the LSD1 H3K4m1/2 demethylase. Thus, changes to chromatin modifications indirectly mediated by Set7 require further investigation. While mechanisms independent of

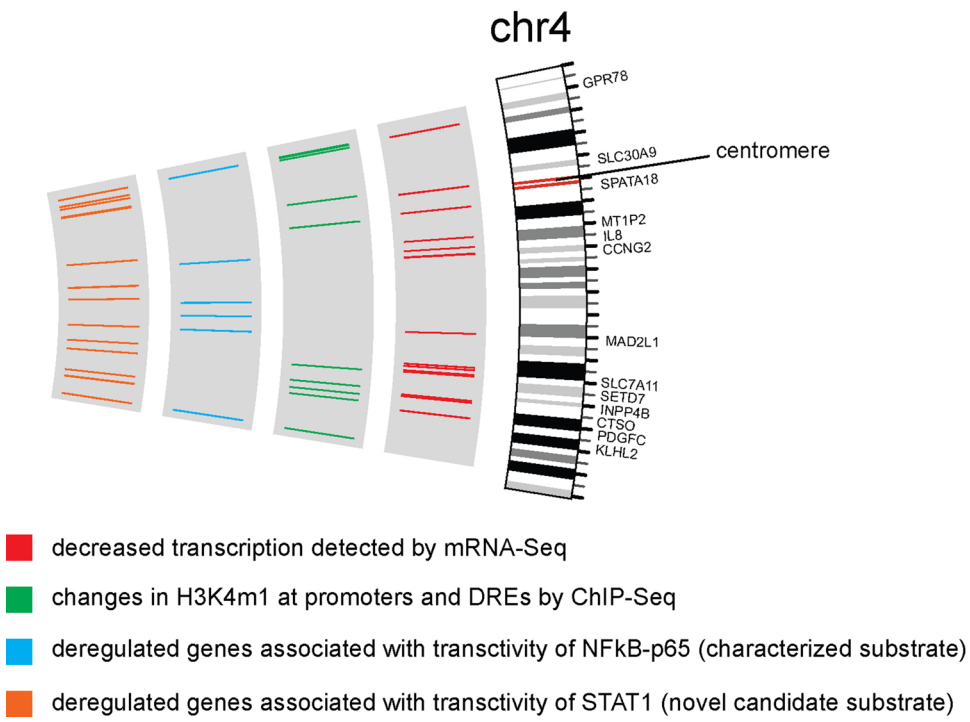

Fig. 9 Extensive changes to gene expression associated with Set7 are regulated by H3K4m1 and transcription factor-dependent mechanisms. The positions of genes down-regulated by Set7KD are shown for chromosome 4. *Red bars* depicting transcriptional repression events identified by mRNA-Seq show numerous changes in Set7KD cells. *Green bars* represent the association of gene expression to changes to H3K4m1 patterns determined by ChIP-Seq. Deregulated genes associated with transactivity of characterized Set7 substrate NFκB-p65 are represented by *blue bars*. *Orange bars* depict deregulated genes associated with transactivity of the novel candidate Set7 substrate STAT1

H3K4m1 clearly regulate gene expression in Set7KD cells, activation of discrete promoters by this modification in endothelial and other cell types [8–10, 37, 39–41] may be more prominent in the context of inducible gene expression. The data shown here indicate the strong contribution of histone-independent mechanisms of Set7 activity in vascular endothelial cells, which could be mediated by lysine methylation to signaling proteins and transcription factors. Figure 9 describes the associations to down-regulated gene expression for H3K4m1 enrichment as well as NFκB-p65 (characterized substrate) and STAT1 (novel candidate substrate) for chromosome 4. The full circos plot describing these associations genome-wide is included as Supplemental Figure 2. Down-regulated genes that are associated with NFκB-p65 and STAT1 are described in Table S4.

Transcription factors with novel and previously characterized associations to Set7

Recent association of lysine methylation to transcription factor stability and function [16, 17] has revealed the dual role of KMTs such as Set7 in histone-dependent and histone-independent mechanisms of transcriptional regulation, although the relative contribution of each of these mechanisms remains to be determined. Gene ontology resources mapped a large proportion of differential expression identified by mRNA-Seq to a list of transcription factors that importantly included six factors with previously described associations to Set7. Enrichment of STAT3, IRF1, p53, ERα, TAF7, as well as the particularly strong connectivity of NFκB to the data set and independent validated gene expression changes downstream of NFκB-p65 endorse our approach to identification of factors associated with Set7 function and regulation of transcription. Intriguingly the contribution of these and other factors not previously linked with Set7 function was associated with gene repression and not significantly with genes that were increased for expression in Set7KD cells. Accordingly, loss of KMT activity by Set7 can repress transcription by altering the transactivity of NFκB-p65 [19] and STAT3 [27], as well as reducing the stability of p53 [17]. In addition to the potential role for Set7-mediated methylation of p65 in positive regulation of innate immunity and inflammation as indicated by down-regulation of these pathways in Set7KD cells, the robust association of AP-1 components to changes in gene expression further implicates Set7 in regulation of these processes as well as immediate early response genes. Genes encoding several AP-1 factors exhibited transcriptional deregulation in Set7KD cells (for instance *JUN* and *FOS*). Furthermore, the enrichment of deregulated genes mapped to differentiation-associated FOXA1, as well as the lymphoid-specific IRF4, highlights potential roles for Set7 in mediating signatures in cells other than the endothelial cell [41].

Predicting novel methylation substrates

Protein interaction data have accurately predicted several biochemically validated in vivo non-histone substrates for methylation by Set7 [3]. Integration of the transcription factor enrichment data extrapolated from the Set7KD

transcriptome profile and the motif recognition formula revealed putative methylation sites several factors not previously associated with Set7. By focusing on factors associated with transcriptional changes, we were able to infer protein interactions from gene expression data and identify putative methylation sites in factors associated with proliferative, inflammatory and general transcriptional pathways in vascular endothelial cells. To increase the validity of our predictions, stringent application of the consensus formula meant substrate motifs were highly predictive for these residues. Indeed, not all characterized methylation substrates for Set7 strictly adhere to this formula. For instance, the motif surrounding K372 of p53 includes threonine and glutamate at −2 and −1 relative to the target lysine. Thus, accommodating sequences that diverge from the consensus at individual residues could expand the list of candidate substrates proposed by the current study. To this end, three putative methylation sites of the glucocorticoid receptor (K256, K494 and K496) were excluded because of lysine residues at the +2 position, though this amino acid position may exhibit a level of redundancy in Set7 binding [3]. Similarly, the motif sequences at lysines of the p65 protein targeted by Set7 diverge considerably from the formula. Alternative prediction sequences have been proposed that may prove useful for further identification of Set7 methylation substrates [42]. Amino hydrophobicity analysis indicated most predicted motifs to be accessible to post-translational modification and strong surface accessibility prediction of characterized methylation motifs of known substrates implicated in the gene deregulation validated the peptide modeling approach. Our in vitro STAT1 peptide methylation assays accordingly demonstrate the predictive value of our method to identify novel candidate substrates to analyze in vivo. STAT1-dependent transcription and signaling have been strongly linked with vascular dysfunction and inflammation [43]. Specifically, pro-inflammatory signal integration of IFNγ, IL-6, and TLR4 ligand such as lipopolysaccharide is dependent on STAT1 in endothelial and smooth muscle cells [44, 45]. Further connecting to previously described roles for Set7 [9, 34], altered regulation of STAT1 was recently associated with hyperglycemic memory in monocytes derived from patients with type 1 diabetes [46].

Conclusions and future considerations

The current study employed the strategy of mapping statistically significant gene expression changes to transcription factor networks in Set7KD cells, thereby focusing on biological relevant substrates. The close correlation between these motifs and validated methylation sites of previously described substrates provides strong predictive data for novel substrate identification by biochemical investigation. Further characterization of candidate transcription factor substrates, including in vivo validation of our peptide methyltransferase findings for STAT1 K685, may reveal novel mechanisms of gene regulation by Set7 including functional modulation of transcription factors, as well as substrate-driven co-recruitment of the enzyme to specific promoters to potentiate H3K4m1 enrichment [27, 47]. With associations to inflammatory vascular complications and the potential for targeted therapeutic interventions, characterization of the histone-dependent and transcription factor-dependent mechanisms regulating gene expression is central to a better understanding the role of Set7 [48, 49]. The inability to explain transcriptional up-regulation by these mechanisms is perhaps associated with factors beyond the scope of the current study such as changes in kinase activity or other signaling pathways. As public data accumulates, the potential to identify novel associations and mechanisms of transcriptional control by intersecting ever-increasing data sets such as transcriptome and ChIP-Seq profiles grows rapidly. Because the dependency of the ChIP assay on antibody specificity and quality can impact the reliability of findings, intersection of our ChIP-Seq data with data from similar studies of H3K4m1 can be used to verify our observations. Exemplified by adoption of nomenclature for KMTs reflecting the ubiquity of lysine modifications historically described for histones [50], the method employed by this study is applicable to substrate prediction for other broad-substrate modifications such as lysine acetylation [51].

Acknowledgments The authors are thankful for expert bioinformatic support by Antony Kaspi and Dr. Ross Lazarus. The authors acknowledge grant and fellowship support from the Juvenile Diabetes Research Foundation International, the Diabetes Australia Research Trust, the National Health and Medical Research Council (NHMRC) and the National Heart Foundation of Australia. STK is supported by an Australian Postgraduate Award. AE-O is a senior research fellow supported by the NHMRC. Supported in part by the Victorian Government's Operational Infrastructure Support program.

References

1. Schwartz D, Chou MF, Church GM (2009) Predicting protein post-translational modifications using meta-analysis of proteome scale data sets. Mol Cell Proteomics 8(2):365–379
2. Basu A et al (2009) Proteome-wide prediction of acetylation substrates. Proc Natl Acad Sci USA 106(33):13785–13790
3. Dhayalan A et al (2011) Specificity analysis-based identification of new methylation targets of the SET7/9 protein lysine methyltransferase. Chem Biol 18(1):111–120
4. Rathert P et al (2008) Protein lysine methyltransferase G9a acts on non-histone targets. Nat Chem Biol 4(6):344–346
5. Rafehi H et al (2014) Vascular histone deacetylation by pharmacological HDAC inhibition. Genome Res
6. Wang H et al (2001) Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase. Mol Cell 8(6):1207–1217

7. Martens JH et al (2003) Cascade of distinct histone modifications during collagenase gene activation. Mol Cell Biol 23(5):1808–1816
8. Brasacchio D et al (2009) Hyperglycemia induces a dynamic cooperativity of histone methylase and demethylase enzymes associated with gene-activating epigenetic marks that coexist on the lysine tail. Diabetes 58(5):1229–1236
9. El-Osta A et al (2008) Transient high glucose causes persistent epigenetic changes and altered gene expression during subsequent normoglycemia. J Exp Med 205(10):2409–2417
10. Okabe J et al (2012) Distinguishing hyperglycemic changes by Set7 in vascular endothelial cells. Circ Res 110(8):1067–1076
11. Heintzman ND et al (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459(7243):108–112
12. Koch CM et al (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome Res 17(6):691–707
13. Heintzman ND et al (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39(3):311–318
14. Nishioka K et al (2002) Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev 16(4):479–489
15. Wang D et al (2013) Methylation of SUV39H1 by SET7/9 results in heterochromatin relaxation and genome instability. Proc Natl Acad Sci USA 110(14):5516–5521
16. Chuikov S et al (2004) Regulation of p53 activity through lysine methylation. Nature 432(7015):353–360
17. Ivanov GS et al (2007) Methylation-acetylation interplay activates p53 in response to DNA damage. Mol Cell Biol 27(19):6756–6769
18. Kurash JK et al (2008) Methylation of p53 by Set7/9 mediates p53 acetylation and activity in vivo. Mol Cell 29(3):392–400
19. Ea CK, Baltimore D (2009) Regulation of NF-kappaB activity through lysine monomethylation of p65. Proc Natl Acad Sci USA 106(45):18972–18977
20. Yang XD et al (2009) Negative regulation of NF-kappaB action by Set9-mediated lysine methylation of the RelA subunit. EMBO J 28(8):1055–1066
21. Calnan DR et al (2012) Methylation by Set9 modulates FoxO3 stability and transcriptional activity. Aging (Albany NY) 4(7):462–479
22. Keating S, El-Osta A (2013) Transcriptional regulation by the Set7 lysine methyltransferase. Epigenetics 8(4)
23. Subramanian A et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102(43):15545–15550
24. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760
25. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6):841–842
26. Zhang Y et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9(9):R137
27. Yang J et al (2010) Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. Proc Natl Acad Sci USA 107(50):21499–21504
28. Thurman RE et al (2012) The accessible chromatin landscape of the human genome. Nature 489(7414):75–82
29. Petersen B et al (2009) A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol 9:51
30. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140
31. Herz HM et al (2012) Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev 26(23):2604–2620
32. Subramanian K et al (2008) Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase. Mol Cell 30(3):336–347
33. Couture JF et al (2006) Structural basis for the methylation site specificity of SET7/9. Nat Struct Mol Biol 13(2):140–146
34. Li Y et al (2008) Role of the histone H3 lysine 4 methyltransferase, SET7/9, in the regulation of NF-kappaB-dependent inflammatory genes. Relevance to diabetes and inflammation. J Biol Chem 283(39):26771–26781
35. Lamonica JM, Vakoc CR, Blobel GA (2006) Acetylation of GATA-1 is required for chromatin occupancy. Blood 108(12):3736–3738
36. Barski A et al (2007) High-resolution profiling of histone methylations in the human genome. Cell 129(4):823–837
37. Gaughan L et al (2011) Regulation of the androgen receptor by SET9-mediated methylation. Nucleic Acids Res 39(4):1266–1279
38. Lehnertz B et al (2011) p53-dependent transcription and tumor suppression are not affected in Set7/9-deficient mice. Mol Cell 43(4):673–680
39. Deering TG et al (2009) Methyltransferase Set7/9 maintains transcription and euchromatin structure at islet-enriched genes. Diabetes 58(1):185–193
40. Sun G et al (2010) Epigenetic histone methylation modulates fibrotic gene expression. J Am Soc Nephrol 21(12):2069–2080
41. Tao Y et al (2011) The histone methyltransferase Set7/9 promotes myoblast differentiation and myofibril assembly. J Cell Biol 194(4):551–565
42. Masatsugu T, Yamamoto K (2009) Multiple lysine methylation of PCAF by Set9 methyltransferase. Biochem Biophys Res Commun 381(1):22–26
43. Sikorski K et al (2012) STAT1 as a central mediator of IFNgamma and TLR4 signal integration in vascular dysfunction. JAKSTAT 1(4):241–249
44. Sikorski K et al (2011) STAT1 as a novel therapeutical target in pro-atherogenic signal integration of IFNgamma, TLR4 and IL-6 in vascular disease. Cytokine Growth Factor Rev 22(4):211–219
45. Sikorski K et al (2011) STAT1-mediated signal integration between IFNgamma and LPS leads to increased EC and SMC activation and monocyte adhesion. Am J Physiol Cell Physiol 300(6):C1337–C1344
46. Miao F et al (2014) Evaluating the role of epigenetic histone modifications in the metabolic memory of type 1 diabetes. Diabetes 63(5):1748–1762
47. Ko S et al (2011) Lysine methylation and functional modulation of androgen receptor by Set9 methyltransferase. Mol Endocrinol 25(3):433–444
48. Keating ST, El-Osta A (2012) Chromatin modifications associated with diabetes. J Cardiovasc Transl Res 5(4):399–412
49. Keating ST, El-Osta A (2013) Glycemic memories and the epigenetic component of diabetic nephropathy. Curr Diab Rep 13(4):574–581
50. Allis CD et al (2007) New nomenclature for chromatin-modifying enzymes. Cell 131(4):633–636
51. Glozak MA et al (2005) Acetylation and deacetylation of non-histone proteins. Gene 363:15–23
52. GENCODE v14 annotation URL ftp.sanger.ac.uk/pub/gencode/release_14/gencode.v14.annotation.gtf.gz