Aus dem Institut für Molekularbiologie und Tumorforschung

Geschäftsführender Direktor: Prof. Dr. Alexander Brehm

des Fachbereichs Medizin der Philipps-Universität Marburg

Characterization of the transcriptional repressor SAMD1
in hepatocellular carcinoma
and pancreatic ductal adenocarcinoma

Inaugural-Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

dem Fachbereich Medizin der Philipps-Universität Marburg

vorgelegt von

Clara Simon

aus Neuwied

Marburg 2023

Angenommen vom Fachbereich Medizin der Philipps-Universität Marburg am: 09.11.2023

Gedruckt mit Genehmigung des Fachbereichs Medizin

Dekanin: Prof. Dr. Denise Hilfiker-Kleiner

Referent: PD Dr. Robert Liefke

Korreferent: Prof. Dr. Matthias Lauth

Für meine Familie

Table of contents

1. Summary

Until now, the protein SAMD1 had remained largely unexplored. However, recently, we identified SAMD1 as a novel CpG island-binding protein. SAMD1 directly interferes with unmethylated CpG-rich DNA via an atypical winged-helix domain, simultaneously engaging both the major and minor DNA groove. This interaction enables SAMD1 to exert a repressive function on active CpG-islands by recruiting the histone demethylase KDM1A. Moreover, SAMD1 features a SAM domain, facilitating interactions with other SAM domain-containing proteins, and pivotal for homopolymerization. Deletion of SAMD1 in mouse embryonic stem cells leads to augmented H3K4me2 levels and dysregulation of various biological pathways.

SAMD1 exerts a pleiotropic function in cancer, and high *SAMD1* expression can correlate with a favorable or unfavorable prognosis. For instance, in hepatocellular carcinoma, the deletion of SAMD1 reduces clonogenicity and promotes a more advantageous transcriptional network. In contrast, in pancreatic ductal adenocarcinoma (PDAC), SAMD1 operates as a repressor of genes related to epithelial-mesenchymal transition. Consequently, in PDAC cells, SAMD1 knockout prompts a shift towards a more mesenchymal phenotype and accelerates migration rates. *CDH2*, encoding for N-cadherin, is a crucial downstream target of SAMD1 and the main driver of migration after SAMD1 deletion. The DNA binding capability of SAMD1 is modulated by the FBXO11-E3-ubiquitin ligase complex, leading to an overall reduction in SAMD1's chromatin recruitment, specifically in pancreatic ductal adenocarcinoma.

In conclusion, SAMD1's function as a transcriptional repressor is highly context-dependent, as its target genes differ among distinct cell types. Intriguingly, *SAMD1* expression often exhibits prognostic significance across various tumor types, positioning it as a promising candidate gene for further exploration in cancer research.

2. Zusammenfassung

Das Protein SAMD1 wurde bisher nur wenig untersucht. In der folgenden Arbeit konnten wir SAMD1 als neues, CpG Inseln-bindendes Protein identifizieren. Über eine untypische winged-helix Domäne bindet SAMD1 direkt an unmethylierte, CpG-reiche DNA, indem es sowohl mit der großen, als auch der kleinen Furche der DNA interagiert. Mithilfe dieser Interaktion rekrutiert SAMD1 die Histon Demethylase KDM1A und übt seine repressive Funktion aus. Die zweite Domäne von SAMD1 ist eine SAM Domäne, die wichtig für die Interaktion mit anderen Proteinen ist, die ebenfalls SAM Domänen enthalten, und außerdem zur Homopolymerisierung dient. Wenn SAMD1 in murinen embryonischen Stammzellen deletiert wird, führt dies zu erhöhten H3K4me2 Leveln und einer Deregulierung vieler biologischen Signalwege.

In Tumorerkrankungen hat SAMD1 eine pleiotrope Funktion, da die Expression von *SAMD1* sowohl mit guter, als auch mit schlechter Prognose korrelieren kann. Unter anderem führt die Deletion von SAMD1 im hepatozellulären Karzinom zur reduzierter Klonogenität und zu einem vorteilhafteren transkriptionellen Netzwerk. Im Gegensatz dazu hat SAMD1 im duktalen Adenokarzinom des Pankreas eine repressive Wirkung auf Gene, die eine Rolle in epithelial mesenchymaler Transition spielen. Daher führt der Knockout von SAMD1 zu einer Veränderung hin zu einem mesenchymalen Phänotyp und zu erhöhten Migrationsraten. Das Gen *CDH2*, das für N-Cadherin kodiert, ist ein essentielles Zielgen von SAMD1 und führt nach der Deletion von SAMD1 zu erhöhten Migrationsraten. Die DNA-Bindefähigkeit von SAMD1 wird durch den FBXO11-E3-Ubiquitin Ligase Komplex reguliert. Dieser führt zu einer generellen Reduktion in der Chromatinrekrutierung von SAMD1, besonders im duktalen Adenokarzinom des Pankreas.

Zusammenfassend konnte herausgefunden werden, dass die Funktion von SAMD1 als transkriptioneller Repressor sehr kontextspezifisch ist, da sich die Zielgene zwischen verschiedenen Zelltypen stark unterscheiden. Dass die Expression von *SAMD1* jedoch oft in einem prognostischen Zusammenhang in verschiedensten Tumorarten steht, macht *SAMD1* zu einem vielversprechenden Gen für weitere Untersuchungen in der Tumorforschung.

3. Introduction

3.1 CpG islands

CpG islands (CGIs) are prevalent regulatory elements, typically around ~1000 base pairs in length, situated within the promoter regions of vertebrate genomes (Bird, 1987). Remarkably, CGIs are present in approximately 70% of all annotated vertebrate promoters (Sazonov et al., 2006).

In addition to CGIs located at transcriptional start sites, a distinct class known as "orphan" CGIs exists, predominantly situated within intergenic or intragenic regions. Despite their non-conventional positioning, these orphan CGIs are imbued with regulatory significance and hold the potential to serve as alternative promoters (Illingworth et al., 2010; Maunakea et al., 2010).

In contrast to the broader genomic landscape, CpG dinucleotides within CGIs predominantly exist in an unmethylated state. This epigenetic status holds importance, as the hypermethylation of CGIs has been linked to the repression of target genes by preventing the binding of numerous transcription factors (Bell & Felsenfeld, 2000; Mancini et al., 1999; Renda et al., 2007; W.-G. Zhu et al., 2003). Notably, almost all housekeeping and various tissue-specific genes have been described to contain a CGI within their promoter region (Larsen et al., 1992; Zhu et al., 2008).

While recognition motifs like the TATA-Box offer specific cues for transcription initiation, CGIs, in contrast, emerge as more generalized initiators of transcription, thereby assuming a "transcriptionally permissive" role (Deaton & Bird, 2011; Sandelin et al., 2007). This is further supported by the observation that RNA polymerase II (RNAP II) is recruited to promoters containing CGIs, even in the case of inactive genes in mouse embryonic stem cells (Guenther et al., 2007). A notable characteristic of CGIs is their low density of nucleosomes, rendering them highly accessible (Ramirez-Carrozzi et al., 2009). Abundant histone marks on CGIs are acetylation of histone H3 and H4 (Tazi & Bird, 1990), as well as histone H3 trimethylation (H3K4me3). The latter mark often persists even in cases where the corresponding gene is not actively transcribed (Guenther et al., 2007; Mikkelsen et al., 2007). Moreover, H3K4me3 seems to anticorrelate with CpG methylation (Ciccone et al., 2009; Gahurova et al., 2017; Greenfield et al., 2018; Stewart et al., 2015), and it was suggested that H3K4me3 prevents methyltransferase activity (Hu et al., 2009; Morselli et al., 2015).

Only a limited number of proteins interacting directly with unmethylated CGIs have been identified. Among them are the CXXC domain-containing proteins, a group of 12 proteins that can be further classified into four subgroups based on their DNA-binding motif (C. Xu et al.,

2018). These CXXC domain-containing proteins recognize unmethylated CpG-rich DNA through a short zinc finger containing two conserved CXXCXXC motifs (Shin Voo et al., 2000). Some proteins in this category also recruit other proteins, subsequently modifying nucleosomes. For instance, CXXC1 cooperates with the SET1 methyltransferase complex (Lee & Skalnik, 2005), directly influencing the chromatin state by facilitating the addition of H3K4me3 marks (Thomson et al., 2010).

On the other hand, some CXXC domain-containing proteins, such as KDM2A (Lysine Demethylase 2A), exhibit both DNA-binding and nucleosome-modifying capabilities (Blackledge et al., 2010; Tsukada et al., 2006). As a result, CXXC domain-containing proteins can influence the methylation of H3K4 (Thomson et al., 2010) or the demethylation of H3K36 directly or indirectly (Tsukada et al., 2006).

Another group of proteins that interact directly with CGIs are Polycomb-like proteins (PCLs) (H. Li et al., 2017). In vertebrates, three homologs of PCLs have been identified: PHF1, MTF2, and PHF19, also known as PCL 1-3. Each homolog consists of a C-terminal Chromo-like domain and a larger domain comprising a Tudor domain, two PHD fingers, and a winged-helix domain (Fischer et al., 2022; Fischer & Liefke, 2023; H. Li et al., 2017). The winged-helix domain has been shown to interfere directly with CpG-rich DNA by accessing the major groove, making it essential for PCL DNA binding (H. Li et al., 2017).

The binding of PCLs to CGIs results in the recruitment of PRC2 (Polycomb Repressive Complex), leading to the deposition of H3K27me3 (H. Li et al., 2017; Nekrasov et al., 2007). This histone mark is then recognized by the PRC1 complex, which in turn deposits H2AK119Ub1 (de Napoles et al., 2004; H. Wang et al., 2004), resulting in the inhibition of RNA Pol II elongation (Stock et al., 2007; Zhou et al., 2008). PRC1 is also recruited to unmethylated CGIs by KDM2B (Lysine Demethylase 2B), which directly recognizes unmethylated CpG-rich DNA through a CXXC domain (Farcas et al., 2012; X. Wu et al., 2013).

Together, CXXC domain-containing proteins and PCLs play a role in regulating CGIs, leading to active, repressed, or bivalent gene states (Voigt et al., 2013). Recently, a novel class of CGI-binding proteins was discovered. These proteins employ an unconventional winged-helix domain to interact with unmethylated CpG-rich DNA, signifying a novel mechanism of DNA-protein interaction (Stielow, Simon, et al., 2021; Stielow, Zhou, et al., 2021; Weber et al., 2023). This work focuses on one of these novel CGI-binding proteins, called SAMD1 (Sterile Alpha Motif Domain-containing protein 1).

3.2 Hepatocellular carcinoma

Hepatocellular carcinoma (HCC) is the predominant form among all liver cancers, globally ranking as the seventh most frequently diagnosed and fourth most lethal tumor type (International Agency for Research on Cancer, World Health Organization, Cancer Today). A five-year survival rate of only 18% makes HCC the second most lethal tumor besides pancreatic ductal adenocarcinoma (Jemal et al., 2017). In the majority of cases, HCC arises from liver infections. These infections can be attributed to HBV or HCV infections, while alcohol abuse is a notable contributor to liver cancer (Villanueva, 2019). Another risk factor is exposure to aflatoxin, produced by fungi of the *Aspergillus* species (Bressac et al., 1991; Turner et al., 2002). Additionally, the prevalence of nonalcoholic fatty liver disease (NAFLD) is increasing, emerging as a primary driver of HCC in Western nations (Younossi et al., 2019).

The progression of HCC typically initiates with the development of cirrhosis, followed by the emergence of dysplastic nodules, ultimately resulting in the formation of early HCC (Villanueva, 2019). A predominant portion of these tumors manifests a mutation in the *TERT* (Telomerase Reverse Transcriptase) promoter, alongside mutations in *TP53* and *CTNNB1* are common (Schulze et al., 2016). HCC can be further categorized into two distinct molecular subtypes: the proliferative subtype and the non-proliferative subtype. The proliferative category, frequently instigated by HBV infection, exhibits increased chromosomal instability, presenting as a more aggressive variant linked to an unfavorable prognosis. In contrast, the non-proliferative class, characterized by a prevalence of *CTNNB1* mutations, displays a lower degree of dedifferentiation (Villanueva et al., 2011; Zucman-Rossi et al., 2015).

Given that most HCC cases originate from inflammatory conditions arising due to infection, alcohol misuse, or NAFLD, the immune system plays a crucial role in the progression of HCC (Llovet et al., 2016). Within the tumor microenvironment (TME), immune cells undertake diverse roles, either promoting or impeding tumor growth. Notably, myeloid-derived suppressor cells have been demonstrated to repress various other immune cells and foster metastasis and angiogenesis. Conversely, cytotoxic T-cells and natural killer cells emerge as immune cells with anti-tumoral attributes. However, macrophages and B-cells exhibit a dual nature, displaying both pro-tumorigenic and anti-tumorigenic functions (Hao et al., 2021). Of particular significance, the employment of immune checkpoint inhibitors such as anti-PD-1 and anti-CTLA-4 has showcased promising outcomes in clinical trials (Villanueva, 2019).

Other therapy options include surgery or radiotherapy followed by ablation of tumors at lower stages, while advanced cases may necessitate liver transplantation if organ functionality is compromised. Transarterial chemoembolization or selective internal radiation therapy directly

deliver cytotoxic drugs or microspheres laden with radioisotopes to the tumor site, thereby protecting healthy hepatic cells (Villanueva, 2019). Furthermore, therapeutic strategies involving protein kinase inhibitors such as sorafenib and lenvatinib, have led to enhanced survival rates in randomized trials (Kudo et al., 2018; Llovet et al., 2008).

Despite the emergence of novel therapeutic options, there is still a lack of early-detection biomarkers and targeted therapies. Moreover, universal HBV vaccination and better surveillance of early-stage patients could significantly decrease HCC-related deaths (Villanueva, 2019).

3.3   Pancreatic ductal adenocarcinoma

Pancreatic cancer is considered one of the deadliest types of cancer (Halbrook et al., 2023). Even though it is not very common and ranks 15th among all cancers, it still causes a significant number of cancer-related deaths, ranking 9th in that category (International Agency for Research on Cancer, World Health Organization, Cancer Today). Projections suggest that by 2030, pancreatic cancer will ascend to the second most prominent cause of cancer-related deaths in the United States (Rahib et al., 2014). Furthermore, the 5-year survival rate for pancreatic cancer patients is only 10% due to late diagnosis and early metastasis (Siegel et al., 2021). Among pancreatic cancers, pancreatic ductal adenocarcinoma (PDAC) accounts for 90% of all cases (Park et al., 2021). Various factors are known to increase the risk of pancreatic cancer, including habitual cigarette smoking, alcohol abuse, prolonged chronic pancreatitis, diabetes as well as obesity, and germline mutations (Park et al., 2021).

PDAC typically originates from a condition called Pancreatic Intraepithelial Neoplasias (PanIN), and its progression is influenced by multiple genetic mutations (Wood et al., 2022). Among the most commonly mutated genes in PDAC are the oncogene *KRAS*, as well as the tumor suppressor genes *TP53*, *SMAD4,* and *CDKN2A* (Almoguera et al., 1988; Caldas et al., 1994; Hahn et al., 1996; Scarpa et al., 1993). Additionally, the expression of *CDKN2A* can be abrogated through epigenetic changes, like CGI methylation (Schutte et al., 1997).

Due to the dense and desmoplastic stroma often found in PDACs, the tumor microenvironment (TME) plays a crucial role. Within this environment, cancer-associated fibroblasts (CAFs) have been observed to demonstrate both supportive and inhibitory characteristics for tumor growth (Helms et al., 2020). Unlike other tumor types, PDACs tend to have an anti-inflammatory and immunosuppressive microenvironment. This is marked by increased numbers of immunosuppressive regulatory T-cells and myeloid-derived suppressor cells (Wood et al.,

2022). In contrast, macrophages have been identified to promote neoplastic lesion progression (Lesina et al., 2011).

The primary methods for treating PDAC include surgery and chemotherapy. However, because the tumor is often at an advanced stage when diagnosed, only 15-20% of patients are eligible for surgery (Halbrook et al., 2023). As a result, most patients with advanced PDAC undergo systemic chemotherapy, typically a combination of 5-fluorouracil, leucovorin, irinotecan, and oxaliplatin, known as FOLFIRINOX. A second-line treatment option involves a combination of gemcitabine and nab-paclitaxel (Conroy et al., 2011; Halbrook et al., 2023). Additional treatment approaches, like immune checkpoint inhibitors, have not shown results as promising as in other solid tumors. This might be due to the highly immunosuppressive nature of the TME (Timmer et al., 2021).

The occurrence of PDAC is consistently rising, and the 5-year survival rate remains considerably low compared to other cancers. Hence, there is a critical need for new treatment methods, along with the exploration of early detection biomarkers and personalized treatment approaches (Timmer et al., 2021; Wood et al., 2022).

4. The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands

Bastian Stielow*, Yuqiao Zhou*, Yinghua Cao*, Clara Simon, Hans-Martin Pogoda, Junyi Jiang, Yanpeng Ren, Sabrina Keita Phanor, Iris Rohner, Andrea Nist, Thorsten Stiewe, Matthias Hammerschmidt, Yang Shi, Martha L Bulyk, Zhanxin Wang and Robert Liefke

*These authors contributed equally to this work

4.1 Introduction

The 56 kDa protein SAMD1 (Sterile Alpha Motif Domain-containing protein 1), also known as Atherin, was initially described to play a role in atherosclerosis, where it acts as an LDL-binding protein immobilizing LDL in the arterial wall (Lees et al., 2005). In another study, a microRNA targeting SAMD1 was found to protect against atherosclerosis (Tian et al., 2021).

Conversely, in other investigations, SAMD1 appears in the context of DNA-binding proteins. Specifically, SAMD1 is pulled down by unmethylated DNA (Xiong et al., 2016) as well as by unmethylated nucleosomes (Bartke et al., 2010). Additionally, SAMD1 is enriched in a CpG-rich DNA pulldown followed by mass spectrometry alongside PRC2-related proteins (Perino et al., 2018). Notably, in another experiment, SAMD1 was also enriched when novel CpG motifs were formed, but its presence was diminished when CpG motifs were disrupted (Viturawong et al., 2013). The connection of SAMD1 with CpG motifs is further underscored by a study examining a single nucleotide polymorphism (SNP) associated with a higher risk of atherosclerotic stroke. This SNP results in the destruction of a CpG motif by substitution of adenine for guanine. Intriguingly, SAMD1 was explicitly enriched with the common allele, while its presence was absent with the risk allele (Prestel et al., 2019). Together, these proteomic studies provide strong evidence that SAMD1 might be a CGI-binding protein.

CpG islands (CGIs) are essential genomic elements that play a crucial role in gene regulation, typically existing in an unmethylated state (Deaton & Bird, 2011). CGIs are essential for regulating genes, and understanding how they attract proteins to activate or deactivate specific genes is crucial. By studying unmethylated CGI-binding proteins, we can gain valuable insights into these processes.

Here, we identify SAMD1 as a novel CGI-binding protein interacting with unmethylated CpG-rich DNA via an atypical winged-helix domain. SAMD1 interacts with the histone demethylase KDM1A and the Polycomb-related protein L3MBTL3, resulting in the repression of its target genes. Deletion of SAMD1 in mouse embryonic stem cells leads to dysregulation of a variety of pathways, indicating a vital role of SAMD1 *in vivo*.

4.2   Results

SAMD1 consists of a C-terminal SAM domain and an intrinsically disordered region (IDR) in the center. Interestingly, the N-terminal domain has been predicted to be a winged-helix domain (Figure 1A and Supplementary Figure 1A, B). Firstly, we studied the cellular localization of SAMD1 in mouse embryonic stem cells (mESC). Using immunofluorescence and subcellular fractionation, we found that SAMD1 is located within the nucleus, present in the soluble nuclear fraction, as well as on the chromatin (Figure 1D, E). This prompted us to ask whether SAMD1 might interact directly with DNA. Via electrophoretic mobility shift assays (EMSA), we confirmed that the winged-helix (WH) domain, but not the SAM domain of SAMD1, binds explicitly to CpG-rich DNA. In contrast, no binding could be observed using AT-rich DNA (Figure 1B). To determine the exact binding motif of SAMD1, we used an unbiased protein-binding microarray, revealing that SAMD1 preferentially interacts with a GCGC motif (Figure 1C). This aligns with the results from genome-wide ChIP-sequencing experiments, demonstrating that SAMD1 specifically binds to unmethylated CpG-islands, which show enrichment of a GCGC motif (Figure 1F-I).

The crystal structure of SAMD1-WH, in a complex with CpG-rich DNA, revealed that the WH domain contains two β-strands and three alpha helices interacting with the major as well as with the minor groove of the DNA (Figure 2A). Thereby, SAMD1-WH differs from typical WH domains which usually contain three β-strands (Harami et al., 2013). The amino acids lysine and arginine at positions 45 and 46 are essential for the interaction with the major groove, whereas tyrosine 87 and lysine 88 convey the binding of SAMD1 to the minor groove of the DNA (Figure 2B-E, H). Further EMSA experiments revealed that the binding of SAMD1-WH to CpG-rich DNA is disrupted when the DNA is methylated (Figure 2F) or when either the cytosine or the guanine is exchanged to another base (Figure 2G).

To address the function of SAMD1 at CGIs, we analyzed the genome-wide binding of SAMD1 in more detail. SAMD1 is mainly bound to active promoters marked by H3K4me3 and bivalent promoters decorated with H3K4me3 and H3K27me3. Thereby, the SAMD1-binding pattern has strong similarities with those of the CXXC proteins CXXC1 and KMT2B (Figure 3A, B).

Other CXXC proteins like KDM2A and KDM2B show a broader distribution, present on almost all CGIs (Figure 3A, B). In contrast, the PCL protein MTF2 binds to bivalent and repressed promoters but not to active sites (Figure 3A, B). After SAMD1 KO in mESC, almost 800 genes are significantly dysregulated, with about two-thirds downregulated and one-third upregulated (Figure 3C). Analyzing the dysregulated genes in more detail revealed that 60% of the upregulated genes are direct targets of SAMD1, but only 15% of the downregulated genes are bound by SAMD1 (Figure 3D, E). In line with these findings, we demonstrated that most SAMD1 target genes were upregulated upon SAMD1 KO (Figure 3F). The upregulated genes generally showed higher expression and higher signal of H3K4me3 but lower levels of H3K27me3 compared to the downregulated genes (Figure 3 G, H). Overall, these findings indicate that SAMD1 is mainly bound to the promoters of active genes, where it carries out a repressive function.

To identify interactors of SAMD1 that might lead to this repressive role, we performed mass spectrometry analysis after tandem affinity purification of SAMD1. Here, we identified other SAM domain-containing proteins like L3MBTL3, SFMBT1, and SFMBT2. Moreover, we found several members of the KDM1A complex (Figure 4A). We confirmed the interaction between SAMD1 and L3MBTL3/KDM1A using ectopic co-immunoprecipitation (Co-IP) (Figure 4B). To investigate the interaction between SAMD1 and other SAM domain-containing proteins, we applied a mammalian-two-hybrid approach using the SAM domains only to further confirm the interaction with L3MBTL3. Furthermore, we found an interaction between SAMD1 and itself and L3MBTL4 (Figure 4C). In addition, we verified the interaction between SAMD1 and L3MBTL3 endogenously (Supplementary Figure 5A). To confirm the homopolymerization of SAMD1, we conducted Co-IP experiments, demonstrating that SAMD1 indeed interacts with itself via the SAM domain (Supplementary Figure 5E, F). Next, we addressed which domains of SAMD1 are required for the interaction with L3MBTL3 and KDM1A. Mutating the major groove interaction site of the WH domain did not impact the binding of SAMD1 to L3MBTL3 or KDM1A (Figure 4D-F), suggesting that the DNA binding of SAMD1 is not essential for those interactions. In contrast, deleting the entire WH domain increased the interaction with L3MBTL3 but abolished the interaction with KDM1A (Figure 4D-F). In line with our findings from mammalian-two-hybrid experiments (Figure 4C), we found that mutating or deleting the SAM domain of SAMD1 strongly decreases the interaction with L3MBTL3 (Figure 4D, E). Conversely, these mutants increase the interaction with KDM1A (Figure 4D, F), suggesting that L3MBTL3 and KDM1A interact with SAMD1 via different binding sites in an antagonistic manner.

Next, we performed ChIP-sequencing to investigate how SAMD1, L3MBTL3, and KDM1A might cooperate on CGIs. L3MBTL3 generally binds to fewer promoters compared to SAMD1, but most of its target genes are also co-occupied by SAMD1, and almost all of them are CGIs (Figure 5A). In contrast, KDM1A can be found at ten times more target genes than SAMD1, but at most SAMD1 target genes, KDM1A is also present (Figure 5B). This suggests that SAMD1, L3MBTL3, and KDM1A are often present together at CGIs. Next, we asked whether SAMD1 deletion might affect the recruitment of its interaction partners. On four different SAMD1 target genes, L3MBTL3 binding was reduced to background level upon SAMD1 KO, suggesting that L3MBTL3 recruitment highly depends on SAMD1 (Figure 5C). However, KDM1A binding is only partially reduced, implying that KDM1A is recruited to chromatin by other proteins as well (Figure 5C). Genome-wide, we confirmed a substantial reduction in L3MBTL3 recruitment upon SAMD1 knockout (Figure 5D, F). On the top 2000 CGIs bound by SAMD1, we also observed a significant decrease in KDM1A binding (Figure 5E, F), which is not detectable at the 2000 bottom CGIs (Figure 5E).

To address whether these two interaction partners contribute to the repressive function of SAMD1 on active CGIs, we created L3MBTL3 and KDM1A KO mESC. After KDM1A deletion, the top 100 SAMD1 target genes became predominantly upregulated, whereas no significant difference was observed after L3MBTL3 deletion (Figure 5G). These results suggest that not L3MBTL3 but rather KDM1A is required to convey the repressive role of SAMD1. As KDM1A demethylates H3K4me2, we asked whether this histone mark might be affected after SAMD1 deletion. Indeed, we observed a significant increase in H3K4me2 at the top 2000 GCIs bound by SAMD1 (Figure 5H), suggesting that SAMD1 might be required for the function of the KDM1A complex at those CGIs. H3K4me2 is not only a substrate of KDM1A but is also necessary for histone methyltransferases depositing H3K4me3. In line with this, H3K4me3 levels also increased slightly after SAMD1 deletion (Figure 5I). Together, these data indicate that KDM1A partially contributes to the repressive role of SAMD1 on active CGIs, potentially together with other chromatin-binding proteins.

Next, we examined how SAMD1 binds to chromatin *in vivo*. Therefore, we used SAMD1 KO mES cells to rescue *SAMD1* expression using either wildtype SAMD1 or point mutants of the SAM or/and the WH domain (Figure 6A). As expected, wildtype SAMD1 strongly binds to chromatin when re-induced in SAMD1 KO cells (Figure 6B). A construct containing a mutation disrupting the interaction between SAMD1 and the major groove of the DNA did not rescue SAMD1 binding (Figure 6B). Moreover, mutating the SAM domain, which disrupts the interaction between SAMD1 and L3MBTL3, almost completely abolishes chromatin recruitment of SAMD1 (Figure 6B). To address whether the interaction with L3MBTL3 is

essential for SAMD1 chromatin binding, we performed ChIP experiments in L3MBTL3 KO mES cells. L3MBTL3 deletion only led to a subtle decrease in SAMD1 recruitment (Supplementary Figure 7A), probably due to decreased SAMD1 protein levels after L3MBTL3 KO (Supplementary Figure 7B). We noticed before that SAMD1 can interact with itself via its SAM domain (Figure 4C), hypothesizing that the homopolymerization of SAMD1 might be essential for its DNA binding. Therefore, we solved the crystal structure of two SAMD1-SAM constructs and found that the SAM domain of SAMD1 forms a pentameric ring structure (Figure 6C-H). Gelfiltration revealed that mutating the SAM domain disrupts the formation of the pentameric ring (Supplementary Figure 7D), suggesting that the reduction in DNA binding might be due to the disruption of homopolymerization.

To gain further insights into the biological role of SAMD1, we conducted undirected differentiation of mES wildtype and SAMD1 KO cells for seven days (Figure 7A). The expression of SAMD1 itself remains stable during differentiation (Figure 7A). Also, we observed only subtle changes between wildtype and SAMD1 KO cells regarding classical pluripotency and differentiation markers (Figure 7B). However, SAMD1 target genes become stronger dysregulated during differentiation (Figure 7B), and we also noticed a stronger difference on the transcriptome level between wildtype and SAMD1 KO cells seven days after differentiation using principal component analysis (PCA) (Figure 7C). Compared to our previous experiments (Figure 3C), we found a larger number of genes significantly dysregulated after seven days of differentiation (Figure 7D, F). Consistent with our first experiments, the top 100 SAMD1 target genes become upregulated in the SAMD1 KO cells, also after differentiation (Figure 7E). Using gene set enrichment analysis (GESA), we found that deletion of SAMD1 impacts several biological pathways (Figure 7G). Among the downregulated pathways in SAMD1 KO cells were mainly metabolism-related and immune response gene sets. In contrast, pathways related to neuronal development and heart morphogenesis became upregulated (Figure 7G). These data indicate that SAMD1 has a rather pleiotropic function and regulates several biological pathways.

Overall, this study demonstrates that SAMD1 is a novel CpG-island binding protein. We identified the histone demethylase KDM1A and the Polycomb-related protein L3MBTL3 as the main interactors of SAMD1. SAMD1 exhibits a repressive function on active CGIs, probably due to the recruitment of KDM1A leading to the demethylation of H3K4me2 (Figure 7H). Furthermore, the SAM domain of SAMD1 homopolymerizes in a pentameric ring structure, which seems essential for proper DNA binding (Figure 6). Deletion of SAMD1 leads to dysregulation of various biological pathways (Figure 7), and SAMD1 is commonly upregulated in cancer (Supplementary Figure 8D), indicating a pleiotropic biological function.

4.3   Discussion

Only two known classes of proteins are described to bind to CGIs directly: CXXC domain-containing proteins and Polycomb-like proteins (PCLs). CXXC domain-containing proteins establish interactions with CGIs through a small zinc finger, performing diverse functions that govern active or repressed chromatin states through nucleosome modifications (Shin Voo et al., 2000; C. Xu et al., 2018). Additional members of the CXXC domain-containing protein group can modify DNA itself via DNA methylation (Rhee et al., 2002) or hydroxymethylation (Gu et al., 2011; Ko et al., 2013; Tahiliani et al., 2009). The second group, Polycomb-like proteins, directly engage with CGIs via a winged-helix domain (H. Li et al., 2017). PCLs hold a pivotal role in guiding the PRC2 complex to active CGIs, thereby resulting in suppressed transcription by introducing H3K27me3 (H. Li et al., 2017). Thus, CGIs host both activating and repressive players, contributing significantly to the orchestration of gene regulation (Deaton & Bird, 2011).

Besides CXXC proteins and Polycomb-like proteins, the winged-helix domain of SAMD1 emerged as a third class of CGI-binding domains. This observation was followed by the identification of the winged-helix domains of the histone acetyltransferases KAT6A and KAT6B (Becht et al., 2023; Weber et al., 2023), demonstrating a novel mechanism of how histone acetyltransferases are recruited to chromatin. Together, these three proteins, SAMD1, KAT6A, and KAT6B, represent a common class of proteins that possess a SAMD1-like WH domain (Becht et al., 2023; Weber et al., 2023).

Another domain of SAMD1, the C-terminal SAM domain, plays a pivotal role in two aspects: homopolymerization (Figure 6) and interaction with other SAM domain-containing proteins, such as L3MBTL3 and L3MBTL4 (Figure 4C). In a previous study, the SAM domain of SAMD1 displayed a high score for polymerization (Meruelo & Bowie, 2009) and formed polymers *in vitro* (Knight et al., 2011). SAM domains exhibit versatile molecular interactions, including proteins, RNA, and lipids (Vincenzi et al., 2020). In this study, we confirmed the homopolymerization of SAMD1-SAM into a pentameric ring (Figure 6) and revealed that the SAM-domains of SAMD1 and L3MBTL3 heteropolymerize (Figure 4c).

L3MBTL3, a Polycomb-related protein, and the histone demethylase KDM1A were identified as the major interaction partners of SAMD1 (Figure 4). L3MBTL3 was often enriched together with SAMD1 in studies investigating unmethylated CpG-binding proteins (Bartke et al., 2010; Perino et al., 2018; Prestel et al., 2019), already indicating an association between these two proteins. Notably, an interaction between L3MBTL3 and KDM1A was described before in the context of Notch signaling (Hall et al., 2022; T. Xu et al., 2017). In both studies, L3MBTL3 was

described to interact with the transcription factor RBPJ. If RBPJ binds to L3MBTL3, L3MBTL3 recruits KDM1A, leading to transcriptional repression (T. Xu et al., 2017). Moreover, it was described that L3MBTL3 and KDM1A interact via the SAM domain of L3MBTL3 (T. Xu et al., 2017). Conversely, we found that deleting the SAM domain of SAMD1, which prevents the interaction with L3MBTL3, increases the interaction towards KDM1A (Figure 4D-F). This suggests that the interaction between SAMD1 and KDM1A is not conveyed via L3MBTL3.

The role of SAMD1 in cancer is so far poorly studied. TCGA data reveal that *SAMD1* expression is significantly upregulated in various cancer types (Supplementary Figure 8D). Moreover, high *SAMD1* expression correlates with poor prognosis using pan-cancer data. Specifically in hepatocellular carcinoma and acute myeloid leukemia, patients with high *SAMD1* levels display shorter survival (Supplementary Figure 8E). Interestingly, short tandem repeats within the *SAMD1* gene have been demonstrated to be associated with autism disorders (Annear et al., 2022). Furthermore, the knockout of SAMD1 in mice has been shown to result in embryonic lethality, likely attributed to deficiencies in blood vessel maturation and brain development (Campbell et al., 2023). These findings further underscore a crucial role of SAMD1 at the organismal level as well.

Together, these data demonstrate that SAMD1 has multiple functions and is involved in several biological pathways. The tissue-specific function of SAMD1, as well as the role of SAMD1 in cancer, still remains to be investigated.

4.4   Contribution statement

I made the following contributions to this study:

- Immunofluorescence of SAMD1 in mouse embryonic stem cells (Figure 1D).
- Co-immunoprecipitation between SAMD1 and L3MBTL3/KDM1A (Figure 4B).
- Mapping of the interaction between SAMD1 and L3MBTL3/KDM1A (Figure 4D-F).
- Endogenous co-immunoprecipitation between SAMD1 and L3MBTL3 (Supplementary Figure 5A).
- Mapping of the interaction between SAMD1 and itself (Supplementary Figure 5E, F).

5. The CpG island-binding protein SAMD1 contributes to an unfavorable gene signature in HepG2 hepatocellular carcinoma cells

Clara Simon, Bastian Stielow, Andrea Nist, Iris Rohner, Lisa Marie Weber, Merle Geller, Sabrina Fischer, Thorsten Stiewe and Robert Liefke

5.1   Introduction

Hepatocellular carcinoma (HCC), the prevalent form of liver cancer, exhibits a substantial incidence rate coupled with poor overall survival (International Agency for Research on Cancer, World Health Organization, Cancer Today; Jemal et al., 2017).

The recently described CGI-binding protein SAMD1 has remained largely unexplored in the context of cancer so far. Nonetheless, insights from patient-derived data underscore a vital role of SAMD1 across various tumor types. Particularly in the case of HCC, heightened *SAMD1* expression aligns with an unfavorable prognosis, signifying a potential tumor-promoting function within this specific cancer (Stielow, Zhou, et al., 2021). SAMD1 has been observed to exert gene repression through its interaction with the histone demethylase KDM1A. Notably, KDM1A has been implicated in promoting cell growth (Lv et al., 2022) and driving immunosuppression in HCC (Y. Wang & Cao, 2021). Furthermore, akin to SAMD1, elevated KDM1A levels in HCC patients correlate with diminished survival rates (Y. Wang & Cao, 2021).

In this study, we demonstrate that deleting SAMD1 from HepG2 cells leads to a reduction in both proliferation rates and clonogenicity. Moreover, SAMD1 KO instigates the downregulation of MYC target genes and genes associated with a stem cell-like signature. Our findings reveal that SAMD1 target genes exhibit high expression levels and become further upregulated subsequent to SAMD1 deletion, consistent with our earlier observations in mouse embryonic stem cells (Stielow, Zhou, et al., 2021). Intriguingly, we identified *PIK3IP1*, a repressor of the PI3 kinase, as a downstream target of SAMD1, showing an increased expression upon SAMD1 knockout. Additionally, SAMD1-deficient cells exhibit heightened susceptibility to insulin deprivation, in line with the downregulation of mTORC1 signaling and rapamycin-sensitive genes. Together, our data suggest that SAMD1 deletion fosters a more advantageous transcriptional network within HCC cells and could, therefore, be a promising target for further investigations.

5.2   Results

To investigate the role of SAMD1 in cancer, we analyzed *SAMD1* in distinct tumor entities using patient-derived data, where we found that *SAMD1* is commonly upregulated compared to healthy tissue (Figure 1a). In some cancer types like liver hepatocellular carcinoma (LIHC) and kidney renal clear cell carcinoma (KIRC), high *SAMD1* expression is associated with a worse prognosis. In contrast, in other tumors, like thymoma (THYM), it correlates with a better outcome (Figure 1a). As in hepatocellular carcinoma high *SAMD1* expression strongly correlates with overall survival as well as relapse-free survival (Figure 1b), we decided to focus on the function of SAMD1 in this tumor entity. Moreover, the association of *SAMD1* levels and survival does not seem to depend on sex or risk factors like alcohol abuse or HBV infection (Supplementary Figure 1b), indicating a broad role of SAMD1 in HCC.

SAMD1 expression increases with higher tumor stage, indicating a pro-tumorigenic role (Figure 1c). Notably, *SAMD1* expression is higher in *TP53* mutated samples (Figure 1d), a common mutation in HCC (Schulze et al., 2016). Additionally, *SAMD1* correlates positively with the expression of proliferation genes (Figure 1e), further highlighting a tumor-promoting role. Across different tumor cell lines, *SAMD1* seems to be highly expressed (Figure 1f). To study the function of SAMD1 in HCC, we conducted a SAMD1 KO in HepG2 cells using a CRISPR/Cas9 approach (Figure 1h, i). Similar to our previous studies (Stielow, Zhou, et al., 2021), SAMD1 is present in the nucleus and on the chromatin (Figure 1g, i). SAMD1 deleted HepG2 cells show a slightly impaired proliferation (Figure 1j) but substantially reduced clonogenicity (Figure 1k). Together, these data suggest that deletion of SAMD1 might impair proliferation and clonogenicity of HepG2 cells.

To investigate the genome-wide effects after SAMD1 KO, we performed RNA-seq using three clonally independent clones (Figure 2a). More than 1000 genes were significantly dysregulated, with about one-third downregulated and two-thirds upregulated genes (Figure 2b, c). To determine whether SAMD1 has similar functions in different cell lines, we compared our RNA-seq data from HepG2 cells with those from mESC cells (Stielow, Zhou, et al., 2021). Surprisingly, we did not observe any commonly dysregulated genes, except for *SAMD1* itself, which was downregulated after SAMD1 KO, and *L3MBTL3,* which was upregulated (Figure 2d). L3MBTL3 was previously identified as a main interactor and a target gene of SAMD1 (Stielow, Zhou, et al., 2021) and might, therefore, be regulated via a feedback loop. GSEA revealed that many pathways become dysregulated after SAMD1 deletion in HepG2 cells. Specifically, MYC target genes and genes related to the G2M checkpoint and a stem cell-like signature were downregulated, whereas interferon gamma signaling genes became upregulated (Figure 2e). Comparing our data with publicly available datasets revealed that

after SAMD1 deletion, genes associated with good prognosis in HCC become upregulated, whereas genes associated with poor prognosis become downregulated (Figure 2f). This suggests that HepG2 cells with SAMD1 KO might have a more favorable phenotype.

We observed before, using ChIP-qPCR, that SAMD1 binds to the *NANOS1* but not the *CBLN1* promoter in HepG2 cells (Figure 3a), which scored both among the top target genes of SAMD1 in mESC (Stielow, Zhou, et al., 2021). This indicates that SAMD1 target genes might be cell type-specific. To investigate which promoters are bound by SAMD1 in HepG2 cells, we performed ChIP-seq. Here, more than 80% of all SAMD1 peaks colocalize with CGIs and that most of the peaks are present within the promoter region (Figure 3b), confirming the CGI-binding abilities of SAMD1. Analyzing all CGIs, we noticed that almost 20% are bound by SAMD1 (Figure 3c). Similar to our observations in mESC, SAMD1-bound CGIs are active, showing high levels of H3K4me2 and me3 but low levels of H3K4me1 (Figure 3d). Moreover, they are in an unmethylated state and display high Pol II binding, indicating active transcription (Figure 3d). This aligns with expression analysis, revealing that SAMD1-bound genes are generally expressed higher than unbound genes (Figure 3e). The target genes of SAMD1 are mainly involved in transcriptional regulation and chromatin organization but also include developmental processes (Figure 3f). Next, we aimed to study the effect of SAMD1 on CGIs. We found that the top SAMD1 target genes become upregulated after SAMD1 deletion in HepG2 cells (Figure 3g, h), but SAMD1 is also present at the promoters of downregulated genes (Figure 3i). These data confirm our findings from mESC (Stielow, Zhou, et al., 2021), showing that SAMD1 acts as a repressor on active CGIs also in HepG2 cells. Nevertheless, we demonstrated that target genes of SAMD1 seem to be highly cell type-specific, and many genes dysregulated after SAMD1 KO are probably affected due to indirect mechanisms.

Among the upregulated genes that were also direct SAMD1 targets were *CDKN2A* and *PIK3IP1* (Figure 3h, Figure 4a). *CDKN2A* encodes for the tumor suppressor p16[INK4A], commonly downregulated in HCC (Kaneto, 2001; Maeta et al., 2005). PIK3IP1 acts as a repressor of PI3 kinases (Z. Zhu et al., 2007) and was described to suppress the development of HCC (He et al., 2008). We noticed that rapamycin-sensitive genes, as well as MTORC signaling, become downregulated upon SAMD1 deletion in HepG2 cells (Figure 4b). MTOR signaling is activated by PI3 kinase/AKT signaling and, in turn, affects ribosome biogenesis (Gentilella et al., 2015). We noticed that almost all ribosomal genes were downregulated after SAMD1 KO in HepG2 cells (Figure 4c), indicating dysregulation of these pathways. PI3 kinase/AKT signaling is activated by various growth factors (Vara et al., 2004). Therefore, we asked whether starvation might affect HepG2 SAMD1 KO cells differently from wildtype cells. Notably, the SAMD1 deleted cells were susceptible to serum starvation and showed much

lower viability compared to wildtype cells after 72 h of serum deprivation (Figure 4d, e). Intriguingly, we rescued this phenotype by re-adding insulin to the cell culture medium (Figure 4f, g), indicating the HepG2 SAMD1 KO cells are highly dependent on insulin supplementation. Our data suggest that possibly due to the upregulation of *PIK3IP1*, HepG2 SAMD1 KO cells display decreased PI3 kinase/AKT signaling, resulting in a lower activity of MTOR signaling and impaired ribosome biogenesis (Figure 4h).

To address the function of SAMD1 in the chromatin landscape, we performed ChIP-seq of H3K4me2 (Figure 5a). Here, H3K4me2 generally decreased at promoters, independent of whether the respective genes are bound by SAMD1 (Figure 5b). The global reduction in H3K4me2 was also confirmed by western blot (Supplementary Figure 3a, b). Importantly, the changes in H3K4me2 after SAMD1 KO strongly correlated with the binding pattern of the transcription factors MYC, HES1, and ATF4 (Figure 5c). These three, as well as other transcription factors, were downregulated upon SAMD1 deletion in HepG2 cells (Figure 5d), which might be the reason for the observed reduction in H3K4me2 (Figure 5b, c). Notably, many upregulated genes display an increase in H3K4me2 downstream of the transcription start site (TSS), independent of SAMD1 binding (Figure 5e, f). Together, these findings suggest that deletion of SAMD1 generally affects the chromatin landscape in HepG2 cells. Many of the changes are probably indirect effects due to the dysregulation of many transcription factors after SAMD1 deletion.

In contrast to our observations at promoters, we noticed increased H3K4me2 levels at enhancers (Figure 6a). The alterations at enhancers can be clustered into three distinct groups. Within the first cluster, including more than 50% of all enhancers, we observed only a subtle increase. Conversely, we found a more substantial increase in a second cluster, including about 24,000 enhancers, and a decrease in a third cluster (Figure 6b). Intriguingly, the changes in H3K4me2 at enhancers correlated with expression changes of the respective genes, leading to an upregulation of the second cluster and a downregulation of the third cluster compared to the first cluster (Figure 6c). Motif enrichment analysis of the affected enhancer sites revealed that the HNF (hepatocyte nuclear factor) motif was strongly enriched within the first cluster, in which only slight changes in H3K4me2 were observed (Figure 6d). However, for the second cluster, we noticed an increase in the ISRE motif (IFN-stimulated response element), and in the third cluster, the p53 motif was enriched (Figure 6d). At some of the ISRE motif-containing enhancers, we confirmed the increase in H3K4me2 levels after SAMD1 deletion using ChIP-qPCR (Figure 6e, f).

In summary, our study represents the first investigation of SAMD1 in cancer. Our data confirmed that SAMD1 acts as a CGI-binding protein but revealed that expression changes,

as well as changes in the chromatin landscape, upon SAMD1 deletion, are highly cell-type specific and probably involve a lot of indirect mechanisms. Nevertheless, we demonstrated that SAMD1 KO leads to a more favorable transcriptional network, which might pave the way for novel insights into HCC (Figure 6g, h).

5.3 Discussion

In this study, we have unveiled the prevalent upregulation of the CGI-binding protein *SAMD1* in HCC, a finding that significantly correlates with unfavorable prognosis. Across distinct HCC cohorts, we have consistently observed the overall increase of *SAMD1* expression in tumor tissue compared to adjacent healthy tissue (Supplementary Figure 1a). Notably, the link between *SAMD1* expression and survival remains evident across various HCC subclasses, categorized by factors such as sex, viral infection, or alcohol abuse (Supplementary Figure 1b), implying that *SAMD1* expression could serve as a potential prognostic marker in HCC.

Epigenetic mechanisms play a pivotal role in HCC progression. DNA hypermethylation and histone modification changes have been demonstrated to be induced not only by hepatitis B and C infections but also by alcohol abuse and aflatoxin exposure, all frequent triggers of HCC (Herceg & Paliwal, 2011). For instance, early stages of HCC involve the hypermethylation of the *CDKN2A* gene, encoding for the tumor suppressor p16^{INK4A}, leading to diminished expression (Maeta et al., 2005). Hypermethylation of *CDKN2A* has already been identified in samples from chronic hepatitis and cirrhosis, underscoring its significance in tumorigenesis (Kaneto, 2001). Intriguingly, in HepG2 cells with SAMD1 deletion, *CDKN2A* is significantly upregulated (Figure 3h), further emphasizing a more favorable phenotype after SAMD1 deletion.

Notably, HBV infection-related tumors exhibit heightened methylation levels, suggesting a correlation between hepatitis B virus-induced hypermethylation (Song et al., 2016). The structural proteins of HBV have been shown to induce diverse genetic alterations, modulating methyltransferase activity, which subsequently leads to gene silencing and promotes tumor progression (Zhao et al., 2010). Furthermore, HBx stabilizes a histone H3K4me3 methyltransferase subunit, resulting in heightened transcription of HBV itself (Gao et al., 2020). This strategic utilization of cellular machinery by HBV facilitates viral amplification, potentially leading to chronic inflammation and the development of HCC. Additionally, the HBV structural protein, HBx, remodels the PI3 kinase pathway, fostering tumorigenesis by inhibiting PTEN and activating AKT (Kang-Park et al., 2006; Khattar et al., 2012; G. W. Kim et al., 2021). This pathway is aberrantly activated in half of HCC patients, underscoring its essential role (Sun et

al., 2021). We noticed an upregulation of the SAMD1 target gene *PIK3IP1*, a repressor of PI3 kinase signaling, upon SAMD1 deletion (Figure 3h; Figure 4a). In line with this, we observed a downregulation of ribosomal genes (Figure 4c) as well as an increased sensitivity to insulin deprivation (Figure 4d-g), indicating that the PI3 kinase axis is less active in SAMD1 KO cells.

While the role of SAMD1 in cancer has so far been poorly studied, patient-derived data indicate a robust association between *SAMD1* expression and outcomes across various tumor types (Figure 1a). Notably, we also confirmed some of our previous findings from mESC (Stielow, Zhou, et al., 2021) in the liver cancer cell line HepG2. In both cell types, SAMD1 functions as a repressor of highly active CGIs, characterized by an unmethylated state and decorated with active histone marks (Figure 3b-e), (Stielow, Zhou, et al., 2021). Surprisingly, the gene expression changes observed following SAMD1 deletion diverge significantly between these two cell lines (Figure 2d). Nevertheless, we identified a group of genes to which SAMD1 commonly binds in both cell lines (Supplementary Figure 2a-c). Further analysis of this group reveals a high CG content and notably low DNA methylation in both cell types (Supplementary Figure 2c, d). Intriguingly, genes bound by SAMD1 in HepG2, but not in mESC cells, are characterized by larger CGIs in HepG2 cells, suggesting a critical role of CGI size for SAMD1 binding (Supplementary Figure 2b). Additionally, an increase in H3K4me2 following SAMD1 deletion is evident in mESC (Stielow, Zhou, et al., 2021), whereas a decrease is observed in HepG2 cells (Figure 5b, c). Collectively, these findings underscore the highly tissue-specific functions of SAMD1. Although some SAMD1 target genes overlap in both cell lines, the changes in gene expression upon SAMD1 deletion are notably distinct (Figure 2d), implying the involvement of numerous indirect mechanisms.

One explanation for this phenomenon could be the dysregulation of various transcription factors subsequent to SAMD1 deletion in HepG2 cells (Figure 3h, Figure 5d). Generally, the majority of transcription factors became downregulated after SAMD1 deletion (Figure 5d), but some of them, which are direct target genes of SAMD1, exhibited heightened expression (Figure 3h). One example is IRF1 (Interferon Regulatory Factor 1), a ubiquitously expressed transcription factor responsive to various stimuli (Feng et al., 2021). In the context of HCC, IRF1 has been demonstrated to induce the expression of PD-L1 (Programmed Death Ligand 1), implying pro-tumorigenic features (Yan et al., 2020; Zong et al., 2019). In contrast, other studies suggest a tumor-suppressive role for IRF1, as it activates the expression of chemokines or surface receptors, thereby promoting anti-tumor immunity (X. Li et al., 2023; Yan et al., 2021).

In conclusion, our data demonstrate that deletion of SAMD1 in HepG2 alters the chromatin landscape and leads to dysregulation of many genes. Notably, the role of SAMD1 appears to

be distinctly tissue-specific, with diverse effects upon deletion in mESC and HepG2 cells. The precise role of SAMD1 in cancer necessitates further exploration, however, patient-derived data suggest that SAMD1 could exhibit both pro- and anti-tumorigenic properties depending on the tumor entity (Figure 1a). Consequently, an in-depth examination of SAMD1's role in various cancer types holds the potential for further discoveries.

5.4   Contribution statement

For this study, I performed most of the molecular and biological experiments. The bioinformatics analysis was done by Dr. Robert Liefke. In particular, I performed the following experiments:

- Detection of SAMD1 in different cell lines and subcellular fractionation (Figure 1 f, g).
- Knockout of SAMD1 and validation of knockout in HepG2 cells (Figure 1h, i).
- Biological assays to characterize the SAMD1 KO in HepG2 cells (Figure 1 j, k).
- Preparation of RNA for RNA-seq (Figure 2). Next-generation sequencing was conducted by Dr. Andrea Nist and Prof. Dr. Thorsten Stiewe.
- Analysis of gene expression by RT-qPCR (Figure 4a).
- ChIP-qPCR experiments (Figure 3a; Figure 6f) and ChIP for ChIP-sequencing (Figure 3; 5; 6). Next-generation sequencing was conducted by Dr. Andrea Nist and Prof. Dr. Thorsten Stiewe.
- Analysis of histone marks by western blot and quantification (Supplementary Figure 3)
- Investigation of the insulin-dependency of SAMD1 KO cells (Figure 4d-g).

6. SAMD1 suppresses epithelial-mesenchymal transition (EMT) pathways in pancreatic ductal adenocarcinoma

Clara Simon, Inka D. Brunke, Bastian Stielow, Ignasi Forné, Anna Mary Steitz, Merle Geller, Iris Rohner, Lisa M. Weber, Sabrina Fischer, Lea Marie Jeude, Andrea Nist, Thorsten Stiewe, Magdalena Huber, Malte Buchholz and Robert Liefke

6.1 Introduction

Pancreatic cancer has emerged as a very lethal type of cancer, exhibiting a 5-year survival rate of merely 10% (International Agency for Research on Cancer, World Health Organization, Cancer Today; Siegel et al., 2021). Pancreatic ductal adenocarcinoma (PDAC) is the most prevalent and lethal subtype of pancreatic cancer (Park et al., 2021). The role of the CGI-binding protein SAMD1 in PDAC has not been investigated so far. Intriguingly, data derived from patients hint at the potential benefits of heightened *SAMD1* expression for overall survival in PDAC (Simon et al., 2022). This contradicts observations in hepatocellular carcinoma, where SAMD1 was linked to more oncogenic attributes (Simon et al., 2022).

Compared to many other tumor types, PDAC presents a relatively modest mutation burden (H.-X. Wu et al., 2019), indicating a substantial influence of additional processes, such as epigenetics, on tumorigenesis (Pandey et al., 2023). Notably, proteins associated with the SWI/SNF chromatin remodeling complex frequently exhibit mutations within PDAC samples (Waddell et al., 2015). Moreover, chromatin-modifying enzymes like KMT2D (Lysine Methyltransferase 2D) and KDM6A (Lysine Demethylase 6A) are also affected (Raphael et al., 2017; Waddell et al., 2015). Furthermore, recent studies revealed that epigenetic reprogramming plays a crucial role in driving PDAC metastasis (McDonald et al., 2017).

Metastasis within PDAC is orchestrated through a phenomenon known as epithelial-mesenchymal transition (EMT). In this intricate process, cancer cells gradually decrease the expression of epithelial markers while concurrently acquiring the expression of mesenchymal markers (Pastushenko & Blanpain, 2019). Notably, EMT transcription factors are essential in suppressing epithelial gene expression and activating mesenchymal gene expression (S. Wang et al., 2017). Investigations of PDAC samples from patients have underscored a correlation between tumors displaying a mesenchymal phenotype and an increased propensity to form metastases, ultimately contributing to a poorer prognosis (Rasheed et al., 2010).

Moreover, tumors exhibiting mesenchymal characteristics have demonstrated heightened resistance to gemcitabine, a prevailing therapeutic agent for PDAC (Shah et al., 2007).

This study underscores the role of the CGI-binding protein SAMD1 as a suppressor of EMT-associated genes. Analysis of patient-derived data reveals a significant correlation between elevated *SAMD1* expression in PDAC and a better prognosis, coinciding with reduced expression of EMT-associated genes. Moreover, SAMD1 deleted cells display higher migration rates, and we identified the mesenchymal gene *CDH2*, encoding for N-cadherin, as the major driver of migration after SAMD1 KO. Similar to our previous results from mESC (Stielow, Zhou, et al., 2021), SAMD1 cooperates with the histone demethylase KDM1A in PDAC cells, resulting in the repression of its target genes. Intriguingly, our research unveils the FBXO11 complex as a novel regulator, influencing the chromatin binding of SAMD1. Thereby, our data provide insights into a novel interplay involving SAMD1, KDM1A, and FBXO11 and establish SAMD1 as a pivotal regulator of EMT-related genes.

6.2 Results

So far, no studies have been performed addressing the role of the CGI-binding protein SAMD1 in pancreatic ductal adenocarcinoma. However, analysis of expression data revealed that *SAMD1* is significantly upregulated in PDAC cells compared to healthy tissue (Supplementary Figure 1a). Conversely, survival data suggest elevated SAMD1 levels correlate with better prognosis (Figure 1a). We noticed that patients with high *SAMD1* expression show a reduced expression of genes related to epithelial-mesenchymal transition (Figure 1b). Notably, this phenomenon extends beyond PDAC, manifesting in other tumor types where elevated *SAMD1* levels are also associated with a favorable patient outcome (Supplementary Figure 1c, d). In contrast, in tumors in which high *SAMD1* expression is associated with poor prognosis, we noticed that patients with high *SAMD1* levels also display an increased expression of EMT-related genes (Supplementary Figure 1b, d).

These observations suggest that SAMD1 might regulate EMT-related genes. To explore this hypothesis, we employed CRISPR/Cas9 to conduct a SAMD1 knockout in the PDAC cell line PaTu8988t (Figure 1c, d). SAMD1 deletion did not impact proliferation rates (Figure 1e) but resulted in a distinct cellular morphology. Notably, the SAMD1 knockout cells exhibited a more elongated, mesenchymal-like appearance compared to wildtype cells (Figure 1f, g). To assess whether this altered cell shape corresponds with migratory behavior, we performed wound healing assays, unveiling a notable augmentation in migration rates after SAMD1 deletion (Figure h, i). This finding was consistently validated through transwell migration assays

(Supplementary Figure 2a, b) and unbiased time-lapse analysis (Supplementary Figure 2c). Importantly, we demonstrated a comparable phenotype in another cell line, BxPC3, which similarly displayed unaltered proliferation but heightened migratory capacities upon SAMD1 deletion (Supplementary Figure 2d-g). Next, we aimed to confirm that the observed increased migration is indeed a consequence of SAMD1 deletion. Therefore, we applied a system where nuclear localization of SAMD1 can be induced via 4-hydroxytamoxifen (4-OHT) (Supplementary Figure 3a, b). Importantly, inducing SAMD1's nuclear localization led to a reduction in migration rates (Figure 1j, k).

As a next step, we aimed to study the genome-wide consequences after SAMD1 deletion in PaTu8988t cells. PCA analysis revealed that the SAMD1 KO clones are highly distinct from cells transfected with a control sgRNA (Supplementary Figure 4a). Moreover, about 800 genes are significantly dysregulated upon SAMD1 loss, most of them upregulated (Figure 2a; Supplementary Figure 4b). Intriguingly, several cancer-associated pathways were upregulated, including the EMT pathway (Figure 2b, c), confirming our initial observation (Figure 1b). Our previous investigations revealed that SAMD1 acts as a repressive CGI-binding protein in mESC and HepG2 cells (Simon et al., 2022; Stielow, Zhou, et al., 2021). Therefore, we performed ChIP-seq to decipher the function of SAMD1 on chromatin in PDAC. Consistent with our former findings (Simon et al., 2022; Stielow, Zhou, et al., 2021), SAMD1 peaks mostly overlap with CGIs, and the top SAMD1 target genes become upregulated after SAMD1 deletion (Figure 2d, e), suggesting that SAMD1 acts as a repressor on CGIs in PDAC. Gene ontology analysis of SAMD1 target genes revealed that they are primarily involved in regulating gene expression and chromatin organization (Supplementary Figure 4c).

CDH2, a direct target gene of SAMD1 encoding for N-cadherin, emerged as one of the most upregulated genes after SAMD1 deletion (Figure 2f-h). N-cadherin is a well-known driver of EMT and is involved in establishing a mesenchymal phenotype (Mrozik et al., 2018). This prompted us to study whether N-cadherin might be responsible for the enhanced migration rates we observed after SAMD1 KO. To address this question, we established a double KO of SAMD1 and CDH2. SAMD1 and CDH2 double KO clones show a less elongated cell shape (Supplementary Figure 5a, b) and migration rates more akin to control transfected cells (Figure 2i, j), suggesting that N-cadherin is the main driver of migration post SAMD1 deletion. Intriguingly, ADH-1, an inhibitor of N-cadherin-mediated cell adhesion, emerged as an effective suppressor of the mesenchymal phenotype that manifests upon SAMD1 deletion (Supplementary Figure 5c, d). We made use of our inducible SAMD1 rescue system to address whether re-induction of SAMD1 lowers *CDH2* expression. Indeed, we can rescue the binding of SAMD1 to the *CDH2* promoter (Figure 2l) and decrease the expression of *CDH2* upon

nuclear re-induction of SAMD1 (Figure 2k). Notably, the reduction in *CDH2* expression was absent when we employed a DNA-binding deficient mutant (WHmut) of SAMD1 (Figure 2k). Similar results were obtained for *L3MBTL3*, a common target gene of SAMD1 in different cell lines (Supplementary Figure 3c, d). Importantly, the binding of SAMD1 to the *L3MBTL3* and *CDH2* promoters, as well as the upregulation of the respective genes upon SAMD1 deletion, was also confirmed in BxPC3 cells (Supplementary Figure 2i, h). In conclusion, SAMD1 acts as a repressor of *CDH2* in PDAC, and N-cadherin seems to be an essential driver of migration.

Recently, we identified the SAM domain-containing protein L3MBTL3 and the histone demethylase KDM1A as interaction partners of SAMD1 (Stielow, Zhou, et al., 2021). To decipher whether they might be associated with SAMD1 in PDAC, we performed ChIP-qPCR experiments subsequent to SAMD1 deletion in PaTu8988t cells. In line with our previous findings, we noticed that L3MBTL3 chromatin association is highly dependent on SAMD1, and upon loss of SAMD1, L3MBTL3 is no longer recruited to SAMD1 target genes (Figure 3a). KDM1A recruitment to the *CDH2* promoter is strongly reduced but still detectable in the SAMD1 KO cells (Figure 3a). Intriguingly, the recruitment of both proteins can be rescued when re-inducing SAMD1's nuclear localization (Figure 3a). In line with decreased KDM1A recruitment upon SAMD1 KO, we observed increased levels of H3K4me2 and me3 (Figure 3b). Next, we aimed to address the interaction between SAMD1 and KDM1A in PDAC in more detail. Firstly, we confirmed the interaction on endogenous level using Co-IP (Figure 3c). In ectopic Co-IP experiments, deletion of the DNA-binding WH domain abolished the interaction with KDM1A, while removal of the SAM domain conversely increased this interaction (Figure 3e). Notably, the deletion of the SAM domain not only fostered the interaction with KDM1A but also with other components of the complex (Figure 3d; Supplementary Figure 6a).

To investigate which part of KDM1A conveys the interaction with SAMD1, we performed further mapping experiments, indicating that the N-terminal region of KDM1A's amino oxidase domain is required for this interaction (Figure 3f). A KDM1A inhibitor, ORY-1001, was described to bind to the FAD-dependent amino oxidase domain (Fang et al., 2019). We tested whether this inhibitor also prevents the interaction with SAMD1. Notably, ORY-1001 strongly attenuates the interaction between SAMD1 and KDM1A (Supplementary Figure 6b) while leaving the interaction with other KDM1A complex members unaffected (Supplementary Figure 6c, d).

KDM1A is known to demethylate H3K4me2 using its amino oxidase domain (Shi et al., 2004). This prompted us to investigate whether the binding of SAMD1 to this domain might modulate the activity of KDM1A. To address this, we immunoprecipitated KDM1A from either WT or SAMD1 KO cells and subsequently performed a demethylase reaction using calf histones as a substrate. Significantly, KDM1A immunoprecipitated from wildtype cells exhibited robust

histone demethylase activity, whereas we could not detect demethylase activity using KDM1A derived from SAMD1-deleted cells (Figure 3g, h). Together, these results suggests that SAMD1 cooperates with KDM1A also in PDAC cells to demethylate H3K4me2, leading to the suppression of its target genes.

Intriguingly, fractionation experiments in PDAC cell lines revealed remarkably decreased chromatin binding of SAMD1 in contrast to other cell lines (Figure 4a). To address this phenomenon comprehensively, we performed immunoprecipitation of SAMD1, followed by liquid chromatography–mass spectrometry (LC-MS) to analyze new potential interaction partners of SAMD1. Significantly, we were able to confirm previously identified interaction partners of SAMD1 (Stielow, Zhou, et al., 2021): The SAM domain-containing proteins L3MBTL3 and L3MBTL4, as well as the KDM1A-complex (Figure 4b). Moreover, we identified the FBXO11-complex as a novel potential interactor of SAMD1, which has not been reported before. This complex consists of FBXO11 itself, SKP1, CUL1, and RBX1 (Duan et al., 2012; P. Xu et al., 2021), each observed to be enriched alongside SAMD1 (Figure 4b, c). Additionally, the protein NEDD8 was detected (Figure 4b, c), a small protein bound to the FBXO11 complex and pivotal for its activity (Zheng et al., 2002).

We confirmed the interaction between SAMD1 and FBXO11 by co-immunoprecipitation (Figure 4d) and found that SAMD1 and FBXO11 probably interact via multiple sites within SAMD1's N-terminus (Figure 4e). Given that SAMD1 interacts with KDM1A via its N-terminal region (Figure 3e), we asked whether FBXO11 might be associated with KDM1A, as well. Similarly, we found that FBXO11 interacts with the N-terminal amino oxidase domain of KDM1A (Figure 4f), and this interaction was abrogated upon treatment with the KDM1A inhibitor ORY-1001 (Supplementary Figure 6e). The FBXO11 complex is known to ubiquitinate its targets, marking them for degradation (Abbas et al., 2013; Duan et al., 2012; Jin et al., 2015; P. Xu et al., 2021). Indeed, ubiquitination of SAMD1 was demonstrated by treating cells stably expressing HA-tagged ubiquitin with the proteasome inhibitor MG-132 (Figure 4g). Surprisingly, a cycloheximide chase did not show any evident turnover of SAMD1 within six hours, indicating that SAMD1 protein levels are relatively stable (Figure 4h). To address which domain of SAMD1 is ubiquitinated, we used different SAMD1 constructs, revealing that the winged-helix domain of SAMD1 is the primary site of ubiquitination (Figure 4i).

To address the impact of FBXO11 on SAMD1, we conducted an FBXO11 KO in HEK293 and PaTu8988t cells (Figure 5a). The removal of FBXO11 resulted in strongly decreased ubiquitination of SAMD1 (Figure 5b, c), indicating that FBXO11 is an essential regulator of SAMD1. SAMD1 protein levels remain unaltered in FBXO11 KO cells (Figure 5d), suggesting that the ubiquitination of SAMD1 by FBXO11 does not ultimately lead to degradation.

Therefore, we performed fractionation experiments to address whether FBXO11 might impact the cellular localization of SAMD1. Whereas SAMD1 is generally present at low levels on chromatin in PDAC cell lines (Figure 4a), this pattern changes upon FBXO11 KO. In FBXO11 deleted PaTu8988t cells, we noticed that approximately half of SAMD1 is associated with chromatin (Figure 5e), indicating that FBXO11 might modulate the chromatin binding of SAMD1. The PhoshoSitePlus database (Hornbeck et al., 2012) predicts two ubiquitination sites within the WH domain of SAMD1, K75, and K88 (Figure 5f). Using mutants of these amino acids in ubiquitination assays, we noticed a decreased ubiquitination of the K88A mutant (Figure 5g, h), indicating that this site might be ubiquitinated.

In a previous study, we demonstrated that the lysine at position 88 is required for the interaction of SAMD1 with the minor groove of the DNA, and mutating this amino acid slightly reduces DNA binding of SAMD1 (Stielow, Zhou, et al., 2021). This prompted us to ask whether the deletion of FBXO11 affects the DNA binding capabilities of SAMD1. Using ChIP-qPCR, we observed augmented binding of SAMD1 to the *CDH2* promoter upon FBXO11 KO (Figure 6a). Via ChIP-seq, we confirmed this increase genome-wide (Figure 6b; Supplementary Figure 7a). Intriguingly, most promoters show elevated levels of SAMD1 after FBXO11 KO. We identified four groups, each characterized by distinct alterations in SAMD1 recruitment (Figure 6c). Within the first group, SAMD1 was only recruited at low levels in wildtype cells but showed strong enrichment upon FBXO11 KO. In the second group, moderate binding of SAMD1 in wild-type cells and a modest increase after FBXO11 deletion was observed. The third group includes promoters that were already robustly bound by SAMD1. Here, almost no differences were observed due to FBXO11 KO. Conversely, the fourth group comprised only a few promoters that exhibited reduced SAMD1 recruitment subsequent to FBXO11 deletion, an exception to the genome-wide trend (Figure 6c, d; Supplementary Figure 7b).

Notably, the first two groups, which showed heightened SAMD1 recruitment after FBXO11 KO, displayed elevated levels of H3K4me3 and Pol II (Supplementary Figure 7d, e), signifying their heightened transcriptional activity. Moreover, they harbored smaller CGIs compared to groups 3 and 4, suggesting that the effect of FBXO11 deletion on SAMD1 recruitment is most pronounced at active promoters with smaller CGIs (Supplementary Figure 7c). Of note, these distinctions in SAMD1 recruitment did not translate into changes in gene expression levels across the four groups (Supplementary Figure 7f). Next, we asked how the increased chromatin binding of SAMD1 might affect the expression of target genes. Most genes analyzed via qPCR were unaffected by the FBXO11 knockout (Figure 6e). However, a slight reduction of *MMP15* and a substantial reduction of *MTSS1* was detected, both among the top target genes of SAMD1 (Figure 6e).

To address the impact of FBXO11 in PDAC in-depth, we analyzed patient-derived data. Here, we found that *FBXO11* is commonly upregulated, and high expression correlates with unfavorable prognosis in PDAC (Figure 6f, g). Importantly, high *FBXO11* levels correlate positively with the expression of EMT-related genes (Figure 6h), which is the opposite of what we observed for *SAMD1* (Figure 1b).

In summary, our data demonstrate that SAMD1 emerges as a repressor of EMT-related genes in the context of PDAC (Figure 6i). Upon SAMD1 deletion, these genes become upregulated, leading to a more mesenchymal phenotype and enhanced migration. We identified *CDH2*, encoding for N-cadherin, as the primary driver of migration after SAMD1 KO. Intriguingly, we were able to rescue the observed phenotype when re-inducing the nuclear localization of SAMD1. Moreover, the FBXO11 complex is a novel interactor of SAMD1 and modulates the chromatin binding of SAMD1 via ubiquitination of the WH domain (Figure 6j).

6.3   Discussion

This study presents a novel exploration of the interaction between the CGI-binding protein SAMD1, the histone demethylase KDM1A, and FBXO11, an E3 ubiquitin ligase complex member. The interplay between these three proteins orchestrates the expression of EMT-related genes in PDAC. FBXO11 regulates SAMD1's chromatin binding via ubiquitination of the WH domain, and SAMD1 recruits KDM1A to its target promoters, leading to demethylation and subsequent repression of the respective genes (Figure 6i, j).

In our prior experiments (Stielow, Zhou, et al., 2021), we observed that deletion or mutation of the SAM domain of SAMD1 substantially increases its interaction with KDM1A. In this study, we not only confirm this phenomenon but also demonstrate that deletion of SAMD1's SAM domain results in more robust interactions with other KDM1A complex members (Supplementary Figure 6a). As previously reported, the SAM domain is crucial not only for L3MBTL3 interaction but also for SAMD1's homopolymerization (Stielow, Zhou, et al., 2021). In contrast, deleting the WH domain of SAMD1 strengthens the interaction with L3MBTL3 (Stielow, Zhou, et al., 2021), suggesting diverse SAMD1 interactions with L3MBTL3 and KDM1A in differing contexts.

Interestingly, treating cells with the KDM1A inhibitor, ORY-1001, reduces the interaction between SAMD1 and KDM1A (Supplementary Figure 6b) but not with other members of the KDM1A complex (Supplementary Figure 6c, d). ORY-1001 binds to the amino oxidase domain of KDM1A, thereby inhibiting its catalytic activity, and has already been implicated in clinical

trials (Fang et al., 2019; Maes et al., 2018). A recent structural analysis of KDM1A, together with its partner RCOR1 (REST Corepressor 1), revealed their interaction on nucleosomes. KDM1A binds RCOR1 via its tower domain, necessitating RCOR1 for optimal nucleosome binding (S.-A. Kim et al., 2020). Our data suggest that SAMD1 also contributes to recruiting KDM1A to chromatin (Figure 3a) and enhances KDM1A's catalytic activity by binding to its amino oxidase domain (Figure 3f-h).

The role of KDM1A in pancreatic ductal adenocarcinoma is poorly studied. In one study, the knockdown of KDM1A reduced proliferation and tumorigenicity, and KDM1A synergized with HIF1α (Qin et al., 2014). Another study demonstrated an oncogenic function of KDM1A by interacting with the long non-coding RNA HOXA-AS2 and the histone methyltransferase EZH2 (Lian et al., 2017). In contrast, our data show that KDM1A is tethered by SAMD1 to the promoters of EMT-related genes, for instance, *CDH2* (Figure 3a), to repress transcription. Nonetheless, our previous experiments have demonstrated that KDM1A binds to more promoters than SAMD1 (Stielow, Zhou, et al., 2021), suggesting that KDM1A exerts many functions independent of SAMD1.

FBXO11, a novel interactor of SAMD1, also interacts with KDM1A, probably via association with SAMD1. Both SAMD1 and FBXO11 bind to KDM1A's amino oxidase domain, and this interaction is disrupted by ORY-1001 treatment (Figure 3f; Figure 4f; Supplementary Figure 6e). We noticed strongly increased chromatin binding of SAMD1 after FBXO11 deletion, but only subtle effects on gene expression. This contradicts our rescue experiments, where we found that increased SAMD1 chromatin binding in wildtype cells results in the repression of the respective genes at the *L3MBTL3* and *CDH2* promoters (Figure 2k, l; Supplementary Figure 3c, d). However, the rescue experiments were only performed for 24 hours, unlike the several-week process of generating FBXO11 KO clones. Cells might adapt to increased SAMD1 recruitment, evading its repressive functions. As our data suggest that SAMD1 acts as a repressor of tumor-promoting genes, preventing SAMD1 from exerting its tumor-repressive function might be essential for the survival of tumor cells.

FBXO11 was previously known for its role in protein degradation through ubiquitination (Abbas et al., 2013; Duan et al., 2012; Jin et al., 2015). Surprisingly, deleting FBXO11 does not alter SAMD1 protein levels (Figure 5d). Instead, FBXO11 seems to modulate chromatin binding via ubiquitination of SAMD1's WH domain (Figure 5f, g; Figure 6a, b). For other DNA-binding proteins like p53, ubiquitination has been described to enhance DNA association and transcription factor activity (Landré et al., 2017). Conversely, our experiments demonstrate that FBXO11 negatively regulates the DNA binding of SAMD1.

In conclusion, our investigation into the SAMD1-FBXO11 axis elucidates the diminished DNA binding of SAMD1 in PDAC. We demonstrate that SAMD1 is a repressor of EMT-related genes, and high *SAMD1* expression correlates with a favorable prognosis. Remarkably, *CDH2* emerges as a pivotal downstream target of SAMD1, positively influencing migration. This study advances our comprehension of EMT-related gene regulation in PDAC and may pave the way for further insights into SAMD1's functional role.

6.4   Contribution statement

For this study, I performed most of the molecular and biological experiments. Together with Dr. Robert Liefke, I planned and designed the experimental setup and prepared the manuscript. The bioinformatics analysis was done by Dr. Robert Liefke. In particular, I performed the following experiments:

- Creation and validation of a SAMD1 KO in PaTu8988t and BxPC3 cells (Figure 1c, d and Supplementary Figure 2d) and analysis of the proliferation rates (Figure 1e and Supplementary Figure 2e).

- Analysis of the migration phenotype observed in PaTu8988t and BxPC3 SAMD1 KO cells (Figure 1h-k and Supplementary Figure 2c, f, g), except for the transwell migration assay shown in Supplementary Figure 2a, b, which was performed by Anna Mary Steitz.

- Quantification of the cell shape of PaTu8988t cells under several conditions (Figure 1f, g and Supplementary Figure 5).

- Creation and validation of PaTu8988t cells with SAMD1 rescue (Supplementary Figure 3a, b).

- RT-qPCR and ChIP-qPCR in BxPC3 and PaTu8988t cells under several conditions were performed with the support of Dr. Bastian Stielow and Inka Brunke (Supplementary Figure 2h, I; Supplementary Figure 3c, d; Figure 2k, l, Figure 3a, b, Figure 6a, e).

- KO of CDH2 in PaTu8988t cells and characterization of the CDH2-related phenotype (Figure 2f-j).

- Preparation of ChIP-DNA and RNA for next-generation sequencing (Supplementary Figure 4; Figure 2a-g; Figure 6b-d; Supplementary Figure 7). Library preparation was performed by Dr. Bastian Stielow. Next-generation sequencing was conducted by Dr. Andrea Nist and Prof. Dr. Thorsten Stiewe.

- Creation and validation of a FBXO11 KO in PaTu8988t and HEK293 cells was performed together with Inka Brunke (Figure 5a).
- Histone demethylase assay and creation of the required cell lines (Figure 3g, h).
- Fractionation assay of SAMD1 in different cell lines (Figure 4a).
- Cycloheximide chase (Figure 4h).
- Preparation of samples for mass spectrometry, analysis was performed by Dr. Ignasi Forné (Figure 4b).
- All Co-IP and Ubiquitination assays were performed with support of Inka Brunke, Merle Geller, and Lea Marie Jeude (Figure 3c-f, Figure 4d-i, Figure 5b-h, Supplementary Figure 6).

7. References

Abbas, T., Mueller, A. C., Shibata, E., Keaton, M., Rossi, M., & Dutta, A. (2013). CRL1-FBXO11 Promotes Cdt2 Ubiquitylation and Degradation and Regulates Pr-Set7/Set8-Mediated Cellular Migration. *Molecular Cell*, *49*(6), 1147–1158. https://doi.org/10.1016/j.molcel.2013.02.003

Almoguera, C., Shibata, D., Forrester, K., Martin, J., Arnheim, N., & Perucho, M. (1988). Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. *Cell*, *53*(4), 549–554. https://doi.org/10.1016/0092-8674(88)90571-5

Annear, D. J., Vandeweyer, G., Sanchis-Juan, A., Raymond, F. L., & Kooy, R. F. (2022). Non-Mendelian inheritance patterns and extreme deviation rates of CGG repeats in autism. *Genome Research*, *32*(11–12), 1967–1980. https://doi.org/10.1101/gr.277011.122

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M., & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, *143*(3), 470–484. https://doi.org/10.1016/J.CELL.2010.10.012

Becht, D. C., Klein, B. J., Kanai, A., Jang, S. M., Cox, K. L., Zhou, B.-R., Phanor, S. K., Zhang, Y., Chen, R.-W., Ebmeier, C. C., Lachance, C., Galloy, M., Fradet-Turcotte, A., Bulyk, M. L., Bai, Y., Poirier, M. G., Côté, J., Yokoyama, A., & Kutateladze, T. G. (2023). MORF and MOZ acetyltransferases target unmethylated CpG islands through the winged helix domain. *Nature Communications*, *14*(1), 697. https://doi.org/10.1038/s41467-023-36368-5

Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, *405*(6785), 482–485. https://doi.org/10.1038/35013100

Bird, A. P. (1987). CpG islands as gene markers in the vertebrate nucleus. *Trends in Genetics*, *3*, 342–347. https://doi.org/10.1016/0168-9525(87)90294-0

Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J., & Klose, R. J. (2010). CpG Islands Recruit a Histone H3 Lysine 36 Demethylase. *Molecular Cell*, *38*(2), 179–190. https://doi.org/10.1016/j.molcel.2010.04.009

Bressac, B., Kew, M., Wands, J., & Ozturk, M. (1991). Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature*, *350*(6317), 429–431. https://doi.org/10.1038/350429A0

Caldas, C., Hahn, S. A., Da-Costa, L. T., Redston, M. S., Schutte, M., Seymour, A. B., Weinstein, C. L., Hruban, R. H., Yeo, C. J., & Kern, S. E. (1994). Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nature Genetics*, *8*(1), 27–32. https://doi.org/10.1038/NG0994-27

Campbell, B., Weber, L. M., Engle, S. J., Ozolinš, T. R. S., Bourassa, P., Aiello, R., & Liefke, R. (2023). Investigation of SAMD1 ablation in mice. *Scientific Reports*, *13*(1). https://doi.org/10.1038/S41598-023-29779-3

Ciccone, D. N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G., Li, E., & Chen, T. (2009). KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature*, *461*(7262), 415–418. https://doi.org/10.1038/NATURE08315

Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., de la Fouchardière, C., Bennouna, J., Bachet, J.-B., Khemissa-Akouz, F., Péré-Vergé, D., Delbaldo, C., Assenat, E., Chauffert, B., Michel, P., Montoto-Grillot, C., & Ducreux, M. (2011). FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. *New England Journal of Medicine*, *364*(19), 1817–1825. https://doi.org/10.1056/NEJMoa1011923

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes & Development*, *25*(10), 1010–1022. https://doi.org/10.1101/gad.2037511

de Napoles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., Nesterova, T. B., Silva, J., Otte, A. P., Vidal, M., Koseki, H., & Brockdorff, N. (2004). Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Developmental Cell*, *7*(5), 663–676. https://doi.org/10.1016/j.devcel.2004.10.005

Duan, S., Cermak, L., Pagan, J. K., Rossi, M., Martinengo, C., di Celle, P. F., Chapuy, B., Shipp, M., Chiarle, R., & Pagano, M. (2012). FBXO11 targets BCL6 for degradation and is inactivated in diffuse large B-cell lymphomas. *Nature*, *481*(7379), 90–93. https://doi.org/10.1038/nature10688

Fang, Y., Liao, G., & Yu, B. (2019). LSD1/KDM1A inhibitors in clinical trials: advances and prospects. *Journal of Hematology & Oncology*, *12*(1), 129. https://doi.org/10.1186/s13045-019-0811-9

Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W., Koseki, H., Brockdorff, N., Ponting, C. P., Kessler, B. M., & Klose, R. J. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *ELife*, *1*. https://doi.org/10.7554/eLife.00205

Feng, H., Zhang, Y.-B., Gui, J.-F., Lemon, S. M., & Yamane, D. (2021). Interferon regulatory factor 1 (IRF1) and anti-pathogen innate immune responses. *PLOS Pathogens*, *17*(1), e1009220. https://doi.org/10.1371/journal.ppat.1009220

Fischer, S., & Liefke, R. (2023). Polycomb-like Proteins in Gene Regulation and Cancer. *Genes*, *14*(4). https://doi.org/10.3390/genes14040938

Fischer, S., Weber, L. M., & Liefke, R. (2022). Evolutionary adaptation of the Polycomb repressive complex 2. *Epigenetics & Chromatin*, *15*(1), 7. https://doi.org/10.1186/s13072-022-00439-6

Gahurova, L., Tomizawa, S. ichi, Smallwood, S. A., Stewart-Morgan, K. R., Saadeh, H., Kim, J., Andrews, S. R., Chen, T., & Kelsey, G. (2017). Transcription and chromatin determinants of de novo DNA methylation timing in oocytes. *Epigenetics & Chromatin*, *10*(1). https://doi.org/10.1186/S13072-017-0133-5

Gao, W., Jia, Z., Tian, Y., Yang, P., Sun, H., Wang, C., Ding, Y., Zhang, M., Zhang, Y., Yang, D., Tian, Z., Zhou, J., Ruan, Z., Wu, Y., & Ni, B. (2020). HBx Protein Contributes to Liver Carcinogenesis by H3K4me3 Modification Through Stabilizing WD Repeat Domain 5 Protein. *Hepatology*, *71*(5), 1678–1695. https://doi.org/10.1002/hep.30947

Gentilella, A., Kozma, S. C., & Thomas, G. (2015). A liaison between mTOR signaling, ribosome biogenesis and cancer. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1849*(7), 812–820. https://doi.org/10.1016/j.bbagrm.2015.02.005

Greenfield, R., Tabib, A., Keshet, I., Moss, J., Sabag, O., Goren, A., & Cedar, H. (2018). Role of transcription complexes in the formation of the basal methylation pattern in early development. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(41), 10387–10391. https://doi.org/10.1073/PNAS.1804755115

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., & Young, R. A. (2007). A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. *Cell*, *130*(1), 77. https://doi.org/10.1016/J.CELL.2007.05.042

Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J., & Xu, G. L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, *477*(7366), 606–612. https://doi.org/10.1038/NATURE10443

Hahn, S. A., Schutte, M., Shamsul Hoque, A. T. M., Moskaluk, C. A., Da Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H., & Kern, S. E. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science (New York, N.Y.)*, *271*(5247), 350–353. https://doi.org/10.1126/SCIENCE.271.5247.350

Halbrook, C. J., Lyssiotis, C. A., Pasca di Magliano, M., & Maitra, A. (2023). Pancreatic cancer: Advances and challenges. *Cell*, *186*(8), 1729–1754. https://doi.org/10.1016/j.cell.2023.02.014

Hall, D., Giaimo, B. D., Park, S.-S., Hemmer, W., Friedrich, T., Ferrante, F., Bartkuhn, M., Yuan, Z., Oswald, F., Borggrefe, T., Rual, J.-F., & Kovall, R. A. (2022). The structure, binding and function of a Notch transcription complex involving RBPJ and the epigenetic reader protein L3MBTL3. *Nucleic Acids Research*, *50*(22), 13083–13099. https://doi.org/10.1093/nar/gkac1137

Hao, X., Sun, G., Zhang, Y., Kong, X., Rong, D., Song, J., Tang, W., & Wang, X. (2021). Targeting Immune Cells in the Tumor Microenvironment of HCC: New Opportunities and Challenges. *Frontiers in Cell and Developmental Biology*, *9*. https://doi.org/10.3389/fcell.2021.775462

Helms, E., Onate, M. K., & Sherman, M. H. (2020). Fibroblast Heterogeneity in the Pancreatic Tumor Microenvironment. *Cancer Discovery*, *10*(5), 648–656. https://doi.org/10.1158/2159-8290.CD-19-1353

Herceg, Z., & Paliwal, A. (2011). Epigenetic mechanisms in hepatocellular carcinoma: How environmental factors influence the epigenome. *Mutation Research/Reviews in Mutation Research*, *727*(3), 55–61. https://doi.org/10.1016/j.mrrev.2011.04.001

He, X., Zhu, Z., Johnson, C., Stoops, J., Eaker, A. E., Bowen, W., & DeFrances, M. C. (2008). PIK3IP1, a Negative Regulator of PI3K, Suppresses the Development of Hepatocellular Carcinoma. *Cancer Research*, *68*(14), 5591–5598. https://doi.org/10.1158/0008-5472.CAN-08-0025

Hornbeck, P. V., Kornhauser, J. M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., & Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Research*, *40*(D1), D261–D270. https://doi.org/10.1093/nar/gkr1122

Hu, J.-L., Zhou, B. O., Zhang, R.-R., Zhang, K.-L., Zhou, J.-Q., & Xu, G.-L. (2009). The N-terminus of histone H3 is required for de novo DNA methylation in chromatin. *Proceedings of the National Academy of Sciences*, *106*(52), 22187–22192. https://doi.org/10.1073/pnas.0905767106

Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R. W., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R., & Bird, A. P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genetics*, *6*(9), e1001134. https://doi.org/10.1371/journal.pgen.1001134

International Agency for Research on Cancer, World Health Organization, Cancer today. https://Gco.Iarc.Fr/Today/Home. Retrieved August 3, 2023, from https://gco.iarc.fr/today/home

Jemal, A., Ward, E. M., Johnson, C. J., Cronin, K. A., Ma, J., Ryerson, A. B., Mariotto, A., Lake, A. J., Wilson, R., Sherman, R. L., Anderson, R. N., Henley, S. J., Kohler, B. A., Penberthy, L., Feuer, E. J., & Weir, H. K. (2017). Annual Report to the Nation on the Status of Cancer, 1975–2014, Featuring Survival. *JNCI: Journal of the National Cancer Institute*, *109*(9). https://doi.org/10.1093/JNCI/DJX030

Jin, Y., Shenoy, A. K., Doernberg, S., Chen, H., Luo, H., Shen, H., Lin, T., Tarrash, M., Cai, Q., Hu, X., Fiske, R., Chen, T., Wu, L., Mohammed, K. A., Rottiers, V., Lee, S. S., & Lu, J. (2015). FBXO11 promotes ubiquitination of the Snail family of transcription factors in cancer progression and epidermal development. *Cancer Letters*, *362*(1), 70–82. https://doi.org/10.1016/j.canlet.2015.03.037

Kaneto, H. (2001). Detection of hypermethylation of the p16INK4A gene promoter in chronic hepatitis and cirrhosis associated with hepatitis B or C virus. *Gut*, *48*(3), 372–377. https://doi.org/10.1136/gut.48.3.372

Kang-Park, S., Im, J. H., Lee, J. H., & Lee, Y. I. (2006). PTEN modulates hepatitis B virus-X protein induced survival signaling in Chang liver cells. *Virus Research*, *122*(1–2), 53–60. https://doi.org/10.1016/j.virusres.2006.06.010

Khattar, E., Mukherji, A., & Kumar, V. (2012). Akt augments the oncogenic potential of the HBx protein of hepatitis B virus by phosphorylation. *The FEBS Journal*, *279*(7), 1220–1230. https://doi.org/10.1111/J.1742-4658.2012.08514.X

Kim, G. W., Imam, H., Khan, M., Mir, S. A., Kim, S. J., Yoon, S. K., Hur, W., & Siddiqui, A. (2021). HBV-Induced Increased N6 Methyladenosine Modification of PTEN RNA Affects Innate Immunity and Contributes to HCC. *Hepatology (Baltimore, Md.)*, *73*(2), 533–547. https://doi.org/10.1002/HEP.31313

Kim, S.-A., Zhu, J., Yennawar, N., Eek, P., & Tan, S. (2020). Crystal Structure of the LSD1/CoREST Histone Demethylase Bound to Its Nucleosome Substrate. *Molecular Cell*, *78*(5), 903-914.e4. https://doi.org/10.1016/j.molcel.2020.04.019

Knight, M. J., Leettola, C., Gingery, M., Li, H., & Bowie, J. U. (2011). A human sterile alpha motif domain polymerizome. *Protein Science*, *20*(10), 1697–1706. https://doi.org/10.1002/pro.703

Ko, M., An, J., Bandukwala, H. S., Chavez, L., Äijö, T., Pastor, W. A., Segal, M. F., Li, H., Koh, K. P., Lähdesmäki, H., Hogan, P. G., Aravind, L., & Rao, A. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature 2013 497:7447*, *497*(7447), 122–126. https://doi.org/10.1038/nature12052

Kudo, M., Finn, R. S., Qin, S., Han, K.-H., Ikeda, K., Piscaglia, F., Baron, A., Park, J.-W., Han, G., Jassem, J., Blanc, J. F., Vogel, A., Komov, D., Evans, T. R. J., Lopez, C., Dutcus, C., Guo, M., Saito, K., Kraljevic, S., … Cheng, A.-L. (2018). Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. *The Lancet*, *391*(10126), 1163–1173. https://doi.org/10.1016/S0140-6736(18)30207-1

Landré, V., Revi, B., Mir, M. G., Verma, C., Hupp, T. R., Gilbert, N., & Ball, K. L. (2017). Regulation of transcriptional activators by DNA-binding domain ubiquitination. *Cell Death & Differentiation*, *24*(5), 903–916. https://doi.org/10.1038/cdd.2017.42

Larsen, F., Gundersen, G., Lopez, R., & Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics*, *13*(4), 1095–1107. https://doi.org/10.1016/0888-7543(92)90024-M

Lee, J.-H., & Skalnik, D. G. (2005). CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. *Journal of Biological Chemistry*, *280*(50), 41725–41731. https://doi.org/10.1074/jbc.M508312200

Lees, A. M., Deconinck, A. E., Campbell, B. D., & Lees, R. S. (2005). Atherin: a newly identified, lesion-specific, LDL-binding protein in human atherosclerosis. *Atherosclerosis*, *182*(2), 219–230. https://doi.org/10.1016/J.ATHEROSCLEROSIS.2005.01.041

Lesina, M., Kurkowski, M. U., Ludes, K., Rose-John, S., Treiber, M., Klöppel, G., Yoshimura, A., Reindl, W., Sipos, B., Akira, S., Schmid, R. M., & Algül, H. (2011). Stat3/Socs3 Activation by IL-6 Transsignaling Promotes Progression of Pancreatic Intraepithelial Neoplasia and Development of Pancreatic Cancer. *Cancer Cell*, *19*(4), 456–469. https://doi.org/10.1016/j.ccr.2011.03.009

Lian, Y., Li, Z., Fan, Y., Huang, Q., Chen, J., Liu, W., Xiao, C., & Xu, H. (2017). The lncRNA-HOXA-AS2/EZH2/LSD1 oncogene complex promotes cell proliferation in pancreatic cancer. *American Journal of Translational Research*, *9*(12), 5496–5506.

Li, H., Liefke, R., Jiang, J., Kurland, J. V., Tian, W., Deng, P., Zhang, W., He, Q., Patel, D. J., Bulyk, M. L., Shi, Y., & Wang, Z. (2017). Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature*, *549*(7671), 287–291. https://doi.org/10.1038/nature23881

Li, X., Huang, J., Wu, Q., Du, Q., Wang, Y., Huang, Y., Cai, X., Geller, D. A., & Yan, Y. (2023). Inhibition of Checkpoint Kinase 1 (CHK1) Upregulates Interferon Regulatory Factor 1 (IRF1) to Promote Apoptosis and Activate Anti-Tumor Immunity via MICA in Hepatocellular Carcinoma (HCC). *Cancers*, *15*(3), 850. https://doi.org/10.3390/cancers15030850

Llovet, J. M., Ricci, S., Mazzaferro, V., Hilgard, P., Gane, E., Blanc, J.-F., de Oliveira, A. C., Santoro, A., Raoul, J.-L., Forner, A., Schwartz, M., Porta, C., Zeuzem, S., Bolondi, L., Greten, T. F., Galle, P. R., Seitz, J.-F., Borbath, I., Häussinger, D., … Bruix, J. (2008). Sorafenib in Advanced Hepatocellular Carcinoma. *New England Journal of Medicine*, *359*(4), 378–390. https://doi.org/10.1056/NEJMoa0708857

Llovet, J. M., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwartz, M., Sherman, M., & Gores, G. (2016). Hepatocellular carcinoma. *Nature Reviews Disease Primers*, *2*(1), 16018. https://doi.org/10.1038/nrdp.2016.18

Lv, S., Zhao, X., Zhang, E., Yan, Y., Ma, X., Li, N., Zou, Q., Sun, L., & Song, T. (2022). Lysine demethylase KDM1A promotes cell growth via FKBP8-BCL2 axis in hepatocellular carcinoma. *The Journal of Biological Chemistry*, *298*(9). https://doi.org/10.1016/J.JBC.2022.102374

Maes, T., Mascaró, C., Tirapu, I., Estiarte, A., Ciceri, F., Lunardi, S., Guibourt, N., Perdones, A., Lufino, M. M. P., Somervaille, T. C. P., Wiseman, D. H., Duy, C., Melnick, A., Willekens, C., Ortega, A., Martinell, M., Valls, N., Kurz, G., Fyfe, M., … Buesa, C. (2018). ORY-1001, a Potent and Selective Covalent KDM1A Inhibitor,

for the Treatment of Acute Leukemia. *Cancer Cell*, *33*(3), 495-511.e12.
https://doi.org/10.1016/j.ccell.2018.02.002

Maeta, Y., Shiota, G., Okano, J. I., & Murawaki, Y. (2005). Effect of promoter methylation of the p16 gene on
phosphorylation of retinoblastoma gene product and growth of hepatocellular carcinoma cells. *Tumour
Biology : The Journal of the International Society for Oncodevelopmental Biology and Medicine*, *26*(6), 300–
305. https://doi.org/10.1159/000089288

Mancini, D. N., Singh, S. M., Archer, T. K., & Rodenhiser, D. I. (1999). Site-specific DNA methylation in the
neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors.
Oncogene, *18*(28), 4108–4119. https://doi.org/10.1038/SJ.ONC.1202764

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., Dsouza, C., Fouse, S. D., Johnson, B. E., Hong,
C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M.,
Rowitch, D. H., Xing, X., Fiore, C., … Costello, J. F. (2010). Conserved role of intragenic DNA methylation
in regulating alternative promoters. *Nature 2010 466:7303*, *466*(7303), 253–257.
https://doi.org/10.1038/nature09165

McDonald, O. G., Li, X., Saunders, T., Tryggvadottir, R., Mentch, S. J., Warmoes, M. O., Word, A. E., Carrer, A.,
Salz, T. H., Natsume, S., Stauffer, K. M., Makohon-Moore, A., Zhong, Y., Wu, H., Wellen, K. E., Locasale,
J. W., Iacobuzio-Donahue, C. A., & Feinberg, A. P. (2017). Epigenomic reprogramming during pancreatic
cancer progression links anabolic glucose metabolism to distant metastasis. *Nature Genetics*, *49*(3), 367–
376. https://doi.org/10.1038/ng.3753

Meruelo, A. D., & Bowie, J. U. (2009). Identifying polymer-forming SAM domains. *Proteins*, *74*(1), 1–5.
https://doi.org/10.1002/prot.22232

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim,
T. K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A.,
Wernig, M., Jaenisch, R., … Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent
and lineage-committed cells. *Nature*, *448*(7153), 553–560. https://doi.org/10.1038/NATURE06008

Morselli, M., Pastor, W. A., Montanini, B., Nee, K., Ferrari, R., Fu, K., Bonora, G., Rubbi, L., Clark, A. T.,
Ottonello, S., Jacobsen, S. E., & Pellegrini, M. (2015). In vivo targeting of de novo DNA methylation by
histone modifications in yeast and mouse. *ELife*, *4*. https://doi.org/10.7554/eLife.06205

Mrozik, K. M., Blaschuk, O. W., Cheong, C. M., Zannettino, A. C. W., & Vandyke, K. (2018). N-cadherin in cancer
metastasis, its emerging role in haematological malignancies and potential as a therapeutic target in
cancer. *BMC Cancer*, *18*(1), 939. https://doi.org/10.1186/s12885-018-4845-0

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H. G.,
Wilm, M., & Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at
Polycomb target genes. *The EMBO Journal*, *26*(18), 4078–4088. https://doi.org/10.1038/sj.emboj.7601837

Pandey, S., Gupta, V. K., & Lavania, S. P. (2023). Role of epigenetics in pancreatic ductal adenocarcinoma.
Epigenomics, *15*(2), 89–110. https://doi.org/10.2217/epi-2022-0177

Park, W., Chawla, A., & O'Reilly, E. M. (2021). Pancreatic Cancer: A Review. *JAMA*, *326*(9), 851–862.
https://doi.org/10.1001/jama.2021.13027

Pastushenko, I., & Blanpain, C. (2019). EMT Transition States during Tumor Progression and Metastasis. *Trends in Cell Biology*, *29*(3), 212–226. https://doi.org/10.1016/j.tcb.2018.12.001

Perino, M., Van Mierlo, G., Karemaker, I. D., Van Genesen, S., Vermeulen, M., Marks, H., Van Heeringen, S. J., & Veenstra, G. J. C. (2018). MTF2 recruits Polycomb Repressive Complex 2 by helical-shape-selective DNA binding. *Nature Genetics*, *50*(7), 1002–1010. https://doi.org/10.1038/S41588-018-0134-8

Prestel, M., Prell-Schicker, C., Webb, T., Malik, R., Lindner, B., Ziesch, N., Rex-Haffner, M., Röh, S., Viturawong, T., Lehm, M., Mokry, M., den Ruijter, H., Haitjema, S., Asare, Y., Söllner, F., Najafabadi, M. G., Aherrahrou, R., Civelek, M., Samani, N. J., … Dichgans, M. (2019). The Atherosclerosis Risk Variant rs2107595 Mediates Allele-Specific Transcriptional Regulation of HDAC9 via E2F3 and Rb1. *Stroke*, *50*(10), 2651–2660. https://doi.org/10.1161/STROKEAHA.119.026112

Qin, Y., Zhu, W., Xu, W., Zhang, B., Shi, S., Ji, S., Liu, J., Long, J., Liu, C., Liu, L., Xu, J., & Yu, X. (2014). LSD1 sustains pancreatic cancer growth via maintaining HIF1α-dependent glycolytic process. *Cancer Letters*, *347*(2), 225–232. https://doi.org/10.1016/j.canlet.2014.02.013

Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., & Matrisian, L. M. (2014). Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Research*, *74*(11), 2913–2921. https://doi.org/10.1158/0008-5472.CAN-14-0155

Ramirez-Carrozzi, V. R., Braas, D., Bhatt, D. M., Cheng, C. S., Hong, C., Doty, K. R., Black, J. C., Hoffmann, A., Carey, M., & Smale, S. T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell*, *138*(1), 114–128. https://doi.org/10.1016/J.CELL.2009.04.020

Raphael, B. J., Hruban, R. H., Aguirre, A. J., Moffitt, R. A., Yeh, J. J., Stewart, C., Robertson, A. G., Cherniack, A. D., Gupta, M., Getz, G., Gabriel, S. B., Meyerson, M., Cibulskis, C., Fei, S. S., Hinoue, T., Shen, H., Laird, P. W., Ling, S., Lu, Y., … Zenklusen, J. C. (2017). Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell*, *32*(2), 185-203.e13. https://doi.org/10.1016/J.CCELL.2017.07.007

Rasheed, Z. A., Yang, J., Wang, Q., Kowalski, J., Freed, I., Murter, C., Hong, S.-M., Koorstra, J.-B., Rajeshkumar, N. V., He, X., Goggins, M., Iacobuzio-Donahue, C., Berman, D. M., Laheru, D., Jimeno, A., Hidalgo, M., Maitra, A., & Matsui, W. (2010). Prognostic Significance of Tumorigenic Cells With Mesenchymal Features in Pancreatic Adenocarcinoma. *JNCI: Journal of the National Cancer Institute*, *102*(5), 340–351. https://doi.org/10.1093/jnci/djp535

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., & Pedone, P. V. (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *The Journal of Biological Chemistry*, *282*(46), 33336–33345. https://doi.org/10.1074/JBC.M706213200

Rhee, I., Bachman, K. E., Park, B. H., Jair, K.-W., Yen, R.-W. C., Schuebel, K. E., Cui, H., Feinberg, A. P., Lengauer, C., Kinzler, K. W., Baylin, S. B., & Vogelstein, B. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, *416*(6880), 552–556. https://doi.org/10.1038/416552a

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., & Hume, D. A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nature Reviews Genetics*, *8*(6), 424–436. https://doi.org/10.1038/nrg2026

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(5), 1412–1417. https://doi.org/10.1073/PNAS.0510310103/SUPPL_FILE/10310TABLE6.XLS

Scarpa, A., Capelli, P., Mukai, K., Zamboni, G., Oda, T., Iacono, C., & Hirohashi, S. (1993). Pancreatic adenocarcinomas frequently show p53 gene mutations. *The American Journal of Pathology*, *142*(5), 1534–1543.

Schulze, K., Nault, J. C., & Villanueva, A. (2016). Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *Journal of Hepatology*, *65*(5), 1031–1042. https://doi.org/10.1016/J.JHEP.2016.05.035

Schutte, M., Hruban, R. H., Geradts, J., Maynard, R., Hilgers, W., Rabindran, S. K., Moskaluk, C. A., Hahn, S. A., Schwarte-Waldhoff, I., Schmiegel, W., Baylin, S. B., Kern, S. E., & Herman, J. G. (1997). Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. *Cancer Research*, *57*(15), 3126–3130.

Shah, A. N., Summy, J. M., Zhang, J., Park, S. I., Parikh, N. U., & Gallick, G. E. (2007). Development and Characterization of Gemcitabine-Resistant Pancreatic Tumor Cells. *Annals of Surgical Oncology*, *14*(12), 3629–3637. https://doi.org/10.1245/s10434-007-9583-5

Shin Voo, K., Carlone, D. L., Jacobsen, B. M., Flodin, A., & Skalnik, D. G. (2000). Cloning of a Mammalian Transcriptional Activator That Binds Unmethylated CpG Motifs and Shares a CXXC Domain with DNA Methyltransferase, Human Trithorax, and Methyl-CpG Binding Domain Protein 1. *Molecular and Cellular Biology*, *20*(6), 2108–2121. https://doi.org/10.1128/MCB.20.6.2108-2121.2000

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., & Shi, Y. (2004). Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell*, *119*(7), 941–953. https://doi.org/10.1016/j.cell.2004.12.012

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians*, *71*(1), 7–33. https://doi.org/10.3322/CAAC.21654

Simon, C., Stielow, B., Nist, A., Rohner, I., Weber, L. M., Geller, M., Fischer, S., Stiewe, T., & Liefke, R. (2022). The CpG Island-Binding Protein SAMD1 Contributes to an Unfavorable Gene Signature in HepG2 Hepatocellular Carcinoma Cells. *Biology*, *11*(4). https://doi.org/10.3390/BIOLOGY11040557

Song, M.-A., Kwee, S. A., Tiirikainen, M., Hernandez, B. Y., Okimoto, G., Tsai, N. C., Wong, L. L., & Yu, H. (2016). Comparison of genome-scale DNA methylation profiles in hepatocellular carcinoma by viral status. *Epigenetics*, *11*(6), 464–474. https://doi.org/10.1080/15592294.2016.1151586

Stewart, K. R., Veselovska, L., Kim, J., Huang, J., Saadeh, H., Tomizawa, S. I., Smallwood, S. A., Chen, T., & Kelsey, G. (2015). Dynamic changes in histone modifications precede de novo DNA methylation in oocytes. *Genes & Development*, *29*(23), 2449–2462. https://doi.org/10.1101/GAD.271353.115

Stielow, B., Simon, C., & Liefke, R. (2021). Making fundamental scientific discoveries by combining information from literature, databases, and computational tools – An example. *Computational and Structural Biotechnology Journal*, *19*, 3027–3033. https://doi.org/10.1016/J.CSBJ.2021.04.052

Stielow, B., Zhou, Y., Cao, Y., Simon, C., Pogoda, H. M., Jiang, J., Ren, Y., Phanor, S. K., Rohner, I., Nist, A., Stiewe, T., Hammerschmidt, M., Shi, Y., Bulyk, M. L., Wang, Z., & Liefke, R. (2021). The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Science Advances*, *7*(20). https://doi.org/10.1126/SCIADV.ABF2229

Stock, J. K., Giadrossi, S., Casanova, M., Brookes, E., Vidal, M., Koseki, H., Brockdorff, N., Fisher, A. G., & Pombo, A. (2007). Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nature Cell Biology*, *9*(12), 1428–1435. https://doi.org/10.1038/ncb1663

Sun, E. J., Wankell, M., Palamuthusingam, P., McFarlane, C., & Hebbard, L. (2021). Targeting the PI3K/Akt/mTOR Pathway in Hepatocellular Carcinoma. *Biomedicines*, *9*(11), 1639. https://doi.org/10.3390/biomedicines9111639

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., & Rao, A. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, *324*(5929), 930–935. https://doi.org/10.1126/science.1170116

Tazi, J., & Bird, A. (1990). Alternative chromatin structure at CpG islands. *Cell*, *60*(6), 909–920. https://doi.org/10.1016/0092-8674(90)90339-G

Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R. W., Deaton, A., Andrews, R., James, K. D., Turner, D. J., Illingworth, R., & Bird, A. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*, *464*(7291), 1082–1086. https://doi.org/10.1038/NATURE08924

Tian, S., Cao, Y., Wang, J., Bi, Y., Zhong, J., Meng, X., Sun, W., Yang, R., Gan, L., Wang, X., Li, H., & Wang, R. (2021). The miR-378c-Samd1 circuit promotes phenotypic modulation of vascular smooth muscle cells and foam cells formation in atherosclerosis lesions. *Scientific Reports 2021 11:1*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-89981-z

Timmer, F. E. F., Geboers, B., Nieuwenhuizen, S., Dijkstra, M., Schouten, E. A. C., Puijk, R. S., de Vries, J. J. J., van den Tol, M. P., Bruynzeel, A. M. E., Streppel, M. M., Wilmink, J. W., van der Vliet, H. J., Meijerink, M. R., Scheffer, H. J., & de Gruijl, T. D. (2021). Pancreatic Cancer and Immunotherapy: A Clinical Overview. *Cancers*, *13*(16), 4138. https://doi.org/10.3390/cancers13164138

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., & Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. *Nature*, *439*(7078), 811–816. https://doi.org/10.1038/nature04433

Turner, P. C., Sylla, A., Diallo, M. S., Castegnaro, J. J., Hall, A. J., & Wild, C. P. (2002). The role of aflatoxins and hepatitis viruses in the etiopathogenesis of hepatocellular carcinoma: A basis for primary prevention in Guinea-Conakry, West Africa. *Journal of Gastroenterology and Hepatology*, *17 Suppl*(SUPPL. 4). https://doi.org/10.1046/J.1440-1746.17.S4.7.X

Vara, J. Á. F., Casado, E., de Castro, J., Cejas, P., Belda-Iniesta, C., & González-Barón, M. (2004). PI3K/Akt signalling pathway and cancer. *Cancer Treatment Reviews*, *30*(2), 193–204. https://doi.org/10.1016/j.ctrv.2003.07.007

Villanueva, A. (2019). Hepatocellular Carcinoma. *New England Journal of Medicine*, *380*(15), 1450–1462. https://doi.org/10.1056/NEJMra1713263

References

Villanueva, A., Hoshida, Y., Battiston, C., Tovar, V., Sia, D., Alsinet, C., Cornella, H., Liberzon, A., Kobayashi, M., Kumada, H., Thung, S. N., Bruix, J., Newell, P., April, C., Fan, J., Roayaie, S., Mazzaferro, V., Schwartz, M. E., & Llovet, J. M. (2011). Combining Clinical, Pathology, and Gene Expression Data to Predict Recurrence of Hepatocellular Carcinoma. *Gastroenterology*, *140*(5), 1501-1512.e2. https://doi.org/10.1053/j.gastro.2011.02.006

Vincenzi, M., Mercurio, F. A., & Leone, M. (2020). Sam Domains in Multiple Diseases. *Current Medicinal Chemistry*, *27*(3), 450–476. https://doi.org/10.2174/0929867325666181009114445

Viturawong, T., Meissner, F., Butter, F., & Mann, M. (2013). A DNA-centric protein interaction map of ultraconserved elements reveals contribution of transcription factor binding hubs to conservation. *Cell Reports*, *5*(2), 531–545. https://doi.org/10.1016/J.CELREP.2013.09.022

Voigt, P., Tee, W.-W., & Reinberg, D. (2013). A double take on bivalent promoters. *Genes & Development*, *27*(12), 1318–1338. https://doi.org/10.1101/gad.219626.113

Waddell, N., Pajic, M., Patch, A.-M., Chang, D. K., Kassahn, K. S., Bailey, P., Johns, A. L., Miller, D., Nones, K., Quek, K., Quinn, M. C. J., Robertson, A. J., Fadlullah, M. Z. H., Bruxner, T. J. C., Christ, A. N., Harliwong, I., Idrisoglu, S., Manning, S., Nourse, C., … Grimmond, S. M. (2015). Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature*, *518*(7540), 495–501. https://doi.org/10.1038/nature14169

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., & Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature*, *431*(7010), 873–878. https://doi.org/10.1038/nature02985

Wang, S., Huang, S., & Sun, Y. L. (2017). Epithelial-Mesenchymal Transition in Pancreatic Cancer: A Review. *BioMed Research International*, *2017*, 1–10. https://doi.org/10.1155/2017/2646148

Wang, Y., & Cao, K. (2021). KDM1A Promotes Immunosuppression in Hepatocellular Carcinoma by Regulating PD-L1 through Demethylating MEF2D. *Journal of Immunology Research*, *2021*. https://doi.org/10.1155/2021/9965099

Weber, L. M., Jia, Y., Stielow, B., Gisselbrecht, S. S., Cao, Y., Ren, Y., Rohner, I., King, J., Rothman, E., Fischer, S., Simon, C., Forné, I., Nist, A., Stiewe, T., Bulyk, M. L., Wang, Z., & Liefke, R. (2023). The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Research*, *51*(2), 574–594. https://doi.org/10.1093/NAR/GKAC1188

Wood, L. D., Canto, M. I., Jaffee, E. M., & Simeone, D. M. (2022). Pancreatic Cancer: Pathogenesis, Screening, Diagnosis, and Treatment. *Gastroenterology*, *163*(2), 386-402.e1. https://doi.org/10.1053/j.gastro.2022.03.056

Wu, H.-X., Wang, Z.-X., Zhao, Q., Chen, D.-L., He, M.-M., Yang, L.-P., Wang, Y.-N., Jin, Y., Ren, C., Luo, H.-Y., Wang, Z.-Q., & Wang, F. (2019). Tumor mutational and indel burden: a systematic pan-cancer evaluation as prognostic biomarkers. *Annals of Translational Medicine*, *7*(22), 640–640. https://doi.org/10.21037/ATM.2019.10.116

Wu, X., Johansen, J. V., & Helin, K. (2013). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. *Molecular Cell*, *49*(6), 1134–1146. https://doi.org/10.1016/j.molcel.2013.01.016

References

Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., Zheng, Y., Nishinakamura, R., Xu, G. L., Wang, H., Chen, S., Gao, S., & Zhu, B. (2016). Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. *Molecular Cell*, *64*(5), 913–925. https://doi.org/10.1016/J.MOLCEL.2016.10.013

Xu, C., Liu, K., Lei, M., Yang, A., Li, Y., Hughes, T. R., & Min, J. (2018). DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure*, *26*(1), 85-95.e3. https://doi.org/10.1016/j.str.2017.11.022

Xu, P., Scott, D. C., Xu, B., Yao, Y., Feng, R., Cheng, L., Mayberry, K., Wang, Y.-D., Bi, W., Palmer, L. E., King, M. T., Wang, H., Li, Y., Fan, Y., Alpi, A. F., Li, C., Peng, J., Papizan, J., Pruett-Miller, S. M., … Weiss, M. J. (2021). FBXO11-mediated proteolysis of BAHD1 relieves PRC2-dependent transcriptional repression in erythropoiesis. *Blood*, *137*(2), 155–167. https://doi.org/10.1182/blood.2020007809

Xu, T., Park, S.-S., Giaimo, B. D., Hall, D., Ferrante, F., Ho, D. M., Hori, K., Anhezini, L., Ertl, I., Bartkuhn, M., Zhang, H., Milon, E., Ha, K., Conlon, K. P., Kuick, R., Govindarajoo, B., Zhang, Y., Sun, Y., Dou, Y., … Rual, J.-F. (2017). RBPJ/CBF1 interacts with L3MBTL3/MBT1 to promote repression of Notch signaling via histone demethylase KDM1A/LSD1. *The EMBO Journal*, *36*(21), 3232–3249. https://doi.org/10.15252/embj.201796525

Yan, Y., Zheng, L., Du, Q., Yan, B., & Geller, D. A. (2020). Interferon regulatory factor 1 (IRF-1) and IRF-2 regulate PD-L1 expression in hepatocellular carcinoma (HCC) cells. *Cancer Immunology, Immunotherapy*, *69*(9), 1891–1903. https://doi.org/10.1007/s00262-020-02586-9

Yan, Y., Zheng, L., Du, Q., Yazdani, H., Dong, K., Guo, Y., & Geller, D. A. (2021). Interferon regulatory factor 1(IRF-1) activates anti-tumor immunity via CXCL10/CXCR3 axis in hepatocellular carcinoma (HCC). *Cancer Letters*, *506*, 95–106. https://doi.org/10.1016/j.canlet.2021.03.002

Younossi, Z., Stepanova, M., Ong, J. P., Jacobson, I. M., Bugianesi, E., Duseja, A., Eguchi, Y., Wong, V. W., Negro, F., Yilmaz, Y., Romero-Gomez, M., George, J., Ahmed, A., Wong, R., Younossi, I., Ziayee, M., & Afendy, A. (2019). Nonalcoholic Steatohepatitis Is the Fastest Growing Cause of Hepatocellular Carcinoma in Liver Transplant Candidates. *Clinical Gastroenterology and Hepatology*, *17*(4), 748-755.e3. https://doi.org/10.1016/j.cgh.2018.05.057

Zhao, J., Wu, G., Bu, F., Lu, B., Liang, A., Cao, L., Tong, X., Lu, X., Wu, M., & Guo, Y. (2010). Epigenetic silence of ankyrin-repeat-containing, SH3-domain-containing, and proline-rich-region- containing protein 1 (ASPP1) and ASPP2 genes promotes tumor growth in hepatitis B virus-positive hepatocellular carcinoma. *Hepatology*, *51*(1), 142–153. https://doi.org/10.1002/hep.23247

Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepp, D. M., Elledge, S. J., Pagano, M., Conaway, R. C., Conaway, J. W., Harper, J. W., & Pavletich, N. P. (2002). Structure of the Cul1–Rbx1–Skp1–F boxSkp2 SCF ubiquitin ligase complex. *Nature*, *416*(6882), 703–709. https://doi.org/10.1038/416703a

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K. A., Lozach, J., Glass, C. K., & Rosenfeld, M. G. (2008). Histone H2A Monoubiquitination Represses Transcription by Inhibiting RNA Polymerase II Transcriptional Elongation. *Molecular Cell*, *29*(1), 69–80. https://doi.org/10.1016/j.molcel.2007.11.002

Zhu, J., He, F., Hu, S., & Yu, J. (2008). On the nature of human housekeeping genes. *Trends in Genetics*, *24*(10), 481–484. https://doi.org/10.1016/j.tig.2008.08.004

Zhu, W.-G., Srinivasan, K., Dai, Z., Duan, W., Druhan, L. J., Ding, H., Yee, L., Villalona-Calero, M. A., Plass, C., & Otterson, G. A. (2003). Methylation of adjacent CpG sites affects Sp1/Sp3 binding and activity in the p21(Cip1) promoter. *Molecular and Cellular Biology*, *23*(12), 4056–4065. https://doi.org/10.1128/MCB.23.12.4056-4065.2003

Zhu, Z., He, X., Johnson, C., Stoops, J., Eaker, A. E., Stoffer, D. S., Bell, A., Zarnegar, R., & DeFrances, M. C. (2007). PI3K is negatively regulated by PIK3IP1, a novel p110 interacting protein. *Biochemical and Biophysical Research Communications*, *358*(1), 66–72. https://doi.org/10.1016/j.bbrc.2007.04.096

Zong, Z., Zou, J., Mao, R., Ma, C., Li, N., Wang, J., Wang, X., Zhou, H., Zhang, L., & Shi, Y. (2019). M1 Macrophages Induce PD-L1 Expression in Hepatocellular Carcinoma Cells Through IL-1β Signaling. *Frontiers in Immunology*, *10*, 1643. https://doi.org/10.3389/fimmu.2019.01643

Zucman-Rossi, J., Villanueva, A., Nault, J.-C., & Llovet, J. M. (2015). Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology*, *149*(5), 1226-1239.e4. https://doi.org/10.1053/j.gastro.2015.05.061

8. Appendix

List of abbreviations

4-OHT	4-Hydroxytamoxifen
AKT	Serine/Threonine Kinase
ATF4	Activating Transcription Factor 4
C	cytosine
CAF	Cancer-Associated Fibroblast
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
CGI	CpG island
ChIP	Chromatin Immunoprecipitation
Co-IP	Co-Immunoprecipitation
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR-associated 9
CTLA4	Cytotoxic T-Lymphocyte Associated Protein 4
CTNNB1	Catenin Beta 1
CUL1	Cullin 1
DNA	Desoxyribonucleic Acid
EMSA	Electrophoretic Mobility Shift Assay
EMT	Epithelial-Mesenchymal Transition
EZH2	Enhancer of Zeste Homolog 2
G	guanine
GSEA	Gene Set Enrichment Analysis
H2AK119Ub1	Histone H2A lysine 119 mono ubiquitination
H3K27me3	histone H3 lysine 27 trimethylation
H3K36me3	histone H3 lysine 36 trimethylation
H3K4me2/3	histone H3 lysine 4 di-/trimethylation
HCC	Hepatocellular Carcinoma
HIF1α	Hypoxia Inducible Factor 1 Subunit Alpha
HNF	Hepatocyte Nuclear Factor
HOXA-AS2	HOXA Cluster Antisense RNA 2
IDR	Intrinsically Unstructured Region
IRF1	Interferon Regulatory Factor 1
ISRE	IFN-Stimulated Response Element
KAT	Lysine Acetyl Transferase

KDM	Lysine Demethylase
KDM1A	Lysine-Specific Histone Demethylase 1A
KMT	Lysine Methyltransferase
KO	knockout
KRAS	Kirsten Rat Sarcoma
L3MBTL3	Lethal(3)Malignant Brain Tumor-Like Protein 3
LC-MS	Liquid Chromatography–Mass Spectrometry
LDL	Low-Density Lipoprotein
mESC	mouse embryonic stem cells
MTOR	Mechanistic Target Of Rapamycin Kinase
PanIN	Pancreatic Intraepithelial Neoplasia
PCA	Principal Component Analysis
PCL	Polycomb-like protein
PDAC	Pancreatic Ductal Adenocarcinoma
PD-L1	Programmed Death Ligand 1
PIK3IP1	Phosphoinositide-3-Kinase Interacting Protein 1
PRC	Polycomb Repressive Complex
PTEN	Phosphatase And Tensin Homolog
qPCR	quantitative Polymerase Chain Reaction
RBPJ	Recombination signal Binding Protein for immunoglobulin kappa J region
RBX1	Ring-Box 1
RCOR1	REST Corepressor 1
RNA	Ribonucleic Acid
RNAP II	RNA polymerase II
SAMD1	Sterile Alpha Motif Domain-containing protein 1
SET1	SET Domain Containing 1A
SFBMT1/2	Scm Like With Four Mbt Domains ½
sgRNA	Single guide RNA
SKP1	S-Phase Kinase Associated Protein 1
SMAD4	Mothers Against Decapentaplegic Homolog 4
SNP	single nucleotide polymorphism
SWI/SNF	SWItch/sucrose non-fermentable
TME	Tumor Microenvironment
TP53	Tumor Protein 53
TSS	Transcription Start Site
WH	Winged-Helix

List of academic teachers

My academic teachers at the Philipps-University Marburg were:

Adhikary, Till

Bartsch, Jörg Walter

Bauer, Uta Maria

Brandt, Dominique

Brehm, Alexander

Brendel, Cornelia

Brichkina, Anna

Buchholz, Malte

Burchert, Andreas

Chung, Ho Ryun

Conrad, Matthias

Decher, Niels

Elsässer, Hans Peter

Exner, Cornelia

Feuser, Beate

Frech, Miriam

Fritz, Barbara

Greene, Brandon

Hänze, Jörg

Huber, Magdalena

Jacob, Ralf

Lauth, Matthias

Lill, Roland

Linklater, Nicole

Maisner, Andrea

Milani, Wiebke

Müller, Rolf

Müller Brüsselbach, Sabine

Mühlenhoff, Ulrich

Neubauer, Andreas

Oberwinkler, Johannes

Oliver, Dominik

Plant, Timothy David

Pogge von Strandmann, Elke

Preisig Müller, Regina

Reinartz, Silke

Schmeck, Bernd

Schnarre, Markus

Stehling, Oliver

Steinhoff, Ulrich

Stiewe, Thorsten

Strauer, Dorothea

Suske, Guntram

Timofeev, Oleg

Visekruna, Alexander

Wanzel, Michael

Worzfeld, Thomas

Westermann, Reiner

Wrocklage, Christian

Publications

The following publications were discussed in this thesis:

1. The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands

2. The CpG island-binding protein SAMD1 contributes to an unfavorable gene signature in HepG2 hepatocellular carcinoma cells

3. SAMD1 suppresses epithelial-mesenchymal transition (EMT) pathways in pancreatic ductal adenocarcinoma

STRUCTURAL BIOLOGY

The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands

Bastian Stielow[1†], Yuqiao Zhou[2†], Yinghua Cao[2†], Clara Simon[1], Hans-Martin Pogoda[3,4], Junyi Jiang[2], Yanpeng Ren[2], Sabrina Keita Phanor[5], Iris Rohner[1], Andrea Nist[6], Thorsten Stiewe[6], Matthias Hammerschmidt[3,4], Yang Shi[7,8], Martha L. Bulyk[5,9], Zhanxin Wang[2]*, Robert Liefke[1,10]*

CpG islands (CGIs) are key regulatory DNA elements at most promoters, but how they influence the chromatin status and transcription remains elusive. Here, we identify and characterize SAMD1 (SAM domain-containing protein 1) as an unmethylated CGI-binding protein. SAMD1 has an atypical winged-helix domain that directly recognizes unmethylated CpG-containing DNA via simultaneous interactions with both the major and the minor groove. The SAM domain interacts with L3MBTL3, but it can also homopolymerize into a closed pentameric ring. At a genome-wide level, SAMD1 localizes to H3K4me3-decorated CGIs, where it acts as a repressor. SAMD1 tethers L3MBTL3 to chromatin and interacts with the KDM1A histone demethylase complex to modulate H3K4me2 and H3K4me3 levels at CGIs, thereby providing a mechanism for SAMD1-mediated transcriptional repression. The absence of SAMD1 impairs ES cell differentiation processes, leading to misregulation of key biological pathways. Together, our work establishes SAMD1 as a newly identified chromatin regulator acting at unmethylated CGIs.

INTRODUCTION

Vertebrate CpG islands (CGIs) are specific genomic regions characterized by the accumulation of CpG dinucleotides. They are commonly found at gene promoters and play important roles in gene regulation (1). Most of the CGIs are in an unmethylated state, and the associated genes are typically actively transcribed. The CXXC domain, which contains two zinc fingers, was first identified to specifically recognize unmethylated but not methylated CpG motifs (2). Proteins having CXXC domains are often subunits of larger protein complexes involved in modifying the chromatin state. For example, CXXC1 (CFP1) is part of an H3K4me3 methyltransferase complex (3) that is important to establish an active chromatin state. In contrast, KDM2B is associated with the Polycomb repressive complex 1 (PRC1) and is involved in establishing a repressive state (4). Recently, we identified the winged-helix (WH) domain of the Polycomb-like (PCL) proteins as a second type of an unmethylated CpG motif–binding domain. The PCL proteins are responsible for the recruitment of PRC2 to Polycomb-targeted CGIs (5). Because of

the action of different CpG-binding proteins, distinct chromatin states can be established at CGIs. Most CGIs are enriched in active H3K4me3 marks and are typically associated with highly expressed genes. In contrast, unmethylated CGIs at the Polycomb target genes are decorated by the repressive H3K27me3 and H2Aub marks, deposited by the PRCs. A third group of CGIs are bivalent in that they have both H3K4me3 and H3K27me3 marks. This bivalent state is proposed to allow a rapid activation during differentiation processes (6). Despite major progresses in understanding the molecular mechanisms that govern the activity of CGIs, many aspects regarding the regulation of these abundant and fundamental genomic elements remain poorly understood (1).

Here, we identify and characterize the essentially uncharacterized protein sterile alpha motif (SAM) domain-containing protein 1 (SAMD1) as a novel unmethylated CGI-binding protein and show that SAMD1 directly binds to unmethylated CpG motifs through an atypical WH domain. SAMD1 acts in concert with chromatin regulators, such as KDM1A and L3MBTL3, to modulate the function of active CGIs in mouse embryonic stem (ES) cells, a process required for proper ES cell differentiation.

RESULTS

SAMD1 is a nuclear protein that binds to unmethylated CGIs

We recently found that the WH domains of the PCL proteins functions as a new type of unmethylated CGI-binding domains (5). Thus, we speculated that there could be other unknown domains or proteins that regulate CGI function through direct interaction with the CpG motifs. To identify potential unmethylated CpG-binding proteins, we surveyed available literature and data. Investigation of mass spectrometry datasets revealed that a protein called SAMD1 (also named Atherin) behaves like known CGI-binding proteins. Specifically, SAMD1 was pulled down with CpG-rich DNA (7, 8) and repelled by methylated DNA (9, 10). SAMD1 is also associated with chromatin and chromatin-related protein complexes (11–13),

[1]Institute of Molecular Biology and Tumor Research (IMT), Philipps University of Marburg, 35043 Marburg, Germany. [2]Key Laboratory of Cell Proliferation and Regulation Biology of Ministry of Education, College of Life Sciences, Beijing Normal University, 19 Xinjiekouwai Avenue, Beijing 100875, China. [3]Institute of Zoology, University of Cologne, Cologne, Germany. [4]Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany. [5]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. [6]Genomics Core Facility, Institute of Molecular Oncology, Member of the German Center for Lung Research (DZL), Philipps University of Marburg, 35043 Marburg, Germany. [7]Division of Newborn Medicine and Epigenetics Program, Department of Pediatrics, Boston Children's Hospital, Boston, Harvard Medical School, Boston, MA 02115, USA. [8]Ludwig Institute for Cancer Research, Oxford University, Oxford, UK. [9]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA. [10]Department of Hematology, Oncology and Immunology, University Hospital Giessen and Marburg, 35043 Marburg, Germany.
*Corresponding author. Email: robert.liefke@imt.uni-marburg.de (R.L.); wangz@bnu.edu.cn (Z.W.)
†These authors contributed equally to this work.

suggesting a function at chromatin. Together, we concluded that SAMD1 may be a potential new CGI regulator.

SAMD1 is a vertebrate-specific protein that has a C-terminal SAM domain, a central unstructured region, and an unannotated globular N-terminal domain, computationally predicted (14, 15) to be a WH domain (Fig. 1A and fig. S1, A and B). Related sequences to this domain were identified in the transcriptional regulators KAT6A, KAT6B, and ZMYND11 (fig. S1, B and C), which we grouped together as a novel class of WH domains (fig. S1D). Since SAMD1 is pulled down by CpG-rich sequences and is repelled from hydroxy-methylated counterparts, similar to the PCL protein MTF2 (7, 9), we hypothesized that the WH domain of SAMD1 might facilitate binding to CpG-rich sequences. Through EMSA (electrophoretic mobility shift assay) experiments, we observed that it is the WH, but not the SAM domain, that is responsible for binding to CpG-containing DNA (Fig. 1B). No binding was observed for AT-rich DNA for both domains. To investigate the DNA binding of the WH domain in an unbiased fashion, we made use of universal protein binding microarrays (PBMs) that represent all possible 10–base pair (bp) sequences (16). This assay identified the GCGC sequence as a preferred motif recognized by the SAMD1-WH domain in vitro (Fig. 1C and fig. S1E), consistent with our hypothesis.

To verify our in vitro findings, we determined the cellular localization and genomic binding loci of SAMD1 in vivo. SAMD1 is expressed to similar levels in different mouse organs, with the strongest expression shown in mouse ES cells (fig. S2A). Thus, we used mouse ES cells as the model system for further investigations. We generated SAMD1 knockout (KO) cells (fig. S2B), which proliferated normally without any obvious phenotype (fig. S2, C and D). Using a custom-made antibody, we found that endogenous SAMD1 is predominantly nuclear localized, with a substantial proportion associated with chromatin (Fig. 1, D and E), supporting a potential chromatin-related function. Subsequently, using chromatin immunoprecipitation sequencing (ChIP-seq), we identified 8733 significant peaks and discovered that they strongly (>90%) overlap with nonmethylated CGIs but not with methylated CGIs (Fig. 1, F to H). The ChIP-seq signal was absent in SAMD1 KO cells, demonstrating the specificity of the antibody (Fig. 1, F and G). SAMD1 is highly enriched at some CGIs such as those of the *Cbln1*, *Nanos1*, and *Pth2* genes, while it shows only a subtle or no binding to other CGIs, suggesting preferential binding to certain CGIs (Fig. 1F). Comparing the sequences of the SAMD1-bound versus the unbound CGIs, a GCGC-containing motif is enriched (Fig. 1I), consistent with the motif identified by the in vitro PBM (Fig. 1C).
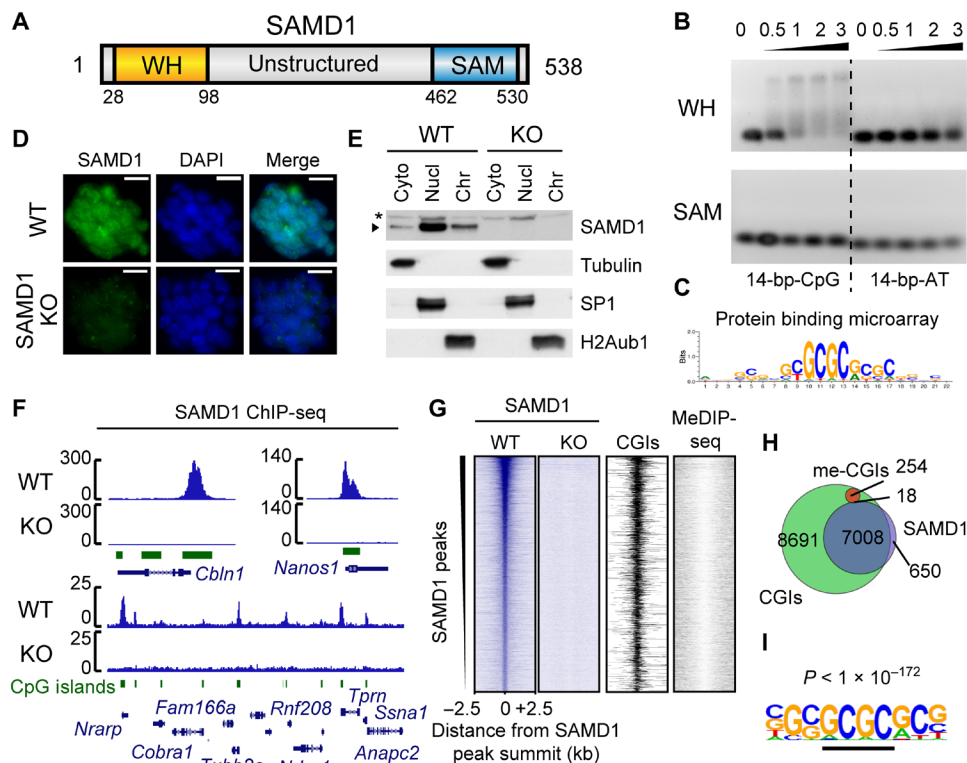
Fig. 1. Characterization of SAMD1 CpG-binding in vitro and in vivo. (**A**) Domain structure of SAMD1. (**B**) EMSA of the SAMD1-WH and SAMD1-SAM domain using CpG-rich and AT-rich DNA. The numbers indicate the molar ratio. (**C**) Representative DNA binding motif of the SAMD1-WH domain, derived from PBM experiment (16). The experiment was performed in two replicates with two distinct WH domain constructs (fig. S1E). (**D**) Immunofluorescence of the endogenous SAMD1 in wild-type (WT) and SAMD1 knockout KO cells. Scale bars, 20 μM. (**E**) Cellular fractionation of mouse ES cells followed by Western blotting. Asterisks indicate a nonspecific band (Cyto, cytoplasm; Nucl, nucleoplasm; Chr, chromatin). (**F**) Example ChIP-seq peaks of the endogenous SAMD1 in mouse ES cells. (**G**) Heatmap showing SAMD1 enrichment at SAMD1 peaks (n = 8733). The heatmaps of the KO, CGI position, and DNA methylation (MeDIP-seq) are shown in comparison. (**H**) Venn diagram showing the overlap of SAMD1 peaks (blue) with all CGIs (green) and methylated CGIs (red). (**I**) Top enriched motif at SAMD1-bound versus unbound CGIs is obtained by HOMER. DAPI, 4′,6-diamidino-2-phenylindole.

SAMD1's WH domain interacts with the minor and major groove of DNA

To address the molecular details of the interaction of SAMD1 with DNA, we solved the crystal structure of the SAMD1-WH domain in complex with 5′-GCGC-3′–containing double-stranded DNA (dsDNA) at a resolution of 1.78 Å (Table 1). Unlike a typical WH domain that contains three β strands and two wing-like loops (named W1 and W2) (17), SAMD1-WH contains only two β strands (named β1 and β2) and the W1 loop connecting both strands in addition to three conserved α helices (Fig. 2A). The C-terminal end of helix α1 and its following loop are inserted into the CpG-containing major groove and make sequence-specific contacts with the CpG-containing region, while the W1 loop reaches deep into the neighboring minor groove to recognize bases flanking the CpG motif (Fig. 2A). In detail,

Arg[45] and Lys[46] at the C-terminal end of α1 are mainly responsible for CpG recognition in the major groove. The main chain carbonyl oxygen atoms of both residues form a hydrogen bond each with bases C6 and its symmetric related C7′, respectively (Fig. 2B). In addition, the side chain of Lys[46] forms hydrogen bonds with G7 from the CpG motif and its flanking base C8. The side chain of Arg[45] also forms a hydrogen bond with the phosphate backbone of T4. The major groove recognition is further strengthened by a hydrogen bond between Arg[56] and the phosphate backbone of C2 (Fig. 2, B and C). The minor groove recognition is mainly mediated by two long side-chain residues, Tyr[87] and Lys[88] from the W1 loop. The side chain of Tyr[87] hydrogen bonds with the G10′ base, while the side chain of Lys[88] forms hydrogen bonds with the bases C10 and T9′, respectively (Fig. 2D). The phosphate backbone connecting

Table 1. Data collection and refinement statistics (molecular replacement). Each dataset is collected from one crystal. RMS, root mean square.

	SAMD1-WH (27–105)/DNA Se-Met–labeled	SAMD1-SAM (459–523)	SAMD1-SAM (459–526)	SAMD1-SAM (459–530) Se-Met–labeled
PDB code	6LUI	6LUJ	6LUK	–
Data collection				
Space group	$P3_221$	$P3_121$	$P2_1$	$P3_2$
Cell dimensions				
a, b, c (Å)	45.93, 45.93, 132.64	69.34, 69.34, 181.89	66.43, 182.84, 66.97	67.75, 67.75, 293.39
α, β, γ (°)	90.00, 90.00, 120.00	90.00, 90.00, 120.00	90.00, 93.32, 90.00	90.00, 90.00, 120.00
Resolution (Å)	50.00–1.78 (1.81–1.78)*	50.00–1.12 (1.14–1.12)	50.00–2.06 (2.12–2.06)	50.00–2.90 (2.95–2.90)
R_{merge}	0.091 (0.797)	0.071 (0.560)	0.099 (0.520)	0.191 (1.262)
$I / \sigma I$	60.0 (3.1)	49.7 (2.3)	12.1 (1.9)	16.3 (1.7)
Completeness (%)	100.0 (100.0)	97.9 (81.1)	99.9 (100.0)	100.0 (100.0)
Redundancy	18.5 (16.8)	12.5 (5.9)	3.4 (3.4)	10.2 (10.4)
Refinement				
Resolution (Å)	25.47–1.78	23.85–1.12	45.71–2.06	–
No. of reflections	16,241	190,644	98,371	–
R_{work} / R_{free}	0.202/0.242	0.175/0.178	0.185/0.226	–
No. atoms				
Protein	598	2640	11,000	–
DNA	527	–	–	–
Ligand/ion	–	80	90	–
Water	103	715	833	–
B-factors (Å^2)				
Protein	37.1	11.4	26.5	–
DNA	42.0	–	–	–
Ligand/ion	–	23.3	43.4	–
Water	40.3	22.3	33.6	–
RMS deviations				
Bond lengths (Å)	0.007	0.004	0.002	–
Bond angles (°)	0.963	0.913	0.515	–

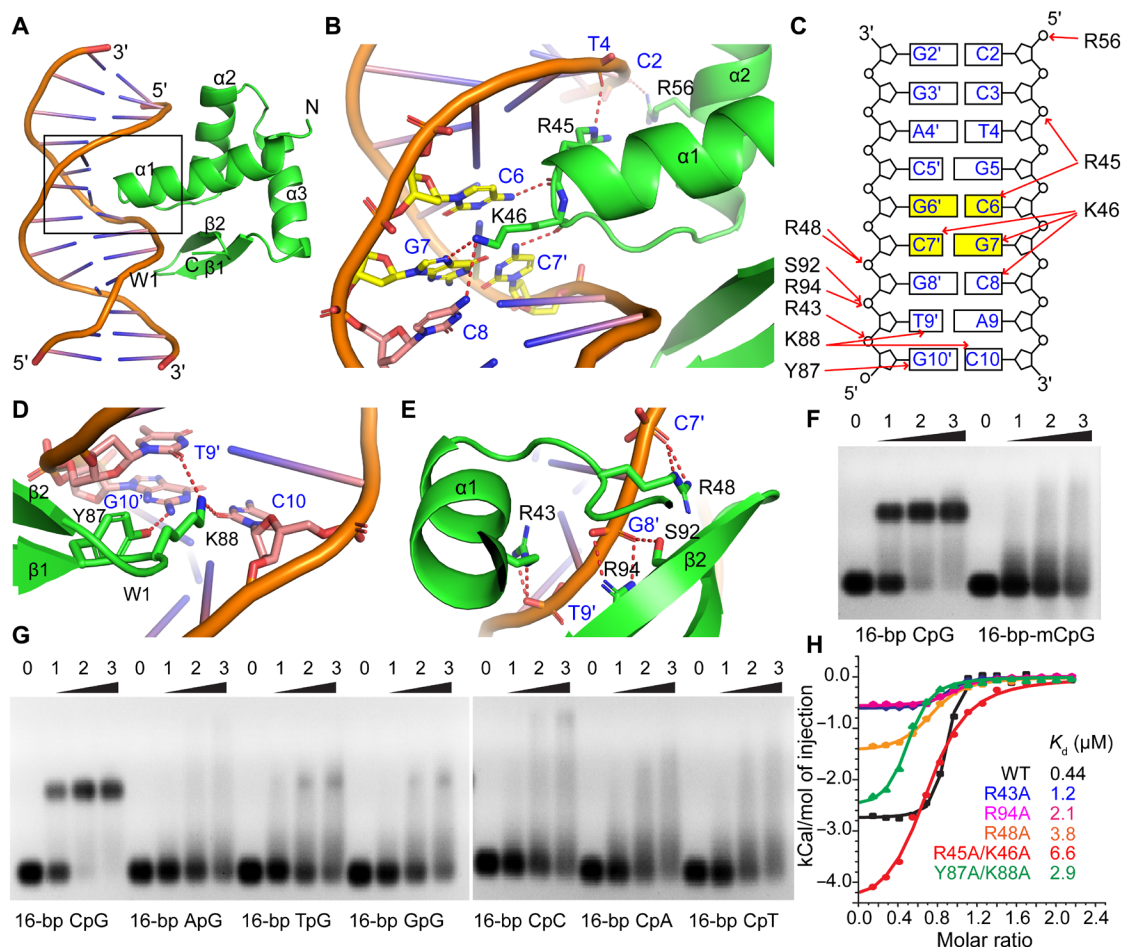*Values in parentheses are for highest-resolution shell.

Fig. 2. Structural and biochemical analysis of recognition between the SAMD1-WH domain and various DNA substrates. (**A**) An overall view of the SAMD1-WH domain/DNA complex [Protein Data Bank (PDB): 6LUI]. SAMD1-WH is colored in green with the secondary structural elements labeled. (**B**) A zoom-in view of the bases in the DNA major groove recognized by SAMD1-WH. (**C**) A schematic representation of the interaction network between SAMD1-WH and the target DNA. (**D** and **E**) Zoom-in views of the recognition in the minor groove (D) and the phosphate backbone (E) by SAMD1-WH. (**F** and **G**) EMSA analysis of the binding affinity of SAMD1-WH for a 16-bp CpG-containing DNA and its CpG-methylated (F) or substituted (G) counterpart. Numbers indicate molar ratio. (**H**) ITC measurements of the binding affinities for a 16-bp CpG-containing DNA by the WT SAMD1-WH domain and its mutants. Dissociation constants (K_d) are shown as inserts.

both the major and the minor grooves is also recognized. Arg[43] from α1 and Arg[48] from the loop following α1 each form a pair of hydrogen bonds with the phosphate backbone of T9′ and C7′, respectively (Fig. 2E). The side chains of Ser[92] and Arg[94] from β2 form one and a pair of hydrogen bonds with the phosphate backbone of G8′, respectively. Overall, SAMD1-WH recognizes the CpG-containing target DNA over a 9-bp footprint, with the major groove, the minor groove, and the phosphate backbone in-between both grooves extensively recognized (Fig. 2C).

The DNA binding mode is different from that shown for PCL proteins (5), as well as those exhibited by the other WH domains (17). Since three bases from the CpG motif are specifically recognized (Fig. 2, B and C), methylation of either cytosine, or replacement of either base from the CpG duplex would disrupt the interaction or cause a steric clash with the main-chain backbone of the SAMD1-WH domain. Consistently, through EMSA analysis, we found that SAMD1-WH showed a marked loss of binding affinity toward the DNA substrates when the cytosines of the CpG motif are fully methylated (Fig. 2F). Furthermore, replacement of either the

cytosine or the guanine of the CpG motif in the substrate DNA resulted in a marked loss of binding affinity by SAMD1-WH (Fig. 2G). In contrast, replacement of single bases flanking either side of the CpG motifs did not substantially lower the binding affinity of the SAMD1-WH domain (fig. S3), further confirming the importance of an unmethylated CpG motif for recognition. Through isothermal titration calorimetry (ITC) measurements, we found that wild-type (WT) SAMD1-WH binds a 16-bp CpG-containing DNA with a dissociation constant (K_d) of 0.44 μM (Fig. 2H and Table 2). The double mutant R45A/K46A that disrupts the recognition in the DNA major groove displayed a 15-fold weaker binding affinity (K_d = 6.6 μM). A double mutation Y87A/K88A that disrupts recognition at the minor groove decreased the binding for the target DNA by 6.6-fold. R43A, R94A, and R48A mutations that disrupt recognition of the phosphate backbone between the major and minor grooves reduce the binding affinity of SAMD1-WH to the target DNA by 2.7-, 4.8-, and 8.6-fold, respectively (Fig. 2H and Table 2). This indicates that in addition to the CpG motif, recognition of the minor groove and the phosphate backbone between the

Table 2. ITC-based measurements of K_d between the SAMD1-WH domain or its mutants with 16-bp CpG DNA.

DNA	Protein sample	K_d (µM)	ΔH (kcal/mol)
16-bp CpG	SAMD1-WH(16–110)	0.44 ± 0.11	−2.75 ± 0.04
16-bp CpG	SAMD1-WH(16–110)-R45A/K46A	6.6 ± 0.9	−4.60 ± 0.14
16-bp CpG	SAMD1-WH(16–110)-Y87A/K88A	2.9 ± 0.5	−2.60 ± 0.08
16-bp CpG	SAMD1-WH(16–110)-R43A	1.2 ± 0.2	−0.62 ± 0.01
16-bp CpG	SAMD1-WH(16–110)-R48A	3.8 ± 0.4	−1.48 ± 0.10
16-bp CpG	SAMD1-WH(16–110)-R94A	2.1 ± 0.4	−0.58 ± 0.02

major and minor grooves also play important roles for high binding affinity of SAMD1-WH domain to CpG-containing DNA.

SAMD1 acts as a repressor at active CGIs

To gain insights into the potential function of SAMD1 at CGIs, we compared the genomic binding pattern of SAMD1 with those of other CpG-binding proteins, such as MTF2 (PCL2) (5), KMT2B (MLL2) (18), KDM2A (FBXL11), KDM2B (FBXL10) (4), and CXXC1 (CFP1) (3) and related histone modifications. Using correlation analysis, we found that SAMD1 clusters together with CXXC1, KMT2B, KDM2A, KDM2B, and H3K4me3 but is distant from MTF2 and H3K27me3 (Fig. 3A). To investigate this further, we categorized all CGIs into active (H3K4me3 only), repressed (H3K27me3 only), bivalent (H3K4me3 and H3K27me3), and undecorated (neither) CGIs. Comparison of the respective heatmaps revealed that the binding pattern of SAMD1 is most similar to CXXC1 and KMT2B, which selectively bind to CGIs that have the active histone mark H3K4me3, while KDM2A and KDM2B are more broadly distributed across all CGI categories (Fig. 3B and fig. S4A). SAMD1 overlaps with the PRC2-associated factor MTF2 only at bivalent CGIs but not at repressed CGIs. Overall, SAMD1-bound genes are predominantly highly expressed (fig. S4B) and belong to many distinct biological processes, such as transcription, translation, cell cycle, intracellular transport, and development (fig. S4C).

To gather information about the functional role of SAMD1 at those genes, we performed RNA sequencing (RNA-seq) in SAMD1 KO versus WT ES cells. This experiment identified 524 significantly ($P < 0.01$) down-regulated and 257 up-regulated genes (Fig. 3C). Further investigation showed that the up-regulated but not the down-regulated genes are strongly occupied by SAMD1 (Fig. 3, D and E), suggesting that direct targets of SAMD1 become derepressed upon SAMD1 deletion. We confirmed via gene set enrichment analysis (GSEA) that the 100 genes with the highest levels of SAMD1 are, on average, significantly up-regulated upon SAMD1 KO (Fig. 3F), further supporting a repressive role for SAMD1. Gene Ontology analysis of the genes that are up-regulated and bound by SAMD1 showed that they are related to transcription, cell division, chromatin remodeling, and developmental processes (fig. S4D). These genes are generally highly expressed and rich in H3K4me3 but lack H3K27me3 modifications (Fig. 3, G and H), suggesting that SAMD1 restricts the expression of a subset of highly active genes.

SAMD1 interacts with L3MBTL3 and the KDM1A complex

To determine the molecular mechanism of the repressive function of SAMD1, we purified SAMD1 from HeLa-S cells and identified SAMD1-associated proteins by mass spectrometry (Fig. 4A). Consistent with previous reports, we found that SAMD1 associates with L3MBTL3, SFMBT1, and SFMBT2 (11, 19), as well as the KDM1A complex (12), which demethylates the active H3K4me2 mark (20). Using coimmunoprecipitation experiments after ectopic overexpression, we validated the interaction of SAMD1 with L3MBTL3 and KDM1A (Fig. 4B). The endogenous interaction between SAMD1 and L3MBTL3 could also be confirmed in mouse ES cells (fig. S5A).

SAMD1, L3MBTL3, SFMBT1, and SFMBT2 all have a SAM domain at their C termini. This domain is also present in several Polycomb-related proteins, such as the PHC proteins (fig. S5B), and is important for protein-protein interactions by forming polymers, which contribute to the formation of Polycomb bodies (21, 22). To address the interplay of SAMD1 with other SAM domain–containing proteins in more detail, we performed mammalian-two-hybrid experiments and examined the association of distinct SAM domains (Fig. 4C, top, and fig. S5B). We used the SAM domain of the Polycomb protein PHC1 as a positive control, which is known to interact with several other SAM domain proteins (21, 22). We found that the SAM domain of SAMD1 specifically interacts with the SAM domain of L3MBTL3, L3MBTL4, and itself but not with the other investigated SAM domains (Fig. 4C, bottom). These SAM-SAM interactions can be disrupted by mutating critical residues of the SAMD1-SAM domain (fig. S5, C and D). Via coimmunoprecipitation, we confirmed the self-association feature of SAMD1 and validated that this interaction requires an intact SAM domain (fig. S5, E and F). The SAM domain of PHC1 interacts with the SAM domains of several proteins involved in Polycomb repression but not with SAMD1 or L3MBTL3/4 (Fig. 4C, bottom), suggesting a restricted selection of interacting partners among SAM domains. Together, these results support that L3MBTL3 associates with SAMD1 through direct SAM-SAM domain interaction.

As KDM1A appeared to be a strong SAMD1 interactor (Fig. 4, A and B), we also studied the association between SAMD1 and KDM1A. Previous work suggests that L3MBTL3 interacts with KDM1A (19). Thus, SAMD1 may associate with KDM1A either directly or indirectly through L3MBTL3. To test these possibilities, we created different SAMD1 deletion and point mutants (Fig. 4D). Deletion or mutation of the SAM domain strongly reduced the interaction with L3MBTL3 (Fig. 4E) but not the association with KDM1A (Fig. 4F), supporting that the SAMD1-L3MBTL3 association is not required for the interaction between SAMD1 and KDM1A. By contrast, deletion of the WH domain reduced the association with KDM1A but not with L3MBTL3 (Fig. 4, E and F), suggesting that SAMD1 interacts with KDM1A and L3MBTL3 through distinct binding sites.

SAMD1 modulates the chromatin landscape at CGIs

We next addressed how SAMD1, KDM1A, and L3MBTL3 may cooperate at CGIs. First, we investigated the binding pattern of L3MBTL3 and KDM1A relative to that of SAMD1. For L3MBTL3, we generated two polyclonal antibodies directed against the N- or C terminus of L3MBTL3 and determined its genome-wide binding pattern via ChIP-seq. The N-terminal antibody led to more significant peaks ($n = 576$) and was used for this initial analysis. For KDM1A, we used publicly available ChIP-seq data (23). Analysis of the data revealed that in mouse ES cells, L3MBTL3 mainly localizes to
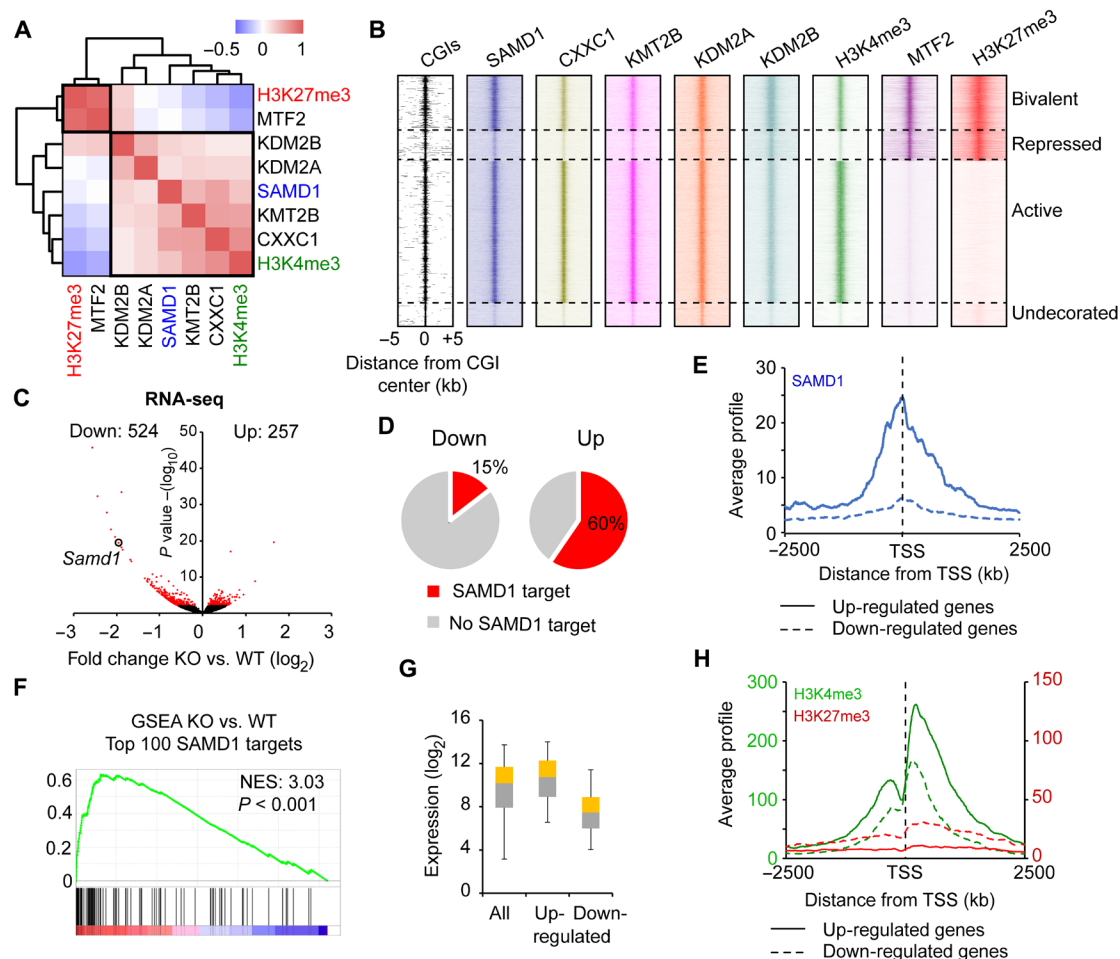
Fig. 3. Investigation of SAMD1's role at CGIs. (**A**) Correlation analysis of SAMD1, KMT2B, KDM2A, KDM2B, CXXC1, MTF2, H3K4me3, and H3K27me3 at CGIs. (**B**) Heatmap showing the distribution of the factors from (A) at distinct CGIs. Based on H3K4me3 and H3K27me3 levels, the CGIs were divided into bivalent (H3K4me3 and H3K27me3), repressed (only H3K27me3), active (only H3K4me3), or undecorated (neither) CGIs. (**C**) Volcano plot of RNA sequencing (RNA-seq) data of SAMD1 KO versus WT cells. Cutoff: $P < 0.01$. Four biological replicates were performed. (**D**) Occupancy of up- and down-regulated genes with SAMD1. (**E**) Promoter profile of SAMD1 at up- and down-regulated genes. (**F**) Gene set enrichment analysis (GSEA) of the top 100 SAMD1-bound genes, using the RNA-seq data. (**G**) Expression level of SAMD1-bound up- and down-regulated genes in comparison to all SAMD1-bound genes. The whisker-box plots represent the lower quartile, median, and upper quartile of the data with 5 and 95% whiskers. (**H**) Promoter profiles of H3K4me3 and H2K27me3 at up- and down-regulated genes. NES, Normalized Enrichment Score.

CGIs (Fig. 5A), with about 57% of its binding sites overlapping with SAMD1. In contrast, KDM1A binds to many non-CGI locations, such as enhancers, that are hardly targeted by SAMD1. However, most of the SAMD1 binding sites (>95%) are also bound by KDM1A (Fig. 5B). At the four CGI categories established above (Fig. 3B), SAMD1, KDM1A, and L3MBTL3 are similarly distributed (fig. S6A), demonstrating that all three proteins preferentially bind to actively transcribed genes. Together, these results suggest that most SAMD1-targeted CGIs are also targeted by KDM1A, while around half of the L3MBTL3 binding sites are cobound by SAMD1.

To determine whether SAMD1 influences the binding of L3MBTL3 and KDM1A to chromatin, we performed ChIP–quantitative polymerase chain reaction (qPCR) experiments at selected SAMD1 target genes in SAMD1 KO cells. SAMD1 deletion strongly reduced the chromatin binding of L3MBTL3 (Fig. 5C), suggesting that L3MBTL3 binding to chromatin might substantially depend on SAMD1 at those locations. Given that L3MBTL3 is related to the

Drosophila Polycomb group protein L(3)mbt (*24*), we also determined the influence on the classical Polycomb group proteins, but we did not observe any obvious consequences on the chromatin binding of those proteins in the absence of SAMD1 (fig. S6B), supporting that SAMD1 is not affecting the canonical Polycomb system. In contrast to the strong reduction of L3MBTL3, the KDM1A levels were only partially reduced upon SAMD1 deletion (Fig. 5C). At a genome-wide level, we confirmed a strong reduction of the L3MBTL3 levels, particularly at CGIs with strong SAMD1 binding (Fig. 5, D and F), while at CGIs where SAMD1 is not present the L3MBTL3 level is not affected (Fig. 5D). This effect can be observed with both L3MBTL3 antibodies (Fig. 5F). For KDM1A, only a subtle reduction can be detected at a genome-wide level (Fig. 5, E and F). These results suggest that SAMD1 is involved in tethering L3MBTL3 to chromatin at SAMD1 binding sites, while SAMD1 may contribute only marginally to KDM1A recruitment. Given that KDM1A interacts with many proteins
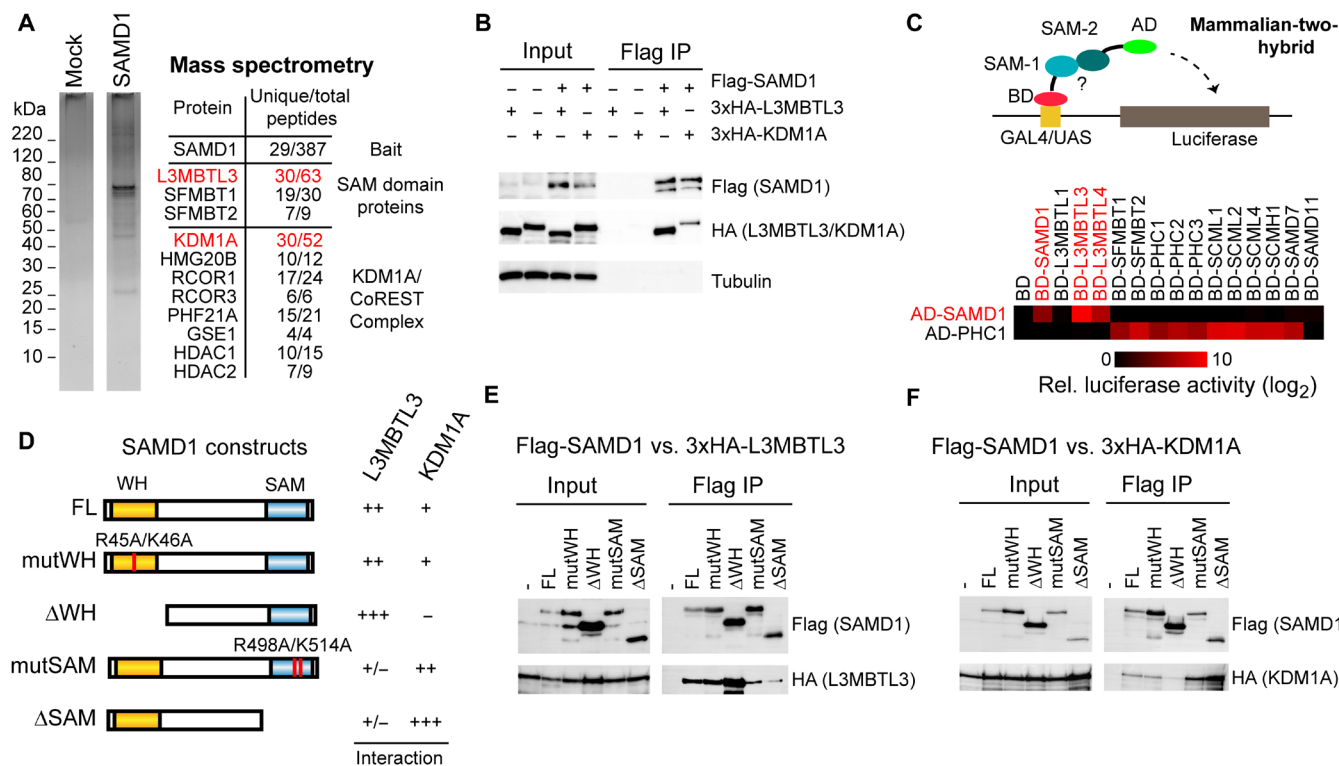
Fig. 4. Identification of SAMD1 interaction partners. (**A**) Tandem affinity purification of Flag-HA-SAMD1 from Hela-S cells followed by mass spectrometry analysis. Shown are unique and total peptide numbers. (**B**) Coimmunoprecipitation experiment in human embryonic kidney–293 cells, demonstrating the interaction of SAMD1 with L3MBTL3 and KDM1A. (**C**) Schematic representation and results of mammalian-two hybrid using various SAM domains (see also fig. S5B). (**D**) Overview of constructs used and the results from mapping experiments in (E) and (F). (**E**) Coimmunoprecipitation of HA-L3MBTL3 with distinct versions of Flag-SAMD1. (**F**) Coimmunoprecipitation of HA-KDM1A with distinct versions of Flag-SAMD1. HDAC, histone deacetylase.

(*12*), the recruitment of KDM1A is likely facilitated also by other chromatin-binding factors.

To assess the contribution of L3MBTL3 and KDM1A to gene repression by SAMD1, we created L3MBTL3 and KDM1A KO cells using CRISPR-Cas9 and performed RNA-seq experiments. Via GSEA, we found that the top SAMD1 targets become up-regulated upon KDM1A but not L3MBTL3 KO (Fig. 5G), suggesting that KDM1A may cooperate with SAMD1 for regulating gene expression, while L3MBTL3 may be less relevant for this aspect. Given that KDM1A demethylates the active H3K4me2 histone mark (*20*), we analyzed the consequence of SAMD1 KO on H3K4me2. We discovered a subtle but significant increase of H3K4me2 at CGIs (Fig. 5H and fig. S6C), which is similar to the consequences observed upon KDM1A knockdown or its chemical inhibition (*25*, *26*). The increase is particularly evident at CGIs with robust SAMD1 binding, suggesting that SAMD1 deletion impairs the function of the KDM1A histone demethylase complex at those CGIs. The H3K4me2 level is significantly reduced at enhancer sites (fig. S6D), where SAMD1 is barely present. This reduction is likely an indirect effect and may explain why more genes become down-regulated than up-regulated upon SAMD1 deletion (Fig. 3C). H3K4me2 serves also as substrate for histone methyltransferases that deposit H3K4me3. Consistently, we also observed a subtle increase of H3K4me3 at CGIs (Fig. 5I and fig. S6E). Given that the KDM1A complex is also associated with histone deacetylases (HDACs), and HDACs were identified as interacting partners by IP-MS (Immunoprecipitation followed by mass-spectrometry)

(Fig. 4A), we asked whether SAMD1 deletion would affect histone acetylation. Via ChIP-seq, we observed no major changes for the histone acetylation mark H3K27ac (fig. S6F), suggesting that SAMD1 does not alter the function or recruitment of those enzymes. Similarly, we found only minimal consequences on H3K27me3 (fig. S6G), supporting the view that SAMD1 is not directly influencing the chromatin regulation by the canonical Polycomb proteins. Together, our results suggest that SAMD1 directly binds to unmethylated CGIs and modulates gene transcription by influencing the function of KDM1A and possibly other chromatin regulators.

SAMD1 requires both the WH and the SAM domain for efficient chromatin binding

To determine the recruitment mechanism of SAMD1 to CGIs in vivo, we performed ChIP-qPCR experiment in SAMD1 KO cells in which SAMD1 expression was restored by ectopically expressed WT or mutated SAMD1 (Fig. 6A). We found that mutation of either the WH domain or the SAM domain reduced the chromatin association of SAMD1 (Fig. 6B). The fact that mutations of the WH domain affect chromatin binding is consistent with our in vitro data showing that this domain is directly involved in DNA binding. This is also in line with what has been shown for the WH domain of MTF2 (*5*). Regarding the SAM domain, we speculated that the SAM domain could mediate the interaction with other SAM domain-containing proteins, such as L3MBTL3, which in turn might contribute to the chromatin association of SAMD1. Given that L3MBTL3
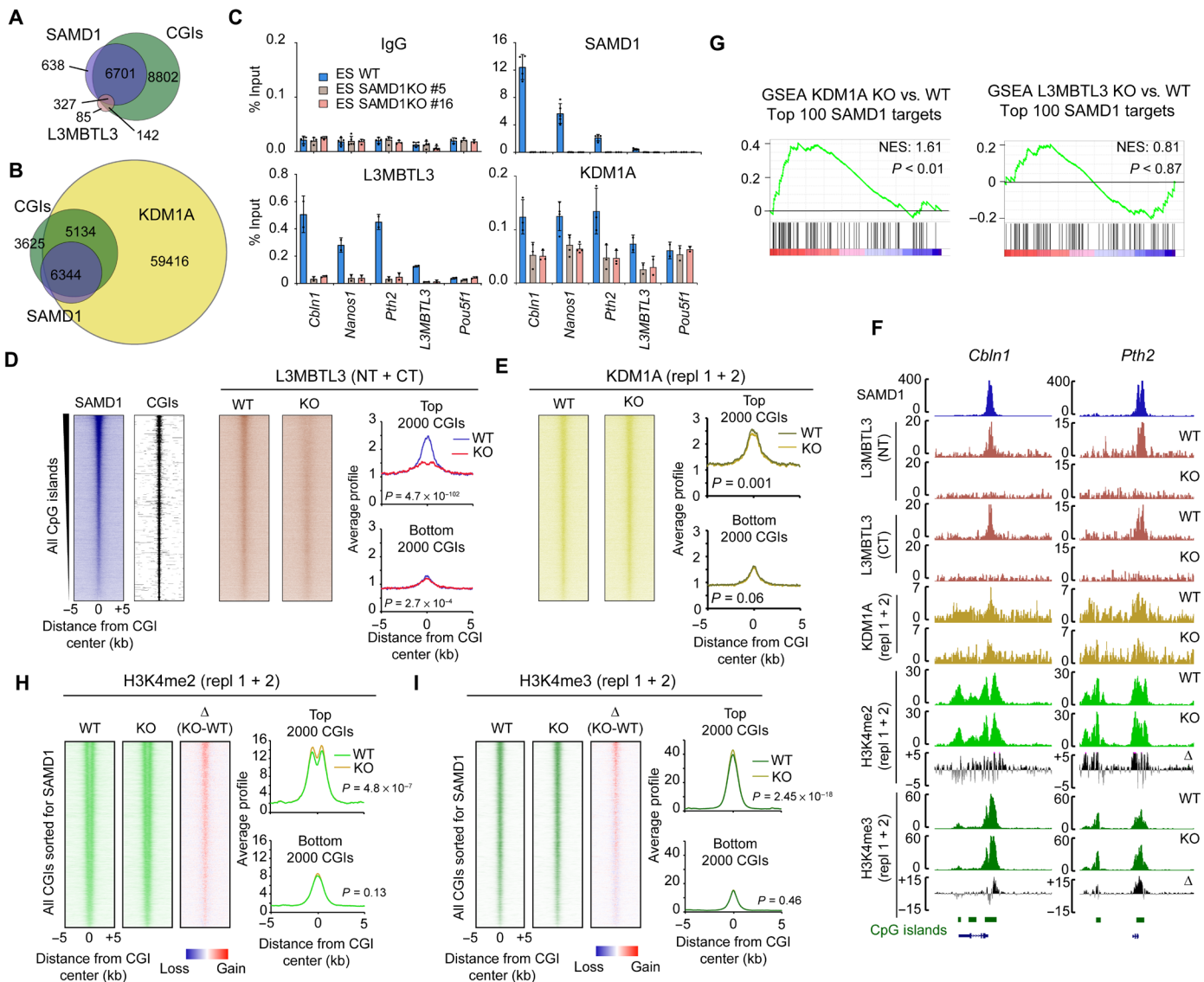
Fig. 5. Chromatin regulation by SAMD1. (**A**) Venn diagram showing the genome-wide overlap of SAMD1 peaks with L3MBTL3 peaks and CGIs (**B**) Overlap of SAMD1 with KDM1A and CGIs. (**C**) ChIP–quantitative polymerase chain reaction (qPCR) experiment for L3MBTL3 and KDM1A in WT and SAMD1 KO cells. (**D**) Heatmap of ChIP-seq experiments for L3MBTL3 in WT and SAMD1 KO cells at all CGIs, sorted after SAMD1 levels. The ChIP-seq results using two distinct L3MBTL3 antibodies were merged. The profiles at the top 2000 and bottom 2000 SAMD1-bound CGIs are shown. (**E**) Heatmap and profiles of ChIP-seq experiments using KDM1A antibody in WT versus SAMD1 KO cells, sorted as in (D). Two replicates for KDM1A were merged. (**F**) Genome browser view of ChIP-seq experiments at two SAMD1 targets. (**G**) GSEA analysis of the top 100 SAMD1 genes in KDM1A (left) and L3MBTL3 (right) KO cells, compared to WT cells. (**H**) Heatmap and profiles of H3K4me2 at CGI, sorted as in (D), in WT and SAMD1 KO cells. The difference is shown in the right heatmap. Two biological replicates were merged. See also fig. S6C. (**I**) Heatmaps and profiles as described in (H) but for H3K4me3. See also fig. S6E. *P* values in (D), (E), (H), and (I) were calculated by were calculated by two-sided Student's *t* tests. NT, N terminus. CT, C terminus.

can bind to chromatin via its MBT domains (*27*), we hypothesized that L3MBTL3 may be relevant for SAMD1 chromatin binding. Unlike the strong effect of an almost complete loss of chromatin association of L3MBTL3 upon SAMD1 KO (Fig. 5C), we found that L3MBTL3 deletion only moderately reduced the association of SAMD1 with chromatin (fig. S7A). This reduction of SAMD1 chromatin binding may also be due to a slightly reduced protein level of SAMD1 in the L3MBTL3 KO cells (fig. S7B). Thus, the association with L3MB-TL3 cannot explain the importance of the SAM domain integrity for SAMD1 chromatin association. We speculated therefore that

SAMD1 may increase its chromatin-binding affinity via SAM homopolymerization, similar to other SAM domain proteins (*21*, *22*, *28*).

The SAMD1-SAM domain homopolymerizes into a pentameric ring

To explore the self-association mechanism of the SAMD1-SAM domain, we crystallized two SAMD1-SAM–containing constructs and solved their structures at the resolution of 1.12 and 2.06 Å (Table 1), respectively. The shorter construct of the SAM domain self-associates into a pentamer (Fig. 6C), while in the structure of the longer construct,
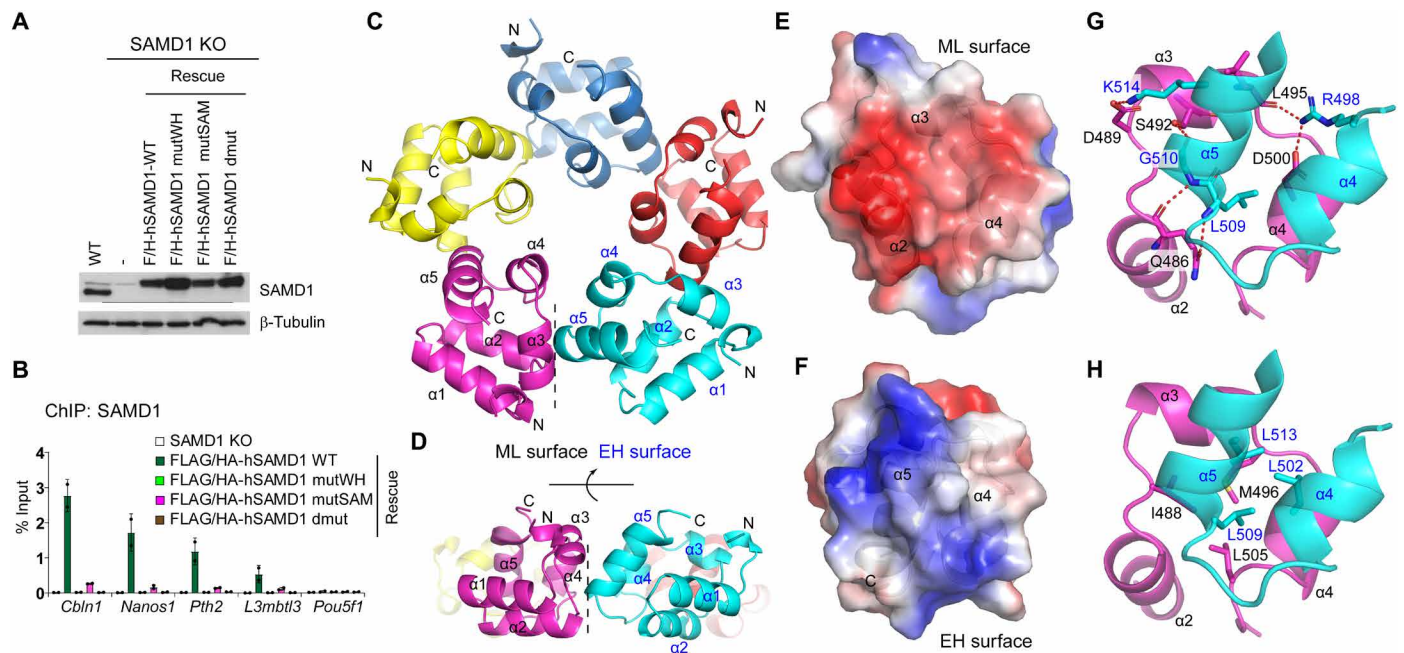
Fig. 6. Structural details of the SAMD1-SAM pentamer. (**A**) Western blot of ectopically expressed SAMD1 in SAMD1 KO ES cells. (**B**) ChIP-qPCR of SAMD1 in rescued cells at SAMD1 target genes. Two biological replicates were performed. Error bars indicate ± SD. (**C** and **D**) Top (C) and side (D) views of the SAMD1-SAM pentamer (PDB: 6LUJ). Five molecules are colored in different colors, with the secondary elements of the SAM domain labeled. The interface between mid-loop (ML) and end-helix (EH) surfaces is indicated by a dotted line. (**E** and **F**) Electrostatic potentials of the ML (E) and EH (F) surfaces. Negative electrostatic potential is colored in red, while the positive electrostatic potential is colored in blue. (**G** and **H**) Details of hydrogen bonding interactions (G) and hydrophobic interactions (H) between ML surface colored in magenta and EH surface colored in cyan.

two pentamers stack together into a decamer (fig. S7C). Gel filtration and mass spectrometry analysis confirmed that WT SAMD1-SAM, but not its mutated version, forms a pentamer in solution (fig. S7, D and E), indicating that the pentameric state is a stable form of the SAMD1-SAM polymer. In the structure of the pentameric SAMD1-SAM polymer, five molecules associate with one another to form a donut-shaped closed ring (Fig. 6C), which is different from the spirally associated SAM domain polymers, such as those of the *Drosophila* Ph-SAM and human Translocated ETS leukemia (TEL)–SAM domains (*29*, *30*). SAMD1-SAM adopts the fold of a typical SAM domain in which five α helices fold into a compact globular structure (Fig. 6, C and D). Similar to the canonical SAM domain polymers (*31*), the SAMD1-SAM pentamer is stabilized through close contacts between the mid-loop (ML) surface of one molecule and the end-helix (EH) surface of a neighboring molecule (Fig. 6, C and D). The ML surface—formed by α helices 2, 3, and 4—exhibits a negative electrostatic potential (Fig. 6E), which is complemented by the EH surface, that is composed of helices 4 and 5 and displays a positive electrostatic potential (Fig. 6F). The interactions between ML and EH surfaces are mediated by both polar and nonpolar interactions. Arg[498] and Lys[514] from EH surface each form a salt bridge with Asp[500] and Asp[489] from the ML surface (Fig. 6G), respectively. Arg[498] also forms a hydrogen bond with Leu[495] from ML surface. Gly[510] from EH surface forms a hydrogen bond each with Ser[492] and Gln[486] from ML surface. The side chain of Gln[486] also hydrogen bonds with the main chain of Leu[509] (Fig. 6G). Intermolecular contacts between two surfaces are further stabilized by hydrophobic interactions, with the long hydrophobic side chains of Leu[502], Leu[509], and Leu[513] from the EH surface pointing toward the hydrophobic patch

composed of Ile[488], Met[496], and Leu[505] from the ML surface (Fig. 6H). Together, these findings demonstrate an unusual self-association of the SAMD1-SAM domain and supports that SAMD1 possibly can interact with CGIs in a multivalent manner. Given that the SAMD1-SAM pentamer is distinct to other nuclear SAM domain polymers (*29*, *30*), it may facilitate unknown chromatin-related functions, which warrants further investigation in the future.

SAMD1 is required for normal ES cell differentiation

Currently, not much is known about the biological role of SAMD1. To address the potential biological function of SAMD1, we performed undirected ES cell differentiation, via leukemia inhibitory factor (LIF) removal, in the presence and absence of SAMD1 and investigated the consequence on the differentiation process. The expression of SAMD1 itself remains largely constant during the differentiation (Fig. 7A). Observation of the cells showed no obvious differences between WT and SAMD1 KO cells, suggesting that the absence of SAMD1 does not impair the general differentiation process. Consistently, in reverse transcription (RT)–qPCR experiments, we observed only minor differences of classical pluripotency and differentiation markers (Fig. 7B). However, we observed that SAMD1-targeted genes, which are slightly up-regulated in SAMD1 KO cells in the undifferentiated state, remain higher expressed in the KO cells throughout the differentiation process (Fig. 7B). To gain deeper insights into the role of SAMD1, we performed RNA-seq after 7 days of differentiation. Principal components analysis (PCA) showed that the difference in the expression pattern between WT and SAMD1 KO cells becomes larger upon differentiation (Fig. 7C), demonstrating that the absence of SAMD1 impairs the differentiation. We
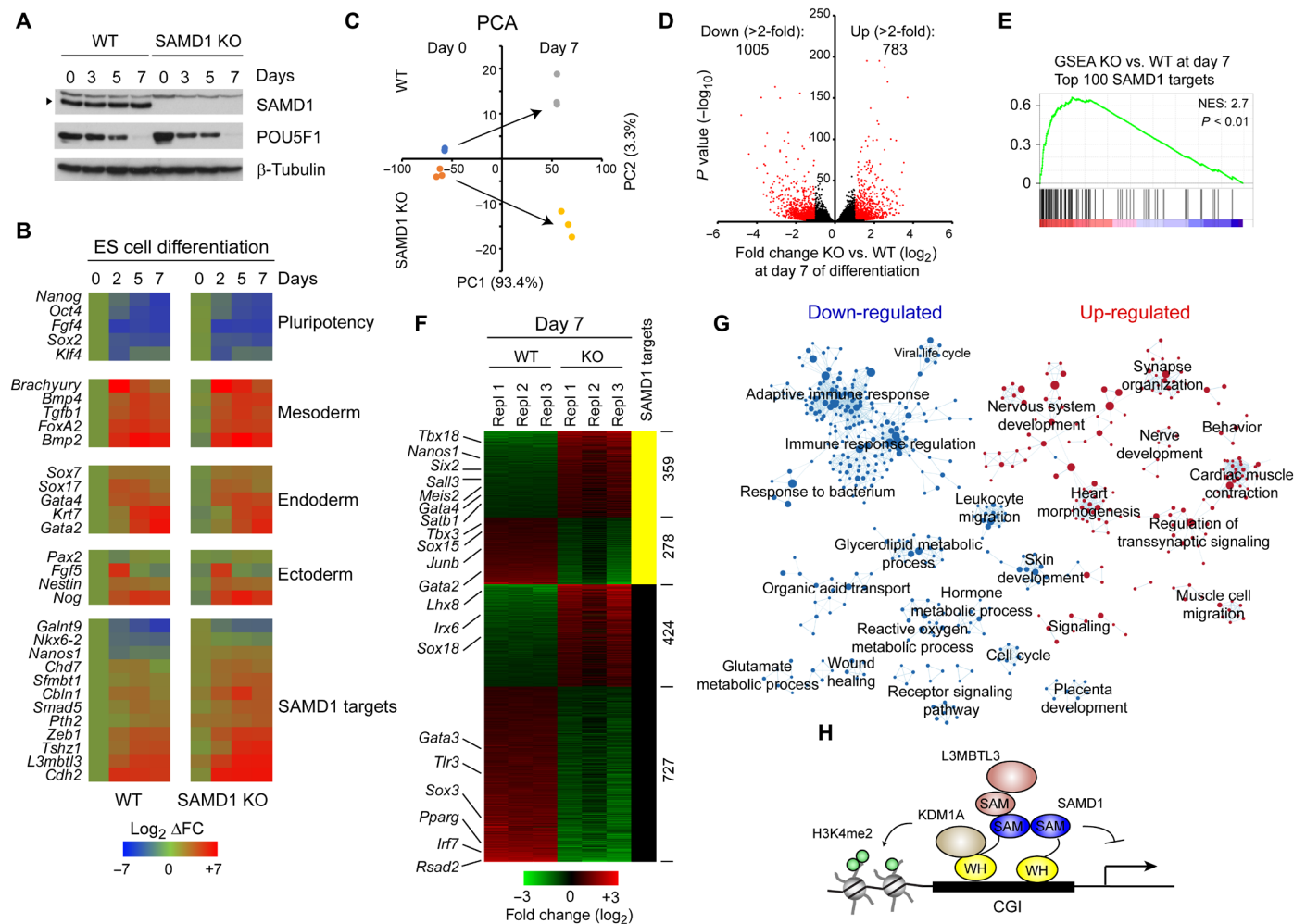
Fig. 7. Investigation of the role of SAMD1 during ES cell differentiation. (**A**) Western blot of SAMD1 and POU5F1 during undirected ES cell differentiation by LIF removal. (**B**) RT-qPCR experiments of stem cell and differentiation markers, as well as SAMD1 target genes during the differentiation. Three biological replicates were performed. (**C**) PCA analysis of RNA-seq data at days 0 and 7 of ES cell differentiation. Three biological replicates were performed. (**D**) Volcano plot of RNA-seq data after 7 days of differentiation in WT and SAMD1 KO cells. In red are genes that are significant ($P < 0.01$) and at least twofold differentially expressed. (**E**) GSEA of top 100 SAMD1 target genes, comparing SAMD1 KO versus WT cells at day 7 of the differentiation. (**F**) Genes from (D) shown has heatmap and sorted for SAMD1-bound and unbound genes. (**G**) GSEA followed by network analysis, demonstrating up- and down-regulated pathways in SAMD1 KO cells after 7 days of differentiation. The circle radius indicates the size of the gene sets. (**H**) Model of putative SAMD1 function at CGIs.

identified more than 5000 genes of which the expression changes significantly ($P < 0.01$), including more than 1500 genes that are more than twofold misregulated (Fig. 7D). Consistent with the RT-qPCR experiment, the top 100 SAMD1 targets remain significantly up-regulated in SAMD1 KO cells (Fig. 7E). Also, the up-regulated, but not the down-regulated, genes are typically occupied by SAMD1 (Fig. 7F and fig. S8A), suggesting that SAMD1 restricts the expression of its target genes throughout the differentiation process. To gain deeper insights into the pathways regulated by SAMD1, we performed unbiased GSEA followed by network analysis. This analysis demonstrates that SAMD1 deletion leads to an up- and down-regulation of numerous biological processes (Fig. 7G). Specifically, pathways related to immune system and metabolism become down-regulated in SAMD1 KO cells (fig. S8B), while pathways related to neuronal function and heart muscle cells become up-regulated (fig. S8C). Key regulators such as *Sall3*, *Nanos1*, *Gata4*, and *Pparg* are

strongly differentially expressed in SAMD1 KO versus WT cells (Fig. 7F). Together, these analyses demonstrate that SAMD1 is required for proper ES cell differentiation, and it suggest that SAMD1 plays a pleiotropic biological role, similar to the role of KDM1A (*32*). Future investigations of the role of SAMD1 during specific differentiation processes and embryogenesis will shed further light on its biological importance.

DISCUSSION

In this study, we uncovered a regulatory mechanism that links unmethylated CGIs to chromatin regulation, mediated by SAMD1. Our in vivo and in vitro data robustly demonstrate that SAMD1 directly interacts with unmethylated CpG motifs and links key chromatin modifiers, such as KDM1A and L3MBTL3, to CGIs. Our findings highlight a previously unknown regulatory pathway at

CGIs and open avenues of investigation in chromatin regulation and translational fields.

In mammals, a large proportion of RNA polymerase II promoters have CGIs (1) but lack other classical promoter elements such as the TATA box. The chromatin environments of the CGIs are largely modulated by CpG motif–binding proteins that recruit or are themselves part of the histone modifying complexes. Our study establishes SAMD1 as a new player in CGI regulation that directly binds to unmethylated CGIs. Several features distinguish SAMD1 from those previously identified unmethylated CpG-binding proteins. First, the SAMD1 WH domain is structurally different from the CXXC domains (Fig. 2), which are well-documented unmethylated CpG-binding domains that adopt a special fold through coordinating two zinc ions (33). Second, although the SAMD1 WH domain also exhibits a WH fold, its DNA recognition mode is completely different from those shown by the WH domains of PCL proteins (5). The PCL WH domain recognizes only the major groove of the unmethylated CpG-containing DNA through the W1 loop whose sequence is conserved only among different PCL proteins. By contrast, the SAMD1 WH domain recognizes the major groove of the CpG-containing DNA through its first α helix, with its W1 loop recognizing the minor groove of the DNA. Last, the SAMD1 WH and PCL WH domains have different DNA sequence preferences (GCGC versus TCGG) (5). Through sequence analysis, we found that the SAMD1 WH domain and the WH domains from several other proteins represent a new subgroup of WH domains that have the potential to recognize unmethylated CpG motifs (fig. S1, C and D). This substantially expands the repertoire of the unmethylated CpG-binding proteins. Given the diversity of the unmethylated CpG-containing motifs and the various binding modes that the CpG-binding domains exhibited, it is quite possible that there are other uncharacterized unmethylated CpG motif–binding domains, which await identification.

Our further work found that SAMD1 requires not only the DNA-recognizing WH domain but also the self-polymerizing SAM domain for efficient chromatin binding. SAM domains are one of the most abundant protein-protein interaction motifs in eukaryotes that can mediate complex formation through homo- or heterodimerization/polymerization (21, 31). In *Drosophila*, the canonical PRC1 is tethered to the DNA binding Pho-repressive complex through SAM-SAM interactions (28). In this study, SAMD1 was also found to interact with several SAM domain-containing proteins through SAM-SAM interactions. Although the physiological function of this interaction remains to be explored, it does provide an explanation for the diverse regulatory function of SAMD1. Unexpectedly, interactions with proteins, such as the interaction with L3MBTL3, contribute only marginally to the recruitment of SAMD1, which prompted us to assess the self-oligomerization feature of the SAMD1-SAM domain. The SAMD1-SAM domain self-associates into a pentameric circle in solution (Fig. 6). This pentamer structure is homogeneous and stable as verified by mass spectrometry. The SAMD1-SAM pentamer is the second ring-shaped SAM domain structure formed by self-association identified in higher eukaryotes and the other being the recently described octameric ring of the SARM1 SAM domain (34); previously described SAM domain structures in eukaryotes tend to form an open-ringed polymer. Oligomerization of the SAMD1 through its SAM domain would most likely enhance the avidity of its WH binding to the target DNAs, thus enhancing its recruitment to the chromatin. Protein oligomerization, including SAM polymerization, has also been shown to play an important role in

phase separation, a process that has been observed for many chromatin regulators, and which contributes to chromatin organization (21, 22). A similar mechanism may also apply to SAMD1. Future work is required to determine whether SAMD1 regulates chromatin structure in vivo.

In the genome of mouse ES cells, SAMD1 is preferentially associated with active CGIs where it cooperates with chromatin regulators, such as KDM1A, to modulate the chromatin landscape and down-regulate its target genes (Figs. 3 to 5). The interplay of SAMD1 with KDM1A at CGIs may contribute to the role of KDM1A at CpG-rich regions in sperm cells (35). The repressive function of SAMD1 places SAMD1 between the CGI-binding proteins that activate transcription, such as CXXC1 (3) and the PCL proteins (5), which represses transcription at Polycomb target genes. SAMD1 binds to over 7000 genes in mouse ES cells, involved in numerous biological pathways (fig. S4C), demonstrating that SAMD1 has a global gene regulatory function. The consequence of loss of function of SAMD1 in undifferentiated ES cells on gene expression is rather minor (Fig. 3C), similar to what is observed upon loss of function of many other DNA binding factors (36). However, upon ES cell differentiation, SAMD1 KO cells exhibited a substantial alteration of their differentiation program, leading to the up- and down-regulation of numerous genes (Fig. 7D). Those genes are involved in many biological pathways (Fig. 7G), supporting a versatile role of SAMD1, and a potential pleiotropic biological function. Notably, SAMD1 is commonly up-regulated in many cancer types (fig. S8D) (37) and its high expression is often associated with a poorer prognosis in patients with cancer, such as in liver cancer and acute myeloid leukemia (fig. S8E), suggesting that SAMD1 could play a role in multiple cancer types. Moreover, SAMD1 has recently been found to be essential for the growth of the K562 erythroleukemic cells (38), supporting a putative protumorigenic function of SAMD1 in this context.

In summary, we have identified SAMD1 as a CGI-binding protein that links H3K4me3-decorated CGIs in mouse ES cells to gene repression. We have provided structural insight into SAMD1's CGI recognition and homopolymerization, which revealed a previously undiscovered mode of binding to CGIs. We provide evidence that SAMD1 acts in concert with the KDM1A demethylase complex, L3MBTL3, and most likely other factors to modulate the chromatin landscape at its target CGIs (Fig. 7H). Future investigations of SAMD1 in a physiological and pathophysiological context will provide further details about the mechanistic and biological function of SAMD1.

MATERIALS AND METHODS
Experimental design
The present study aimed to characterize the molecular mechanisms of the protein SAMD1 at CpG-containing DNA in vitro and in vivo. All experiments performed in vitro and in vivo were carefully controlled. In vitro experiments performed using recombinant proteins or cell lysates were performed independently in at least two replicates. Experiments performed in cells, including chromatin immunoprecipitation and differentiation experiments, were performed in at least two biological replicates.

SAMD1 constructs
Because of high CG content, the annotation of the SAMD1 gene is conflicting. We used constructs corresponding to human SAMD1 as published previously (39) and annotated in UniProt (Q6SPF0),

National Center for Biotechnology Information (NM_138352) and Ensembl (ENSG00000288488). Note that the standard annotation for human SAMD1 in Ensembl (ENSG00000141858) is likely incorrect. To reduce the CG content in our constructs, the sequence of the open reading frame was synonymously mutated. Point mutations were introduced by PCR or by DNA synthesis.

Cell culture
E14 mouse ES cells (E14TG2a) were cultured in Dulbecco's modified Eagle's medium (DMEM) and GlutaMAX (Gibco, catalog no. 61965-026), 15% fetal calf serum (FCS) (Biochrom, S0115, Lot: 1247B), 1× nonessential amino acids (Gibco, 11140-035), 1× sodium pyruvate (Gibco, 11160-039), 1× penicillin/streptomycin (Gibco, 15140-122), 0.15% β-mercaptoethanol, and LIF (1000 U/ml; Millipore, ESG1107, lot: 3060038) on gelatin-coated plates. Human embryonic kidney–293 (HEK293) cells were culture with DMEM/F-12 (Gibco, 31331-028), 1× penicillin/streptomycin, and 10% FCS.

SAMD1, L3MBTL3, and KDM1A mES KO cells were created by transient transfection using the jetPRIME transfection reagent (Polyplus) with LentiCRISPRv2 (Addgene no. 52961) (40) constructs with following guide RNA sequences: mSAMD1 (#1: AGCGCATCTGC-CGGATGGTG; #2: GAGCATCTCGTACCGCAACG), mL3MBTL3 (#1: AGCAGTTGGGACCATCCATG; #2: GCGAAGATCTAAG-CAGCGGT), and mKDM1A (#1: GGAATAGCCGAGACCCCG-GA; #2: GTTCGATCACGGCCTCACCT). After puromycin selection (3 μg/ml) for 3 days, single-cell clones were obtained and further validated. The KO of SAMD1 was confirmed by Western blot, ChIP, and immunofluorescence. Because of the high CG content of the targeted sequence, a validation of potential indels by sequencing was not possible. The KO of KDM1A was confirmed by Western blot, ChIP, and sequencing. The KO of L3MBTL3 was confirmed by sequencing, ChIP, and immunofluorescence. Detection of the endogenous L3MBTL3 by Western blotting was not successful.

Mouse ES cells were undirected differentiated by removal of LIF as described previously (41). In short, $3 × 10^5$ WT and SAMD1 KO ES cells were plated in the absence of LIF on ungelatinized six-well plates and grown to confluency. The cells were then transferred to ungelatinized bacteria petri dishes and grown for 3 days, during which they were unable to adhere and generated spherical aggregates. These aggregates were then replated on gelatin-coated six-well tissue culture plates where cells adhered and formed differentiating outgrowths. At different time points after replating, the total RNA was isolated to analyze gene expression changes. RT-qPCR primers are presented in table S1.

Antibodies
Polyclonal antibodies for SAMD1 and L3MBTL3 were made using purified glutathione S-transferase (GST)–fusion proteins as antigen. For SAMD1, the antibody is directed against the SAM domain of human SAMD1 (amino acids 452 to 538). For L3MBTL3, the antibodies are directed against the N terminus (amino acids 3 to 233) or the C terminus (amino acids 778 to 883) of mouse L3MBTL3. Antibodies were made with Eurogentec using the 28-day speedy protocol. Obtained antibodies were affinity-purified. The following commercial antibodies were used: H3K4me3 (Diagenode, C15410003), H3K4me2 (Diagenode, C15410035), H3K27ac (Active Motif, 39133), H3K27me3 (Diagenode, C15410195), KDM1A (Abcam, 17721), tubulin (Millipore, MAB3408), Pou5f1/Oct4 (Santa Cruz Biotechnology, SC-5279), FLAG (Sigma-Aldrich, F3165), hemagglutinin (Roche, 11867423001),

H2AUb1 (Cell Signaling Technology, 8240), Lamin B (Santa Cruz Biotechnology, sc-6217), EZH2 (Diagenode, C15410039), L3MBTL2 (Active Motif, 39569), PCGF6 (ProteinTech, 24103-1-AP), RING2 (Abcam, ab101273), RYBP (Sigma-Aldrich, PRS2227), SFMBT1 (Bethyl Laboratories, A303-221A), rabbit immunoglobulin G (IgG) control (Diagenode, C15410206), and H2Av (Drosophila) (Active Motif, 39716). The Sp1 antibody was described previously (42).

Mammalian-two-hybrid
Mammalian-two-hybrid was performed using the Mammalian Two-Hybrid Assay Kit from Stratagene/Agilent [catalog no. 211344 (discontinued)]. SAMD1-SAM and PHC1-SAM was cloned into the pCMV-activation domain (AD) vectors. All other SAM domains were cloned into the pCMV-DNA binding domain (BD) vectors. SAM domains from following human proteins were used: SAMD1 (452-538), SAMD7 (307-410), SAMD11 (533-628), L3MBTL1 (743-840), L3MBTL3 (705-780), L3MBTL4 (537-623), SFMBT1 (751-866), SFMBT2 (804-894), SCML1 (240-329), SCML2 (623-700), SCML4 (338-414), and SCMH1 (579-660). For the experiment, 30,000 HEK293 cells were plated into one 24-well plate. Two technical replicates were performed for each experiment. The next day, the cells were transfected with 200 ng of pFR-Luc, 0.5 ng of SV-40-RLuc, and 50 ng of the pCMV-AD and pCMV-BD constructs using FuGeneHD (Promega). Two days after transfection, cells were washed one time with phosphate-buffered saline (PBS) and lysed for 20 min with 1× passive lysis buffer (reagents from the Dual-Luciferase Reporter Assay System Kit, Promega). Firefly luciferase and renilla luciferase activity was determined by using the Dual-Luciferase Reporter Assay System Kit (Promega). Firefly values were normalized to renilla activity.

ChIP, ChIP-seq, and RNA-seq
ChIP experiments were performed in accordance to the One Day ChIP Kit protocol (Diagenode) using antibodies described above. ChIP-qPCRs with gene-specific primers (table S1) were performed using the ImmoMix PCR reagent (Bioline) in the presence of 0.1× SYBR Green (Molecular Probes). ChIP-qPCR experiments have been repeated at least twice. For ChIP-seq, three individual ChIPs were pooled and purified on QIAquick columns (Qiagen). For ChIP-seq of histone marks, 1 μg of Drosophila S2 chromatin (1:200 relative to mouse ES cell chromatin) was added to each reaction as a spike-in control along with 1 μg of a spike-in antibody directed against the Drosophila-specific H2Av variant (Active Motif). Five nanograms of precipitated DNA was used for indexed sequencing library preparation using the Microplex Library Preparation Kit v2 (Diagenode). Libraries were purified on AMPure magnetic beads (Beckman). For RNA-seq, total RNA was extracted from WT mouse ES cells and three different SAMD1 KO clones by using the RNeasy Mini system (QIAGEN) including an on-column deoxyribonuclease I digestion. RNA integrity was assessed on an Experion StdSens RNA Chip (Bio-Rad). RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina). RNA-seq and ChIP-seq libraries were quantified on a Bioanalyzer (Agilent Technologies). Next-generation sequencing was performed on Illumina HiSeq1500 or NextSeq550.

Bioinformatical analysis
ChIP-seq data were aligned to mouse genome mm9 using Bowtie (43), allowing one mismatch. SAM files were converted to BAM using SAMtools (44). Bigwig files were obtained using deepTools/bamCoverage (45)

and normalized based on reads per kilo base per million mapped reads (RPKM). ChIP-seq for Histone marks were normalized according to the spike-in controls (46). Replicates were merged using SAMtools. Downstream data analysis was performed using Galaxy (47), Cistrome (48), and Bioconductor/R (49). SAMD1-bound peaks were identified by MACS2 (50) using standard settings. Heatmaps and profiles were created using deepTools (45). The P values were calculated by performing at two-tailed Student's t test at RPKM-normalized and log$_2$-transformed read counts at top or bottom 2000 SAMD1-bound CGIs, comparing the data from WT and SAMD1 KO cells. Methylated CGIs were identified using Methylated DNA immunoprecipitation (MeDIP)–seq data (GSM881346) (51). CGI and Promoter definitions were downloaded from the University of California, Santa Cruz (UCSC) Table Browser. Motif enrichment was performed using HOMER (52) by using SAMD1-bound CGIs as input and SAMD1-unbound CGIs as background control. Correlation analysis was performed using the "Multiple wiggle files correlation in given regions" tool within the Galaxy/Cistrome platform (48) using all CGIs as given region. Gene ontology analysis was performed using GREAT (53) using all significant SAMD1 peaks as input. Subsequent network analysis was performed using the EnrichmentMap app in Cytoscape (54). Gene Ontology analysis of SAMD1-bound and up-regulated genes was performed using DAVID (55). The phylogenetic tree of the WH domains was made using interactive Tree Of Life (56). Gene expression of mouse SAMD1 was investigated using BioGPS (57). The analysis of the expression of SAMD1 in cancer and survival analysis was performed using gene expression profiling interactive analysis (GEPIA) (37).

RNA-seq data were aligned to mouse transcriptome GenCode. M23 using RNA Star (2.7.2b) (58). Counts per gene were determined using FeatureCounts (1.6.4). Differentially expressed genes and normalized reads were determined using DeSeq2 (2.11.40.6) (59). GSEA (60) was performed with standard setting. For top SAMD1 target genes, the 100 genes with the highest SAMD1 promoter occupancy, excluding Samd1 itself, were used as gene set. Network analysis was performed using the EnrichmentMap app in Cytoscape (54).

The following publicly available datasets were used: CXXC1 [GSM2454338 and GSM2454339 (61)], MTF2 [GSM2472747 and GSM2472748 (7)], KMT2B [GSM2073033 (18)], KDM2A (GSM1003593) and KDM2B [GSM1003594 (4)], KDM1A [GSM2630507 (23)], H3K27me3 [GSM2472743 and GSM2472744 (7)], H3K4me3 [GSM2472745 and GSM2472746 (7)], and MeDIP-seq [GSM881346 (51)].

Coimmunoprecipitation

All ectopic coimmunoprecipitation (Co-IP) experiments were performed in HEK293 cells. Cells were seeded in 10-cm dishes with 2×10^6 cells per dish. One day later, the expression constructs for 3xHA or N-FLAG–tagged proteins were transfected using FuGENE HD Transfection Reagent (Promega, E2311). Two days after transfection, cell lysis was done using Co-IP lysis buffer [50 mM tris-Cl (pH 7.5), 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 10% Glycerol, 1× protease inhibitor cocktail, and 0.5 mM phenylmethylsulfonyl fluoride]. Cells were shaken for 30 min at 4°C followed by centrifugation for 10 min at 13,000 rpm at 4°C. Protein concentration was determined with the DC Protein Assay (Bio-Rad, 5000116). For each IP, 1 mg of protein was applied, and extract was filled up to a total volume of 500 μl using Co-IP lysis buffer. To remove unspecific binding proteins, a preclearing was performed for 1 hour using mouse IgG–Agarose (Merck, A0919). Beads were equilibrated by

washing two times with 1× tris-buffered saline and one time with Co-IP lysis buffer. To bind FLAG-tagged proteins, precleared extracts were added to 50 μl of ANTI-FLAG M2 Affinity Gel (Merck, A2220) and incubated for approximately 3 hours at 4°C. After incubation, three washing steps with Co-IP lysis buffer were performed. The FLAG beads were boiled 3 min in 30 μl of 2× Laemmli buffer without β-mercaptoethanol. Afterward, 1 μl of β-mercaptoethanol was added and the supernatants were cooked again for 5 min. Detection of proteins in the input, supernatant, and IP fractions was conducted via Western blotting using an 8% gel. Co-IP experiments were repeated at least two times. Endogenous Co-IP was performed in mouse ES cells, comparing WT and SAMD1 KO cells.

Complex purification and mass spectrometry

Flag-HA–tagged human SAMD1 was expressed after retroviral infection of HeLa-S. Nuclear extract was prepared from the established stable cell lines, and the SAMD1 complex was purified using anti-Flag (M2)–conjugated agarose beads (Sigma-Aldrich, A2220) by incubation in Tandem Affinity Purification (TAP) buffer [50 mM tris-HCl (pH 7.9), 100 mM KCl, 5 mM MgCl2, 10% glycerol, 0.1% NP-40, 1 mM dithiothreitol (DTT), and protease inhibitors] for 4 hours and three times washing with TAP buffer. Proteins were eluted with Flag peptides. A second purification was performed using anti-HA–conjugated agarose beads (Santa Cruz Biotechnology, sc-7392), followed by elution with HA peptides. For mass spectrometry, the sample was TCA-precipitated and peptides were identified via liquid chromatography–tandem mass spectrometry at the Taplin Core facility/Harvard Medical School.

Immunofluorescence

WT or SAMD1 KO mouse ES cells were seeded (5×10^5 cells) on coverslips coated with gelatine in six-well plates; after 1 day, cells were washed three times with PBS and incubated for 25 min in 4% paraformaldehyde (in PBS). Cells were washed one time with wash buffer (0.5% Triton in PBS), permeabilized with wash buffer for 25 min, and blocked (wash buffer and 10% FCS) for 1 hour. One-hundred fifty microliters of primary antibody (1:1000 in wash buffer and 10% FCS) was added and incubated for 1 hour. Cells were washed with wash buffer three times for 10 min each, and 150 μl of secondary antibody [1:2000, goat anti-rabbit IgG H&L (Alexa Fluor 488, Thermo Fisher Scientific, A-11008) in wash buffer and 10 % FCS] was added and incubated for 1 hour (dark). Cells were washed with wash buffer three times for 10 min each and subsequently one time with PBS for 10 min. Coverslips were mounted with VECTASHIELD antifade mounting medium with 4′,6-diamidino-2-phenylindole (Vector Laboratories, H1200), transferred to microscope slides, and sealed. Microscopy was performed using a Leica DM5500 microscope, and data were analyzed using ImageJ (Fiji).

Cellular fractionation

Cellular fractionations were performed using "Subcellular Protein Fractionation Kit for Cultured Cells" (Thermo Fisher Scientific, 78840) according to the manufacturer's instructions, followed by Western blotting.

PBM experiments and analysis

Sequences of two distinct hSAMD1-WH regions [amino acids 1 to 110 (WH1) and 28 to 110 (WH2)] were cloned into the pT7CFE1-NHis-GST-CHA plasmid (Thermo Fisher Scientific, 88871). GST-fusion

proteins were expressed using the 1-Step Human Coupled IVT Kit (Thermo Fisher Scientific). Expressed protein concentrations were estimated from anti-GST Western blots. Subsequently, custom-designed "all-10mer" universal oligonucleotide arrays in 8 × 60 K GSE array format (Agilent Technologies, AMADID 030236) were double-stranded, and PBM experiments were performed essentially as described previously (16) with Alexa 488–conjugated anti-GST antibody (Invitrogen, A-11131). Each of the two WH domain constructs (hSAMD1-WH1 and hSAMD1-WH2) was assayed in duplicate at a final concentration of 600 nM in PBS-based binding and wash buffers on fresh slides. Scans were acquired using a GenePix 4400A (Molecular Devices) microarray scanner. Microarray data quantification, normalization, and motif derivation were performed essentially as described previously using the Universal PBM Analysis Suite and the Seed-and-Wobble motif-derivation algorithm (16).

Protein expression and purification

Open reading frame of human SAMD1-WH is chemically synthesized with codon optimized for efficient bacterial expression. SAMD1-WH–containing fragments (residues 27 to 105, 16 to 110, and its mutants) and SAMD1-SAM–containing fragments (residues 459 to 523, 459 to 526 and 459 to 530) were inserted into a hexahistidine-SUMO–tagged pRSFDuet-1 vector. The target proteins were expressed in *Escherichia coli* strain BL21(DE3) cells, which were shaken at 37°C until the OD_{600} (optical density at 600 nm) reached around 1.0, and then cooled at 20°C for around 1 hour before 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) were added to induce expression overnight. Cells were collected by centrifugation at 5000g for 10 min. Cell pellets were resuspended with the initial buffer containing 20 mM tris at pH 7.0, 500 mM NaCl, and 20 mM imidazole and then sonicated for around 5 min. The supernatant was collected by centrifugation of the cell lysate at 25,000g for 1 hour. Histidine-SUMO–tagged target protein was isolated through a nickel-charged HiTrap Chelating FF column (GE Healthcare). The histidine-SUMO tag was cleaved by incubating with a histidine-tagged ubiquitin-like-specific protease 1 (ULP1) protease and then dialyzed with the initial buffer at 4°C overnight.

For SAMD1-WH, the dialyzed solution was reloaded onto a nickel-charged chelating column to remove both the histidine-tagged SUMO and ULP1 protease. The flow-through was diluted twofold with 20 mM tris at pH 7.0 and 2 mM DTT to yield a solution at half the initial salt concentration (250 mM NaCl), which was then loaded directly onto a heparin column (GE Healthcare) to remove bound DNA. Target protein was separated by increasing the salt concentration of the low-salt buffer (20 mM tris at pH 7.0, 250 mM NaCl, and 2 mM DTT) from 250 mM to 1 M NaCl through a linear gradient. The target protein was further purified by a HiLoad 200 16/600 gel filtration column (GE Healthcare) equilibrated with the low-salt buffer through which the resulting product was pooled. Purified proteins were concentrated to around 20 mg/ml and stored in a −80°C freezer.

For SAMD1-SAM, after elution of the histidine-SUMO–tagged protein, SAMD1-SAM was incubated with histidine-tagged ULP1 protease and dialyzed with the low-salt buffer containing 20 mM tris at pH 8.0, 100 mM NaCl, and 2 mM DTT at 4°C overnight. The dialyzed sample was loaded onto a HiTrap Q FF column (GE Healthcare) and then eluted by increasing the salt concentration from 100 mM NaCl to 1 M NaCl to remove histidine-SUMO tag.

The eluted target protein was then purified through a HiLoad 200 16/600 gel filtration column before loading onto a Mono Q 5/50 column (GE Healthcare) for further purification. After these steps, the target protein was concentrated to around 16 mg/ml and stored in a freezer.

The selenomethionine-labeled SAMD1-WH and SAMD1-SAM were expressed in the methionine auxotrophic B834 (DE3) strain. Cells (1 liter) grown in LB media at the OD_{600} of 1.2 were harvested by centrifugation at 4000 rpm for 10 min. The cells were washed twice with M9 media and then were used to inoculate 2 liter of methionine-depleted medium supplemented with L-selenomethionine (50 mg/liter) and nutrient mix (SelenoMet, Molecular Dimensions). After shaking for an additional 30 min at 37°C, the cells were induced with 0.2 mM IPTG and shaken overnight at 20°C. Selenomethionine-labeled proteins were purified similarly as those of the native proteins, except that 2 mM β-mercaptoethanol was added in the buffer in the initial stage of protein purification. All mutations of SAMD1-WH and SAMD1-SAM were generated by PCR-based method.

Crystallization and structure resolution

Crystallization was carried out using the hanging-drop, vapor-diffusion method through mixing equal volume of protein and well solution. The complex of SAMD1-WH (27 to 105), and DNA was prepared by mixing the target protein with a 13-bp CpG-containing dsDNA (5′-ACCTGCGCACCAT-3′ as the sequence of one strand) at the molar ratio of 2:1.1.

The crystals of both native and selenomethionine-labeled SAMD1-WH/DNA complex were grown in the solution containing 0.2 M calcium acetate hydrate, 20% (w/v) polyethylene glycol 3350 at 4°C. Crystals were flash-frozen in the cryoprotectant composed of crystallization buffer containing 12% 2,3-butanediol.

The crystals of SAMD1-SAM (459 to 523) were grown in the solution containing 2.1 M ammonium sulfate and 0.2 M magnesium chloride hexahydrate at 20°C. Crystals of SAMD1-SAM (459 to 530) were grown in the solution of 0.1 M Hepes at pH 7.0, 23% (w/v) polyethylene glycol 3350 at 20°C. Crystallization buffer with addition of 10% 2,3-butanediol was used as the cryoprotectant for both crystals. The crystals of SAMD1-SAM (459 to 526) were grown in the solution containing 0.1 M bis-tris at pH 7.5 and 2.1 M ammonium sulfate at 20°C. Crystallization buffer containing 20% glycerol was used as the cryoprotectant.

All the datasets were collected at the Shanghai Synchrotron Radiation Facility beamlines in China at the temperature of −196°C. Datasets for selenomethionine-labeled SAMD1-WH/DNA complex crystals were collected at the beamline BL19U1 at the wavelength of 0.97855 Å. Datasets for SAMD1-SAM (459 to 523) crystals were collected at the beamline BL17U1 at the wavelength of 0.97922 Å. Datasets for SAMD1-SAM (459 to 526) crystals were collected at the beamline BL19U1 at the wavelength of 0.97891 Å. Datasets for selenomethionine-labeled SAMD1-SAM (459 to 530) crystals were collected at the beamline BL19U1 at the wavelength of 0.97917. The datasets were processed using the program HKL2000 (62). Structures of SAMD1-WH/DNA complex and SAMD1-SAM (459 to 530) were solved by PHENIX (63) using the SAD method with the anomalous signals from selenomethionine-labeled crystals. The initial partial model was manually built in Coot (64) and further refined by PHENIX. High-resolution structures of SAMD1-SAM (459 to 523) and SAMD1-SAM (459 to 526) were solved by molecular replacement method using the model of SAMD1-SAM (459 to

530). There is one SAMD1-WH/DNA complex molecule in one crystallographic asymmetric unit. In the final model, 98.55 and 1.45% residues are refined in the favored and allowed regions in the Ramachandran plot, respectively. There is one SAMD1-SAM pentamer in one asymmetric unit of the SAMD1-SAM (459 to 523) crystals. In the final model, 99.38 and 0.62% residues are refined in the favored and allowed regions in the Ramachandran plot, respectively. There are two SAMD1-SAM decamers in one asymmetric unit of the SAMD1-SAM (459 to 530) and SAMD1-SAM (459 to 526) crystals. In both crystals, each decamer is formed by two head-to-head stacked SAMD1-SAM pentamers. In the final model of the SAMD1-SAM (459 to 526) structure, 98.28 and 1.72% residues are refined in the favored and allowed regions in the Ramachandran plot, respectively. X-ray statistics are listed in Table 1.

ITC measurement
Calorimetric experiments were carried out at 20°C with a MicroCal iTC200 instrument. Purified WT or mutant proteins and dsDNA molecules were dialyzed overnight at 4°C in the titration buffer containing 20 mM tris at pH 7.5, 150 mM NaCl, and 2 mM β-mercaptoethanol. Titration was performed by injecting DNA molecules into the protein samples. Calorimetric titration data were fitted with the Origin software under the algorithm of one binding site model. All ITC measurements have been repeated at least twice. ITC binding parameters are listed in Table 2.

Electrophoretic mobility shift assay
dsDNA (50 pmol) was mixed with increasing amounts of recombinant SAMD1-WH proteins in the reaction buffer containing 20 mM tris at pH 7.0, 200 mM NaCl, and 2 mM DTT and incubated at 4°C for 10 min. The mixture was then loaded onto a 1.2% agarose gel in the tris-acetate-EDTA buffer for electrophoresis and detected by ethidium bromide staining. SAMD1-WH (16 to 110), and its mutants were used for the analysis. All EMSA experiments were repeated at least three times. DNA names and sequences are listed in table S2.

Statistical analysis
Statistical analyses were performed as indicated in the figure legends or Materials and Methods.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/20/eabf2229/DC1

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES
1. A. M. Deaton, A. Bird, CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
2. K. S. Voo, D. L. Carlone, B. M. Jacobsen, A. Flodin, D. G. Skalnik, Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* **20**, 2108–2121 (2000).
3. J. P. Thomson, P. J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A. R. Kerr, A. Deaton, R. Andrews, K. D. James, D. J. Turner, R. Illingworth, A. Bird, CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082–1086 (2010).
4. A. M. Farcas, N. P. Blackledge, I. Sudbery, H. K. Long, J. F. McGouran, N. R. Rose, S. Lee, D. Sims, A. Cerase, T. W. Sheahan, H. Koseki, N. Brockdorff, C. P. Ponting, B. M. Kessler, R. J. Klose, KDM2B links the Polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *eLife* **1**, e00205 (2012).
5. H. Li, R. Liefke, J. Jiang, J. V. Kurland, W. Tian, P. Deng, W. Zhang, Q. He, D. J. Patel, M. L. Bulyk, Y. Shi, Z. Wang, Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* **549**, 287–291 (2017).
6. P. Voigt, W. W. Tee, D. Reinberg, A double take on bivalent promoters. *Genes Dev.* **27**, 1318–1338 (2013).
7. M. Perino, G. van Mierlo, I. D. Karemaker, S. van Genesen, M. Vermeulen, H. Marks, S. J. van Heeringen, G. J. C. Veenstra, MTF2 recruits Polycomb repressive complex 2 by helical-shape-selective DNA binding. *Nat. Genet.* **50**, 1002–1010 (2018).
8. T. Viturawong, F. Meissner, F. Butter, M. Mann, A DNA-centric protein interaction map of ultraconserved elements reveals contribution of transcription factor binding hubs to conservation. *Cell Rep.* **5**, 531–545 (2013).
9. J. Xiong, Z. Zhang, J. Chen, H. Huang, Y. Xu, X. Ding, Y. Zheng, R. Nishinakamura, G. L. Xu, H. Wang, S. Chen, S. Gao, B. Zhu, Cooperative action between SALL4A and TET proteins in stepwise oxidation of 5-methylcytosine. *Mol. Cell* **64**, 913–925 (2016).
10. T. Bartke, M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann, T. Kouzarides, Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470–484 (2010).
11. J. Zhang, R. Bonasio, F. Strino, Y. Kluger, J. K. Holloway, A. J. Modzelewski, P. E. Cohen, D. Reinberg, SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes Dev.* **27**, 749–766 (2013).
12. A. Malovannaya, Y. Li, Y. Bulynko, S. Y. Jung, Y. Wang, R. B. Lanz, B. W. O'Malley, J. Qin, Streamlined analysis schema for high-throughput identification of endogenous protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2431–2436 (2010).
13. E. Engelen, J. H. Brandsma, M. J. Moen, L. Signorile, D. H. Dekkers, J. Demmers, C. E. Kockx, Z. Ozgür, W. F. van IJcken, D. L. van den Berg, R. A. Poot, Proteins that bind regulatory regions identified by histone modification chromatin immunoprecipitations and mass spectrometry. *Nat. Commun.* **6**, 7155 (2015).
14. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).
15. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
16. M. F. Berger, A. A. Philippakis, A. M. Qureshi, F. S. He, P. W. Estep III, M. L. Bulyk, Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–1435 (2006).
17. G. M. Harami, M. Gyimesi, M. Kovacs, From keys to bulldozers: Expanding roles for winged helix domains in nucleic-acid-binding proteins. *Trends Biochem. Sci.* **38**, 364–371 (2013).
18. D. Hu, X. Gao, K. Cao, M. A. Morgan, G. Mas, E. R. Smith, A. G. Volk, E. T. Bartom, J. D. Crispino, L. Di Croce, A. Shilatifard, Not All H3K4 methylations are created equal: Mll2/COMPASS dependency in primordial germ cell specification. *Mol. Cell* **65**, 460–475.e6 (2017).
19. T. Xu, S. S. Park, B. D. Giaimo, D. Hall, F. Ferrante, D. M. Ho, K. Hori, L. Anhezini, I. Ertl, M. Bartkuhn, H. Zhang, E. Milon, K. Ha, K. P. Conlon, R. Kuick, B. Govindarajoo, Y. Zhang, Y. Sun, Y. Dou, V. Basrur, K. S. Elenitoba-Johnson, A. I. Nesvizhskii, J. Ceron, C. Y. Lee, T. Borggrefe, R. A. Kovall, J. F. Rual, RBPJ/CBF1 interacts with L3MBTL3/MBT1 to promote repression of Notch signaling via histone demethylase KDM1A/LSD1. *EMBO J.* **36**, 3232–3249 (2017).
20. Y. Shi, F. Lan, C. Matson, P. Mulligan, J. R. Whetstine, P. A. Cole, R. A. Casero, Y. Shi, Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–953 (2004).
21. K. Isono, T. A. Endo, M. Ku, D. Yamada, R. Suzuki, J. Sharif, T. Ishikura, T. Toyoda, B. E. Bernstein, H. Koseki, SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* **26**, 565–577 (2013).
22. A. H. Wani, A. N. Boettiger, P. Schorderet, A. Ergun, C. Munger, R. I. Sadreyev, X. Zhuang, R. E. Kingston, N. J. Francis, Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat. Commun.* **7**, 10291 (2016).
23. K. Cao, C. K. Collings, M. A. Morgan, S. A. Marshall, E. J. Rendleman, P. A. Ozark, E. R. Smith, A. Shilatifard, An Mll4/COMPASS-Lsd1 epigenetic axis governs enhancer function and pluripotency transition in embryonic stem cells. *Sci. Adv.* **4**, eaap8747 (2018).
24. R. Bonasio, E. Lecona, D. Reinberg, MBT domain proteins in development and disease. *Semin. Cell Dev. Biol.* **21**, 221–230 (2010).
25. S. Egolf, Y. Aubert, M. Doepner, A. Anderson, A. Maldonado-Lopez, G. Pacella, J. Lee, E. K. Ko, J. Zou, Y. Lan, C. L. Simpson, T. Ridky, B. C. Capell, LSD1 inhibition promotes epithelial differentiation through derepression of fate-determining transcription factors. *Cell Rep.* **28**, 1981–1992.e7 (2019).
26. W. J. Harris, X. Huang, J. T. Lynch, G. J. Spencer, J. R. Hitchin, Y. Li, F. Ciceri, J. G. Blaser, B. F. Greystoke, A. M. Jordan, C. J. Miller, D. J. Ogilvie, T. C. Somervaille, The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells. *Cancer Cell* **21**, 473–487 (2012).
27. N. Nady, L. Krichevsky, N. Zhong, S. Duan, W. Tempel, M. F. Amaya, M. Ravichandran, C. H. Arrowsmith, Histone recognition by human malignant brain tumor domains. *J. Mol. Biol.* **423**, 702–718 (2012).

28. F. Frey, T. Sheahan, K. Finkl, G. Stoehr, M. Mann, C. Benda, J. Muller, Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. *Genes Dev.* **30**, 1116–1127 (2016).

29. C. A. Kim, M. Gingery, R. M. Pilpa, J. U. Bowie, The SAM domain of polyhomeotic forms a helical polymer. *Nat. Struct. Biol.* **9**, 453–457 (2002).

30. C. A. Kim, M. L. Phillips, W. Kim, M. Gingery, H. H. Tran, M. A. Robinson, S. Faham, J. U. Bowie, Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. *EMBO J.* **20**, 4173–4182 (2001).

31. F. Qiao, J. U. Bowie, The many faces of SAM. *Sci. STKE* **2005**, re7 (2005).

32. J. Wang, K. Scully, X. Zhu, L. Cai, J. Zhang, G. G. Prefontaine, A. Krones, K. A. Ohgi, P. Zhu, I. Garcia-Bassets, F. Liu, H. Taylor, J. Lozach, F. L. Jayes, K. S. Korach, C. K. Glass, X. D. Fu, M. G. Rosenfeld, Opposing LSD1 complexes function in developmental gene activation and repression programmes. *Nature* **446**, 882–887 (2007).

33. H. K. Long, N. P. Blackledge, R. J. Klose, ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Trans.* **41**, 727–740 (2013).

34. S. Horsefield, H. Burdett, X. Zhang, M. K. Manik, Y. Shi, J. Chen, T. Qi, J. Gilley, J. S. Lai, M. X. Rank, L. W. Casey, W. Gu, D. J. Ericsson, G. Foley, R. O. Hughes, T. Bosanac, M. von Itzstein, J. P. Rathjen, J. D. Nanson, M. Boden, I. B. Dry, S. J. Williams, B. J. Staskawicz, M. P. Coleman, T. Ve, P. N. Dodds, B. Kobe, NAD$^+$ cleavage activity by animal and plant TIR domains in cell death pathways. *Science* **365**, 793–799 (2019).

35. K. Siklenka, S. Erkek, M. Godmann, R. Lambrot, S. McGraw, T. Cohen, J. Xia, M. Suderman, J. Hallett, J. Trasler, A. H. Peters, S. Kimmins, Disruption of histone methylation in developing sperm impairs offspring health transgenerationally. *Science* **350**, aab2006 (2015).

36. A. Nishiyama, A. A. Sharov, Y. Piao, M. Amano, T. Amano, H. G. Hoang, B. Y. Binder, R. Tapnio, U. Bassey, J. N. Malinou, L. S. Correa-Cerro, H. Yu, L. Xin, E. Meyers, M. Zalzman, Y. Nakatake, C. Stagg, L. Sharova, Y. Qian, D. Dudekula, S. Sheer, J. S. Cadet, T. Hirata, H. T. Yang, I. Goldberg, M. K. Evans, D. L. Longo, D. Schlessinger, M. S. Ko, Systematic repression of transcription factors reveals limited patterns of gene expression changes in ES cells. *Sci. Rep.* **3**, 1390 (2013).

37. Z. Tang, C. Li, B. Kang, G. Gao, Z. Zhang, GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102 (2017).

38. T. M. Norman, M. A. Horlbeck, J. M. Replogle, A. Y. Ge, A. Xu, M. Jost, L. A. Gilbert, J. S. Weissman, Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science* **365**, 786–793 (2019).

39. A. M. Lees, A. E. Deconinck, B. D. Campbell, R. S. Lees, Atherin: A newly identified, lesion-specific, LDL-binding protein in human atherosclerosis. *Atherosclerosis* **182**, 219–230 (2005).

40. N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

41. J. E. Mermoud, C. Costanzi, J. R. Pehrson, N. Brockdorff, Histone macroH2A1.2 relocates to the inactive X chromosome after initiation and propagation of X-inactivation. *J. Cell. Biol.* **147**, 1399–1408 (1999).

42. S. Volkel, B. Stielow, F. Finkernagel, T. Stiewe, A. Nist, G. Suske, Zinc finger independent genome-wide binding of Sp2 potentiates recruitment of histone-fold protein Nf-y distinguishing it from Sp1 and Sp3. *PLOS Genet.* **11**, e1005102 (2015).

43. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

44. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

45. F. Ramirez, F. Dundar, S. Diehl, B. A. Gruning, T. Manke, deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

46. D. A. Orlando, M. W. Chen, V. E. Brown, S. Solanki, Y. J. Choi, E. R. Olson, C. C. Fritz, J. E. Bradner, M. G. Guenther, Quantitative ChIP-seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170 (2014).

47. E. Afgan, D. Baker, B. Batut, M. van den Beek, D. Bouvier, M. Cech, J. Chilton, D. Clements, N. Coraor, B. A. Gruning, A. Guerler, J. Hillman-Jackson, S. Hiltemann, V. Jalili, H. Rasche, N. Soranzo, J. Goecks, J. Taylor, A. Nekrutenko, D. Blankenberg, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46**, W537–W544 (2018).

48. T. Liu, J. A. Ortiz, L. Taing, C. A. Meyer, B. Lee, Y. Zhang, H. Shin, S. S. Wong, J. Ma, Y. Lei, U. J. Pape, M. Poidinger, Y. Chen, K. Yeung, M. Brown, Y. Turpaz, X. S. Liu, Cistrome: An integrative platform for transcriptional regulation studies. *Genome Biol.* **12**, R83 (2011).

49. W. Huber, V. J. Carey, R. Gentleman, S. Anders, M. Carlson, B. S. Carvalho, H. C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K. D. Hansen, R. A. Irizarry, M. Lawrence, M. I. Love, J. MacDonald, V. Obenchain, A. K. Oles, H. Pages, A. Reyes, P. Shannon, G. K. Smyth, L. Tenenbaum, L. Waldron, M. Morgan, Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).

50. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).

51. P. Yu, S. Xiao, X. Xin, C. X. Song, W. Huang, D. McDee, T. Tanaka, T. Wang, C. He, S. Zhong, Spatiotemporal clustering of the epigenome reveals rules of dynamic gene regulation. *Genome Res.* **23**, 352–364 (2013).

52. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

53. C. Y. McLean, D. Bristor, M. Hiller, S. L. Clarke, B. T. Schaar, C. B. Lowe, A. M. Wenger, G. Bejerano, GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).

54. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

55. G. Dennis Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, R. A. Lempicki, DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol.* **4**, R60 (2003).

56. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).

57. C. Wu, C. Orozco, J. Boyer, M. Leglise, J. Goodale, S. Batalov, C. L. Hodge, J. Haase, J. Janes, J. W. Huss III, A. I. Su, BioGPS: An extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* **10**, R130 (2009).

58. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

59. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

60. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).

61. D. A. Brown, V. Di Cerbo, A. Feldmann, J. Ahn, S. Ito, N. P. Blackledge, M. McClellan, E. Dimitrova, A. H. Turberfield, H. K. Long, H. W. King, S. Kriaucionis, L. Schermelleh, T. G. Kutateladze, H. Koseki, R. J. Klose, The SET1 complex selects actively transcribed target genes via multivalent interaction with CpG island chromatin. *Cell Rep.* **20**, 2313–2327 (2017).

62. Z. Otwinowski, W. Minor, [20] Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

63. P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).

64. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).

Citation: B. Stielow, Y. Zhou, Y. Cao, C. Simon, H.-M. Pogoda, J. Jiang, Y. Ren, S. K. Phanor, I. Rohner, A. Nist, T. Stiewe, M. Hammerschmidt, Y. Shi, M. L. Bulyk, Z. Wang, R. Liefke, The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Sci. Adv.* **7**, eabf2229 (2021).

ScienceAdvances

The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands

Bastian Stielow, Yuqiao Zhou, Yinghua Cao, Clara Simon, Hans-Martin Pogoda, Junyi Jiang, Yanpeng Ren, Sabrina Keita Phanor, Iris Rohner, Andrea Nist, Thorsten Stiewe, Matthias Hammerschmidt, Yang Shi, Martha L. Bulyk, Zhanxin Wang, and Robert Liefke

View the article online
https://www.science.org/doi/10.1126/sciadv.abf2229
Permissions
https://www.science.org/help/reprints-and-permissions

Use of this article is subject to the Terms of service

Science Advances
AAAS

Supplementary Materials for

The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands

Bastian Stielo, Yuqiao Zhou, Yinghua Cao, Clara Simon, Hans-Martin Pogoda, Junyi Jiang, Yanpeng Ren, Sabrina Keita Phanor, Iris Rohner, Andrea Nist, Thorsten Stiewe, Matthias Hammerschmidt, Yang Shi, Martha L. Bulyk, Zhanxin Wang*, Robert Liefke*

*Corresponding author. Email: robert.liefke@imt.uni-marburg.de (R.L.); wangz@bnu.edu.cn (Z.W.)

This PDF file includes:

Figs. S1 to S8
Tables S1 and S2

Supplementary Materials
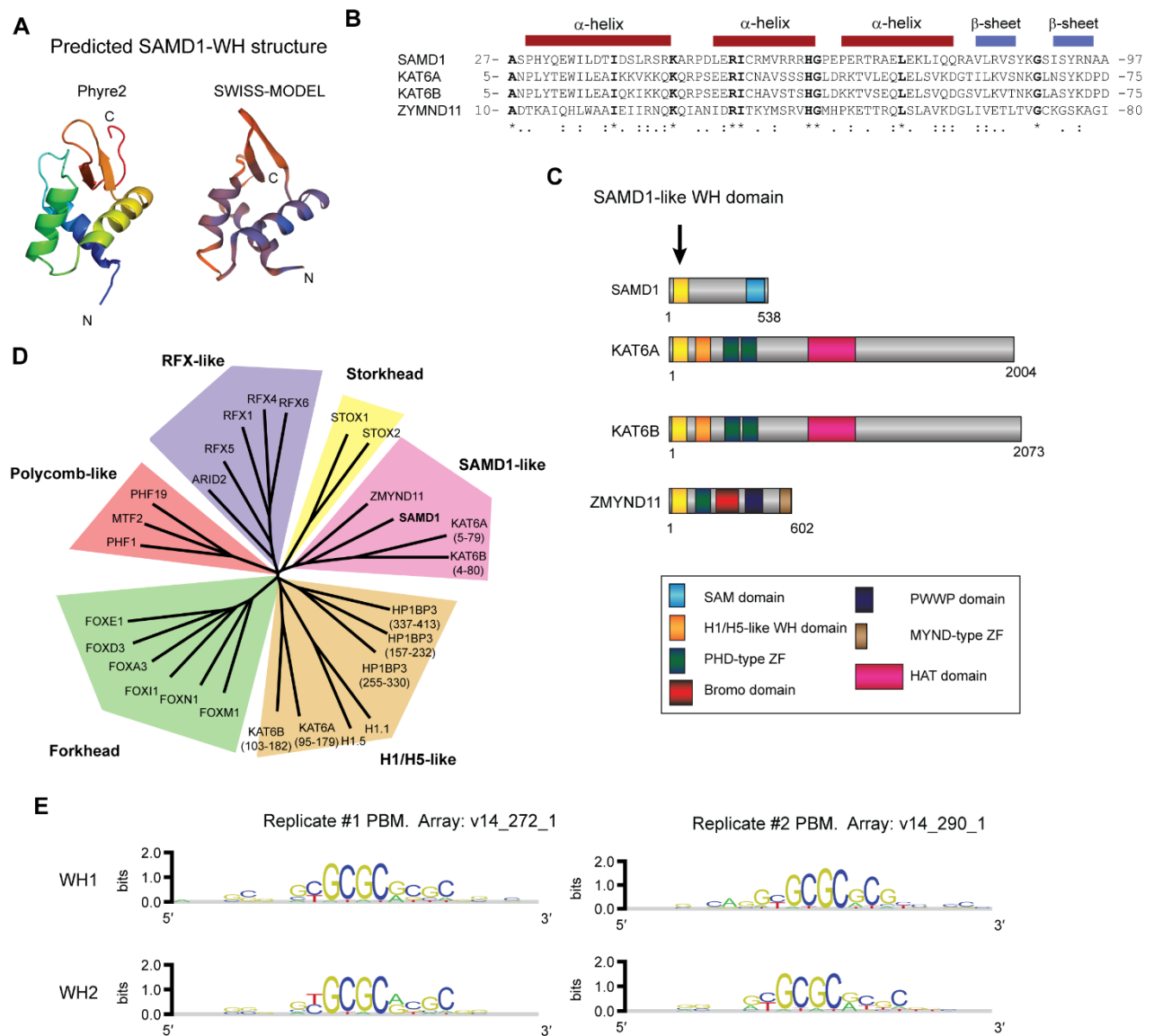
Fig. S1. SAMD1-WH domain is a founding member of a novel class of WH domains.
(**A**) Predicted structures of the SAMD1-WH domain, obtained by Phyre2 (*14*) or SWISS-Model (*15*).
(**B**) Alignment of SAMD1-WH with related sequences in KAT6A, KAT6B and ZMYND11.
(**C**) Domain structure of the proteins containing the SAMD1-like WH domains.
(**D**) Phylogenetic tree of DNA binding WH domains. For the Forkhead proteins selected members are shown. The SAMD1-like WH domains are a distinct class, compared to all other WH domains.
(**E**) Motifs identified in protein binding microarrays (PBMs) (*16*) using the SAMD1-WH domain. Results of two distinct SAMD1 WH domain constructs (WH1 = aa 1-110, WH2 = aa 28-110) and two replicates are shown.
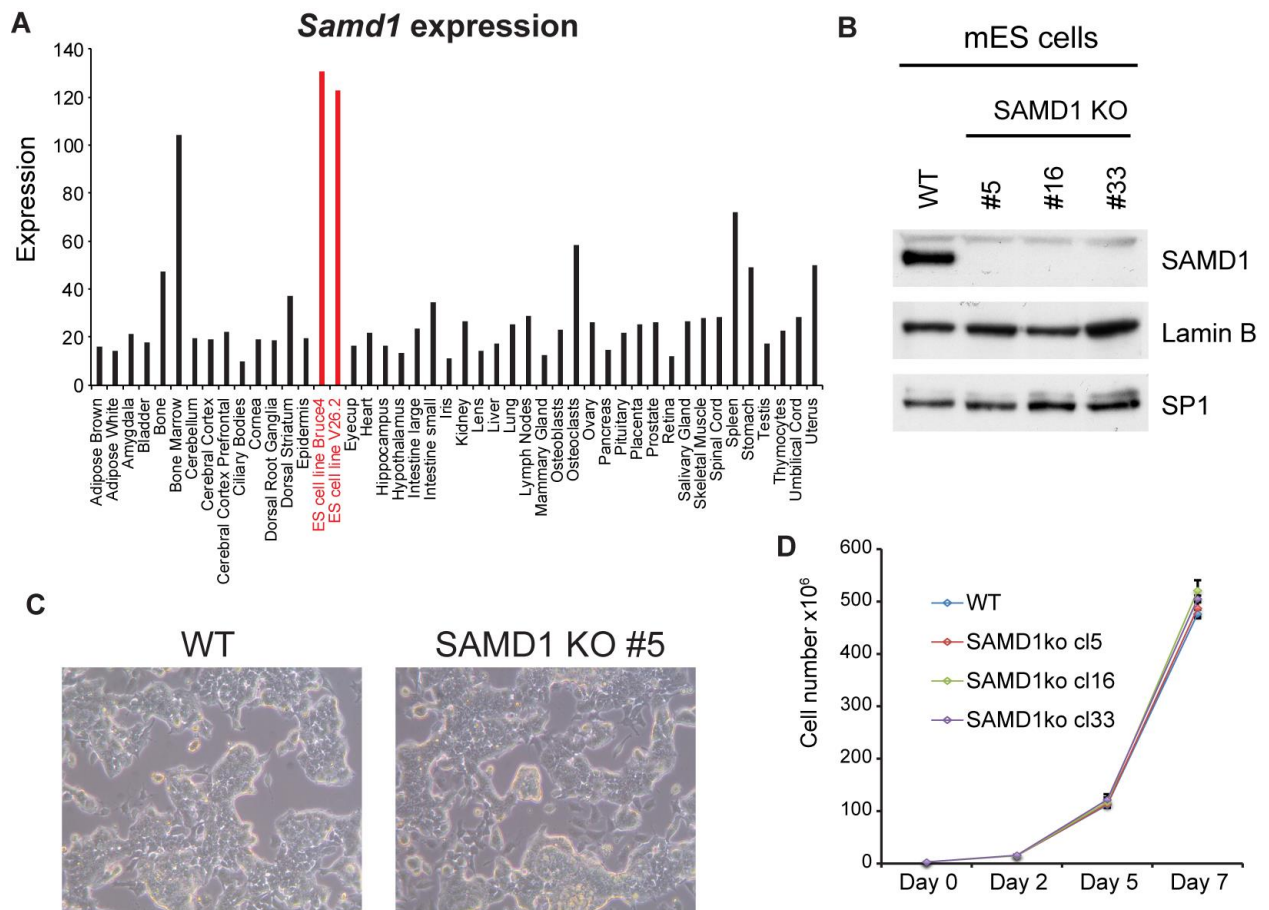
Fig. S2. Creation and characterisation of SAMD1 KO ES cells.

(**A**) Expression of SAMD1 in mouse tissues based on BioGPS microarray data (*57*).

(**B**) Western blot showing three independent SAMD1 KO clones, created via CRISPR/Cas9.

(**C**) Bright field microscopy picture of WT and SAMD1 KO cells.

(**D**) Proliferation of wildtype and SAMD1 KO ES cells. (three biological replicates, Error bars indicate mean ± SD)
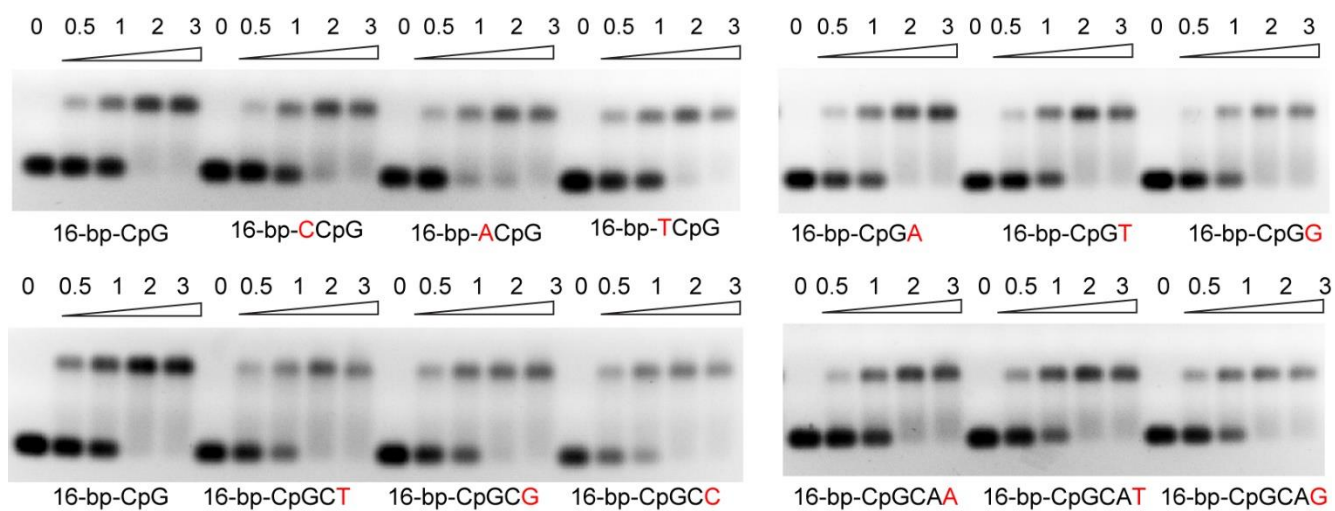
Fig. S3. Extended EMSA analysis of SAMD1-WH
EMSA of the DNA-binding affinities of SAMD1-WH when bases flanking the CpG motif is replaced by other bases.
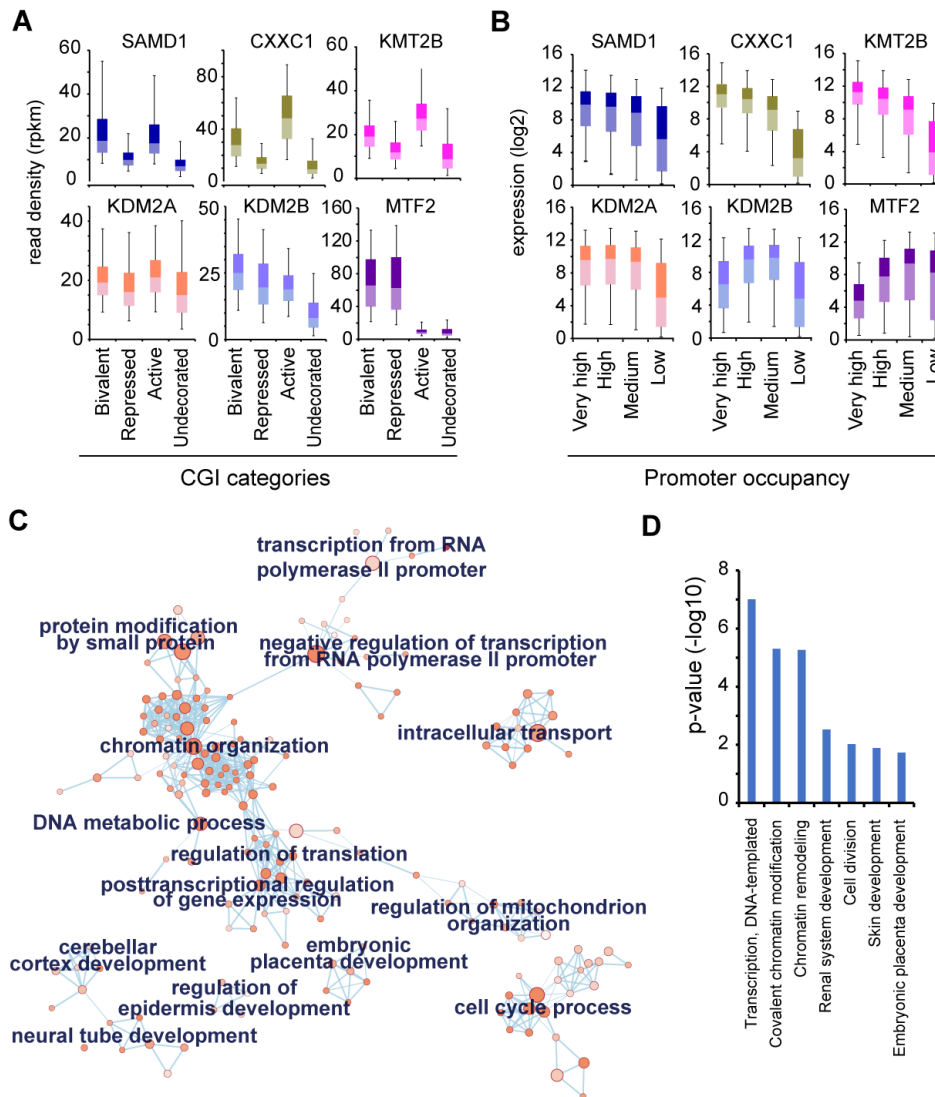
Fig. S4. Comparison of SAMD1 to other CpG binding proteins and gene ontology of its target genes.

(**A**) Read density of SAMD1, CXXC1, KMT2B, KDM2B, KDM2A and MTF2 at the four CGI groups described in **Fig. 3B**.

(**B**) Relationship of promoter occupancy of the six investigated CGI-binding proteins and gene expression. Very high = genes with top 5% highest promoter occupancy, high = 75 - 95%, medium = 50 - 75%, low = 0 - 50%. The whisker-box plots in (A) and (B) represent the lower quartile, median and upper quartile of the data with 5 % and 95 % whiskers.

(**C**) Gene ontologies enriched at all SAMD1 targets. Analysis was performed by GREAT and Cytoscape.

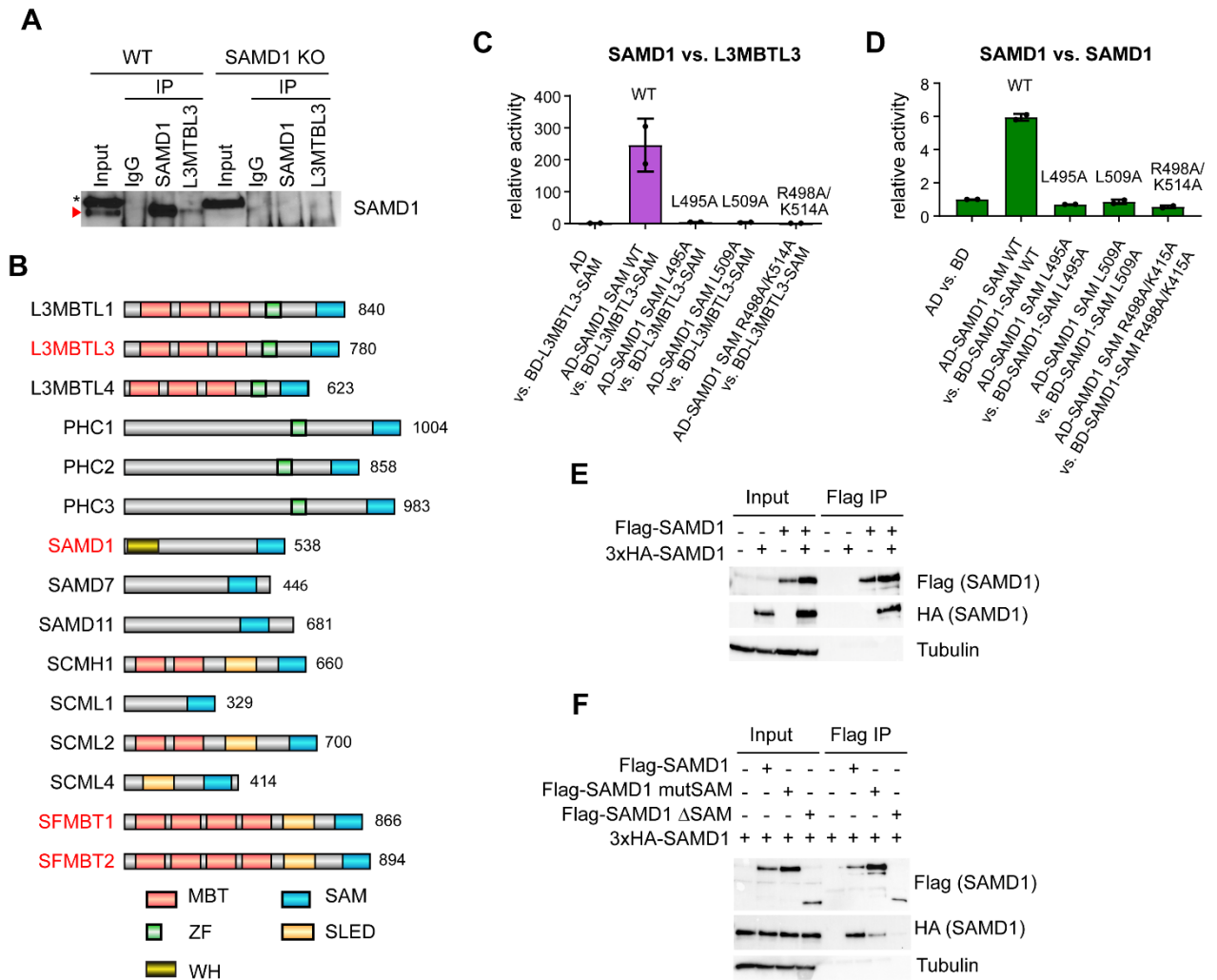(**D**) Gene Ontology analysis of the upregulated and SAMD1 bound genes, using DAVID.

Fig. S5. Disruption of SAM-SAM interactions by point mutations

(**A**) Endogenous co-immunoprecipitation of SAMD1 using an L3MBTL3 antibody, performed in mouse ES cells. Red arrow indicates SAMD1 band. Asterisk indicates an unspecific band.

(**B**) Overview of Polycomb-related SAM domain-containing nuclear proteins investigated in the mammalian-two-hybrid (**Fig. 4C**). Proteins in red were identified to be associated with SAMD1 (**Fig. 4A**). Numbers indicate the length of the human proteins in amino acids.

(**C**, **D**) Mammalian-two-hybrid to investigate the consequence of mutations for the SAMD1-SAM interactions with L3MBTL3 (C) or with itself (D). (two technical replicates, Error bars indicate mean ± SD.)

(**E**) Coimmunoprecipitation of HA-SAMD1 with Flag-SAMD1.

(**F**) Coimmunoprecipitation of wildtype HA-SAMD1 with distinct Flag-SAMD1 constructs (see also **Fig. 4D**).
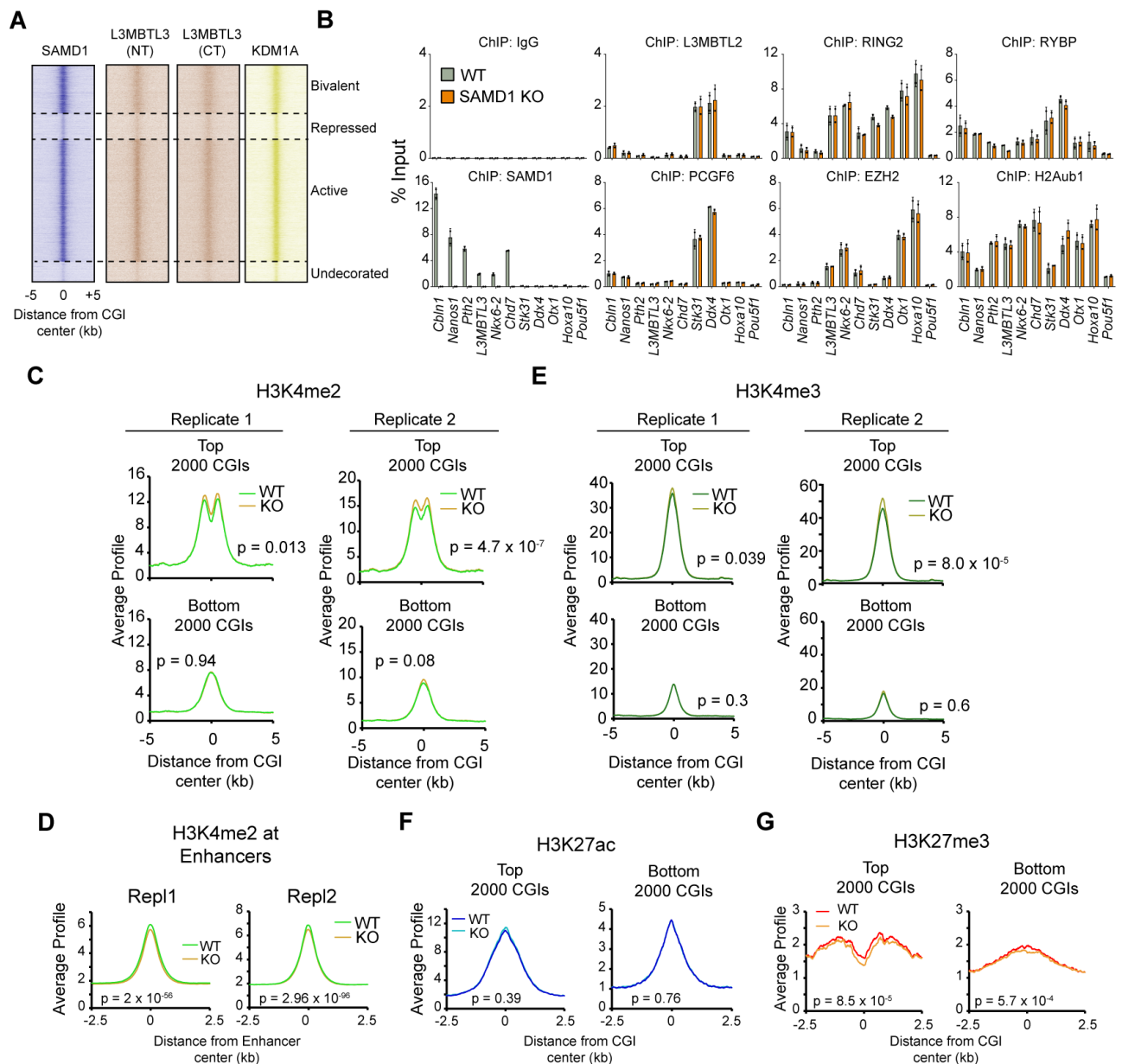
Fig. S6. Detailed investigation of chromatin-related role of SAMD1

(**A**) Heatmaps of binding pattern of L3MBTL3 and KDM1A at the four CGI categories described in **Fig. 3B**.

(**B**) ChIP-qPCR experiments for various canonical and non-canonical Polycomb proteins at SAMD1 and Polycomb target genes. (two biological replicates, Error bars indicate mean ± SD.)

(**C**) Profiles of the two replicates demonstrating the changes of H3K4me2 (C) at CGIs upon SAMD1 knockout, as described in **Fig. 5D**.

(**D**) Profiles of H3K4me2 at enhancer sites in wildtype and SAMD1 KO cells. The two biological replicates are shown.

(**E**) Profiles as in (C) but for H3K4me3.

(**F, G**) Profiles for H3K27ac (F) and H3K27me3 (G) at the 2000 top and bottom SAMD1-bound CGIs in wildtype and SAMD1 KO cells.

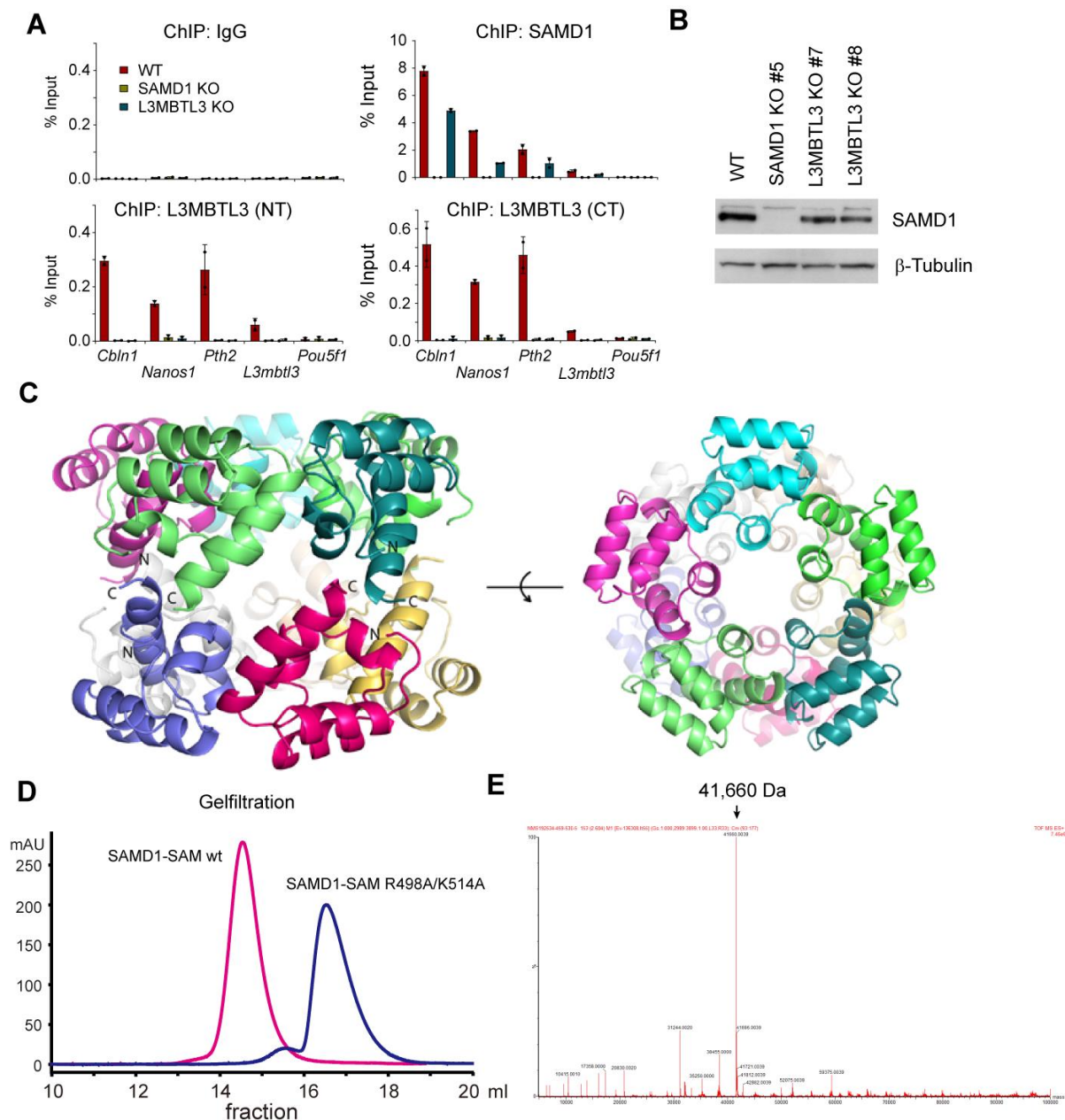P-values in (C)-(G) were calculated by two-sided Student's t-tests.

Fig. S7. Role of L3MBTL3 for SAMD1 chromatin binding and extended investigation of SAMD1 self-association

(**A**) ChIP-qPCR of SAMD1 and L3MBTL3 in SAMD1 and L3MBTL3 KO cells. SAMD1 KO abolishes L3MBTL3 chromatin binding but not vice versa. (two biological replicates, Error bars indicate ± standard deviation).

(**B**) Western of SAMD1 in L3MBTL3 KO cells.

(**C**) Side (left) and top (right) views of the decameric organization of the SAMD1-SAM domain in the second crystal form (PDB: 6LUK). Two pentamers stack head-to-head to form a decamer.

(**D**) Gelfiltration of SAMD1-SAM wt and R498A/K514A mutant, demonstrating that the mutation disrupts a higher order structure of the SAMD1-SAM domain.

(**E**) MS-spectrum of the SAMD1-SAM (459-530) in solution. The highest peak indicates the molecular weight of 41,660 Da, which correlates to the molecular weight of a SAMD1-SAM pentamer.
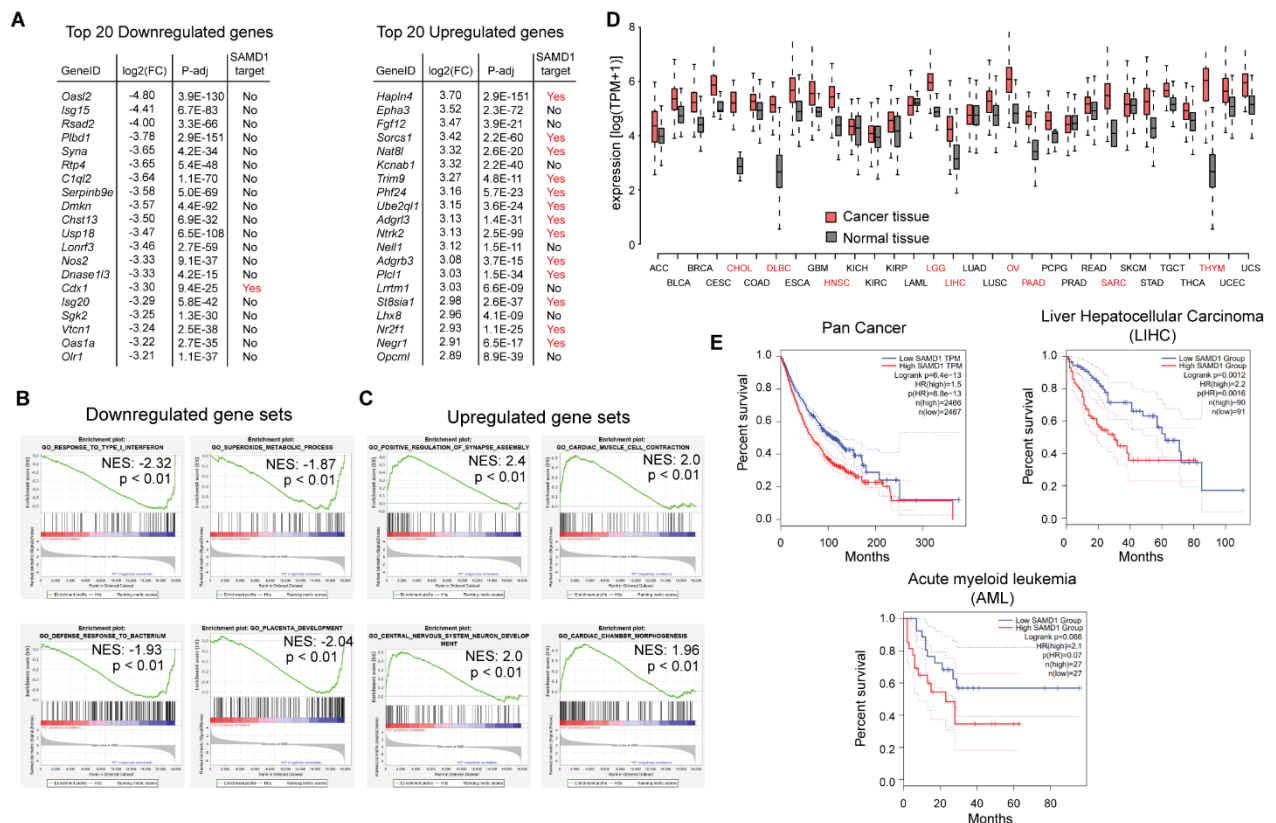
A

Top 20 Downregulated genes

GeneID	log2(FC)	P-adj	SAMD1 target
Oasl2	-4.80	3.9E-130	No
Isg15	-4.41	6.7E-83	No
Rsad2	-4.00	3.3E-66	No
Plbd1	-3.78	2.9E-151	No
Syna	-3.65	4.2E-34	No
Rtp4	-3.65	5.4E-48	No
C1ql2	-3.64	1.1E-70	No
Serpinb9e	-3.58	5.0E-69	No
Dmkn	-3.57	4.4E-92	No
Chst13	-3.50	6.9E-32	No
Usp18	-3.47	6.5E-108	No
Lonrf3	-3.46	2.7E-59	No
Nos2	-3.33	9.1E-37	No
Dnase1l3	-3.33	4.2E-15	No
Cdx1	-3.30	9.4E-25	Yes
Isg20	-3.29	5.8E-42	No
Sgk2	-3.25	1.3E-30	No
Vtcn1	-3.24	2.5E-38	No
Oas1a	-3.22	2.7E-35	No
Olr1	-3.21	1.1E-37	No

Top 20 Upregulated genes

GeneID	log2(FC)	P-adj	SAMD1 target
Hapln4	3.70	2.9E-151	Yes
Epha3	3.52	2.3E-72	No
Fgf12	3.47	3.9E-21	No
Sorcs1	3.42	2.2E-60	Yes
Nat8l	3.32	2.6E-20	Yes
Kcnab1	3.32	2.2E-40	No
Trim9	3.27	4.8E-11	Yes
Phf24	3.16	5.7E-23	Yes
Ube2ql1	3.15	3.6E-24	Yes
Adgrl3	3.13	1.4E-31	Yes
Ntrk2	3.13	2.5E-99	Yes
Nell1	3.12	1.5E-11	No
Adgrb3	3.08	3.7E-15	Yes
Plcl1	3.03	1.5E-34	Yes
Lrrtm1	3.03	6.6E-09	No
St8sia1	2.98	2.6E-37	Yes
Lhx8	2.96	4.1E-09	No
Nr2f1	2.93	1.1E-25	Yes
Negr1	2.91	6.5E-17	Yes
Opcml	2.89	8.9E-39	No

D expression [log(TPM+1)]

Cancer tissue — Normal tissue

ACC BLCA BRCA CESC CHOL COAD DLBC ESCA GBM HNSC KICH KIRC KIRP LAML LGG LIHC LUAD LUSC OV PAAD PCPG PRAD READ SARC SKCM STAD TGCT THCA THYM UCEC UCS

B Downregulated gene sets

Enrichment plot: GO_RESPONSE_TO_TYPE_I_INTERFERON — NES: -2.32, p < 0.01
Enrichment plot: GO_SUPEROXIDE_METABOLIC_PROCESS — NES: -1.87, p < 0.01
Enrichment plot: GO_DEFENSE_RESPONSE_TO_BACTERIUM — NES: -1.93, p < 0.01
Enrichment plot: GO_PLACENTA_DEVELOPMENT — NES: -2.04, p < 0.01

C Upregulated gene sets

Enrichment plot: GO_POSITIVE_REGULATION_OF_SYNAPSE_ASSEMBLY — NES: 2.4, p < 0.01
Enrichment plot: GO_CARDIAC_MUSCLE_CELL_CONTRACTION — NES: 2.0, p < 0.01
Enrichment plot: GO_CENTRAL_NERVOUS_SYSTEM_NEURON_DEVELOPMENT — NES: 2.0, p < 0.01
Enrichment plot: GO_CARDIAC_CHAMBER_MORPHOGENESIS — NES: 1.96, p < 0.01

E

Pan Cancer — Percent survival vs Months
Low SAMD1 TPM / High SAMD1 TPM
Logrank p=6.4e-13
HR(high)=1.5
p(HR)=8.8e-13
n(high)=2466
n(low)=2467

Liver Hepatocellular Carcinoma (LIHC) — Percent survival vs Months
Low SAMD1 Group / High SAMD1 Group
Logrank p=0.0012
HR(high)=2.2
p(HR)=0.0016
n(high)=90
n(low)=91

Acute myeloid leukemia (AML) — Percent survival vs Months
Low SAMD1 Group / High SAMD1 Group
Logrank p=0.066
HR(high)=2.1
p(HR)=0.07
n(high)=27
n(low)=27

Fig. S8. Role of SAMD1 during ES cell differentiation and in cancer
(**A**) Top 20 up- and downregulated genes after 7 days of undirected differentiation.
(**B, C**) Example GSEA of selected gene sets that become down- (B) or upregulated (C) in the SAMD1 KO cells after 7 days of differentiation.
(**D**) Expression of SAMD1 in normal and tumor samples. Cancer names in red have significantly increased SAMD1 expression in the cancer sample.
(**E**) Kaplan-Meier-Survival plot comparing the upper (red) and lower (blue) quartile of SAMD1 expression in all available TCGA cancer types, in Liver Hepatocellular Carcinoma (LIHC) or Acute Myeloid Leukaemia (AML). Plots in (D) and (E) were created using GEPIA (*37*).

Article

The CpG Island-Binding Protein SAMD1 Contributes to an Unfavorable Gene Signature in HepG2 Hepatocellular Carcinoma Cells

Clara Simon [1], Bastian Stielow [1], Andrea Nist [2], Iris Rohner [1], Lisa Marie Weber [1], Merle Geller [1], Sabrina Fischer [1], Thorsten Stiewe [2] and Robert Liefke [1,3,*]

1   Institute of Molecular Biology and Tumor Research (IMT), Faculty of Medicine, Philipps University of Marburg, 35043 Marburg, Germany; clara.simon@imt.uni-marburg.de (C.S.); stielow@imt.uni-marburg.de (B.S.); rohner@staff.uni-marburg.de (I.R.); lisa.weber@imt.uni-marburg.de (L.M.W.); geller@students.uni-marburg.de (M.G.); fische4t@staff.uni-marburg.de (S.F.)
2   Genomics Core Facility, Faculty of Medicine, Institute of Molecular Oncology, Member of the German Center for Lung Research (DZL), Philipps University of Marburg, 35043 Marburg, Germany; andrea.nist@imt.uni-marburg.de (A.N.); stiewe@uni-marburg.de (T.S.)
3   Department of Hematology, Oncology, and Immunology, University Hospital Giessen and Marburg, 35043 Marburg, Germany
*   Correspondence: robert.liefke@imt.uni-marburg.de; Tel.: +49-6421-28-66697

Simple Summary: Hepatocellular carcinoma (HCC) belongs to the most common cancer types and is the third leading cause of cancer-related deaths. To gain insight into the molecular mechanisms of liver cancer cells, we assessed the role of the CpG island regulator SAMD1, which is highly expressed in liver cancer tissues and associated with poor prognosis. We demonstrate that the deletion of SAMD1 in HepG2 cells leads to aberrant gene regulation and to a gene signature linked to a better prognosis. These results establish SAMD1 as a potentially important player in HCC.

Abstract: The unmethylated CpG island-binding protein SAMD1 is upregulated in many human cancer types, but its cancer-related role has not yet been investigated. Here, we used the hepatocellular carcinoma cell line HepG2 as a cancer model and investigated the cellular and transcriptional roles of SAMD1 using ChIP-Seq and RNA-Seq. SAMD1 targets several thousand gene promoters, where it acts predominantly as a transcriptional repressor. HepG2 cells with SAMD1 deletion showed slightly reduced proliferation, but strongly impaired clonogenicity. This phenotype was accompanied by the decreased expression of pro-proliferative genes, including MYC target genes. Consistently, we observed a decrease in the active H3K4me2 histone mark at most promoters, irrespective of SAMD1 binding. Conversely, we noticed an increase in interferon response pathways and a gain of H3K4me2 at a subset of enhancers that were enriched for IFN-stimulated response elements (ISREs). We identified key transcription factor genes, such as *IRF1*, *STAT2*, and *FOSL2*, that were directly repressed by SAMD1. Moreover, SAMD1 deletion also led to the derepression of the PI3K-inhibitor *PIK3IP1*, contributing to diminished mTOR signaling and ribosome biogenesis pathways. Our work suggests that SAMD1 is involved in establishing a pro-proliferative setting in hepatocellular carcinoma cells. Inhibiting SAMD1's function in liver cancer cells may therefore lead to a more favorable gene signature.

Keywords: hepatocellular carcinoma; CpG islands; chromatin; SAMD1; transcription; MYC; mTOR; interferon

1. Introduction

Hepatocellular carcinoma (HCC) globally belongs to the most common cancer types, is associated with major health-related impairments, and is the third leading cause of

cancer-related deaths [1]. Work in recent years has identified numerous molecular targets and signaling pathways that are suitable for treating liver cancer patients [2]. Nonetheless, highly effective antitumor agents are still missing, demonstrating the need to explore further factors as potential therapeutic targets and pharmaceutical treatment options for HCC.

In addition to the aberrant alteration of many signaling pathways [3,4], HCC is also characterized by significant changes in the transcriptional network [5]. In particular, the MYC oncoprotein is often highly expressed in liver cancer cells, which induces aberrant proliferation [6]. MYC affects many different biological processes, such as transcription, translation, and DNA replication [7]. However, despite the important role of MYC, it is considered undruggable [8]. Thus, the identification of molecules and processes that are required for the functioning of MYC may allow the establishment of alternative strategies to target MYC-driven tumors.

We recently identified SAMD1 (SAM domain-containing protein 1) to directly interact with unmethylated CpG islands (CGIs) [9,10], which are key regulatory elements at most gene promoters [11]. SAMD1 is associated with several chromatin modulators, including the histone demethylase KDM1A (lysine-specific histone demethylase 1A) and the chromatin binding protein L3MBTL3 (lethal(3)malignant brain tumor-like protein 3), and it plays a role in gene regulation [9]. In mouse ES cells, the deletion of SAMD1 leads to the derepression of its target genes and an alteration of multiple biological pathways, implicating a pleiotropic role of SAMD1. Indeed, SAMD1 has been proposed to play a role in atherosclerosis [12], was identified in a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) screen to be essential for the growth of K562 cancer cells [13], and via a GWAS study, the nonsynonymous variant E338D of SAMD1 was linked to the immune response after malaria infection [14]. Total knockout of SAMD1 leads to impaired angiogenesis and is embryonic lethal [15]. This suggests a potentially versatile and complex biological function of SAMD1. However, the current literature provides only a very limited picture of SAMD1's role in a physiological and pathophysiological context, prompting us to further examine this underexplored protein.

In the present study, we aimed to gain the first insights into the role of SAMD1 in human cancer cells, with a focus on the hepatocellular carcinoma cell line HepG2, which is a commonly used liver cancer cell line [16,17] and where SAMD1 is strongly expressed. We addressed the consequences of SAMD1 deletion on the biological properties, gene expression, and chromatin landscape of these cells. We found that SAMD1 deletion slightly decreased the proliferation rate, but significantly impaired the clonogenicity of the cells. This phenotype is associated with decreased expression of MYC target genes. Furthermore, we observed an impaired stem cell-like signature. Via genome-wide ChIP-Seq experiments, we confirmed a chromatin regulatory role of SAMD1 in HepG2 cells and showed that it binds to unmethylated CGIs. Moreover, our work indicates that SAMD1 represses the gene transcription of the PI3K (phosphoinositide-3-kinase) interacting protein PIK3IP1 and the tumor suppressor IRF1 (interferon regulatory factor 1), and likely several other factors, to build a transcriptional network that is associated with a poorer prognosis for patients with HCC. Thus, this work supports that interfering with SAMD1's function in human liver cancer cells could be a valid treatment option for hepatocellular carcinomas.

2. Materials and Methods

2.1. Cell Culture

HepG2 cells were cultured with MEM, GlutaMAX™ (Thermo Fisher Scientific, Waltham, MA, USA; 41090036) supplemented with 10% fetal bovine serum (Merck; F7524), 1% penicillin–streptomycin (Thermo Fisher Scientific, Waltham, MA, USA; 15140148), and 1× nonessential amino acids (Thermo Fisher Scientific, Waltham, MA, USA; 11140050).

2.2. SAMD1 Knockout

SAMD1 knockout in HepG2 cells was conducted via the Lenti-CRISPR system. Two different single guide RNAs targeting SAMD1 (sg1: AGCGCATCTGCCGGATGGTG; sg2:

GAGCATCTCGTACCGCAACG) and a nonspecific control single guide RNA were transfected using Opti-MEM™ (Thermo Fisher Scientific, Waltham, MA, USA; 31985062) and FuGENE® HD Transfection Reagent (Promega, Madison, WI, USA; E2311). Single clones were selected using 2 µg/mL puromycin (Merck, Kenilworth, NJ, USA; 58-58-2). The knockout was confirmed by Western blot analysis using a commercial SAMD1 antibody (Bethyl, Montgomery, TX, USA; A303-578A-M).

2.3. Nuclear Extract Preparation

To obtain the nuclear extract, the cytoplasmic fraction was removed by incubating harvested cells for 10 min at 4 °C in low salt buffer (10 mM HEPES/KOH pH = 7.9; 10 mM KCl; 1.5 mM MgCl2; 1xPIC (cOmplete™, Protease Inhibitor Cocktail [Roche, Basel, Switzerland; 04693116001]); 0.5 mM PMSF). After centrifugation, the remaining pellet was dissolved in high salt buffer (20 mM HEPES/KOH pH = 7.9; 420 mM NaCl; 1.5 mM MgCl2; 0.2 mM EDTA; 20% glycerol; 1xPIC; 0.5 mM PMSF) and incubated for 20 min at 4 °C while shaking. Subsequently, the lysates were centrifuged, and the supernatant containing the nuclear fraction was further analyzed by Western blotting.

2.4. Subcellular Fractionation

A subcellular protein fractionation kit for cultured cells (Thermo Fisher Scientific, Waltham, MA, USA; 78840) was used for fractionation experiments according to the manufacturer's instructions. A 10 cm dish format was applied, which corresponded to a packed cell volume of 20 µL per well. Localization of SAMD1 was determined using a homemade SAMD1 antibody recognizing the SAM domain [9]. As loading controls for the respective fractions, a homemade SP1 antibody [18], anti-tubulin (Merck, Kenilworth, NJ, USA; MAB3408), and anti-H2Aub (Cell Signaling Technology, Danvers, MA, USA; 8240) were applied. For the detection of histone marks, the chromatin-bound fraction was used. H3 (Abcam, Cambridge, UK; ab1791), H3K4me2 (Diagenode, Denville, NJ, USA; C15410035), H3K4me3 (Diagenode, Denville, NJ, USA; C15410003) and H3K14ac (Abcam, Cambridge, UK; ab52946) antibodies were applied. The signal was quantified using the ImageLab software (v5.2.1, Bio-Rad, Hercules, CA, USA) and normalized to the H3 signal.

2.5. Immunofluorescence Staining

For immunofluorescence staining, HepG2 cells were seeded on gelatin-coated coverslips. Cells were fixed with 4% formaldehyde (*w/v*), methanol-free (Thermo Fisher Scientific, Waltham, MA, USA; PI28906), and subsequently permeabilized with 0.5% Triton X-100 in PBS. Blocking was performed with 10% FBS in PBS. To detect SAMD1, a homemade SAMD1 antibody recognizing the SAM domain was diluted 1:500 in blocking solution. After primary antibody incubation for 1 h in a wet chamber, the cells were washed three times with 0.5% Triton X-100 in PBS. Secondary antibody incubation was conducted using Alexa Fluor 488 and goat anti-rabbit IgG (H+L) (Thermo Fisher Scientific, Waltham, MA, USA; A-11008) at a 1:1000 dilution. Following three washing steps, the coverslips were mounted onto microscopy slides using VECTASHIELD® Antifade Mounting Medium with DAPI (VECTOR Laboratories, Burlingame, CA, USA; H-1200), and the edges were sealed with nail polish.

2.6. Proliferation Assay

To determine the proliferation rates, cells were seeded on 6-well plates at a density of 1×10^5 cells per well. Cell viability was determined 1, 3, and 7 days after seeding using the MTT assay by adding 90 µL of 5 mg/mL thiazolyl blue $\geq$ 98% (Carl Roth, Karlsruhe, Germany; 4022) to each well. After 1 h, the medium was aspirated, and stained cells were dissolved in 400 µL of lysis buffer (80% isopropanol; 10% 1 M HCl; 10% Triton X-100) and diluted further if necessary. Absorption was measured at 595 nm using a plate reader. All values were normalized to day 1 to compensate for variations in seeding density. The mean value of three biological replicates was determined.

2.7. Colony Formation Assay

To examine the ability of cells to form colonies, the cells were seeded at low density (1×10^3 cells per well on 6-well plates) and cultured for 11 days. Next, the cells were washed once with PBS and then fixed with 100% methanol for 20 min. Afterward, the cells were stained for 5 min with 0.5% crystal violet in 25% methanol. To remove excess color, the plates were washed with dH2O until single colonies were visible. Images were taken, and colonies were counted using ImageJ Fiji (v1.53p). The mean value of three biological replicates was determined.

2.8. Starvation Assay

To investigate the sensitivity of cells to starvation conditions, cells were seeded on 6-well plates at a density of 1×10^5 cells per well and cultured for 24 h. Next, the culture medium was replaced by medium lacking FBS. A control with standard culture medium was included. After 72 h, cell viability was determined using the MTT assay (see Proliferation Assay).

2.9. RNA Preparation

For RNA isolation, cells were cultivated on 6-well plates up to 80–100% confluency. RNA was prepared according to the manufacturer's manual using the RNeasy Mini Kit (Qiagen, Hilden, Germany; 74004) including an on-column DNA digest.

2.10. cDNA Synthesis

The Tetro cDNA Synthesis Kit (Bioline, London, UK; BIO-65043) was used to transcribe mRNA into cDNA according to the manufacturer's manual. Samples were incubated at 45 °C for 50 min followed by 5 min at 85 °C to inactivate Tetro RT. Subsequently, cDNA was diluted 1:20 to be used in RT-qPCR.

2.11. RT-qPCR

For analysis by real-time quantitative PCR, MyTaq™ Mix (Bioline, London, UK; BIO-25041) was used. For gene expression analysis, values were normalized to GAPDH and B2M expression. The qPCR primers used are presented in Supplementary Table S1.

2.12. Chromatin Preparation

To prepare chromatin, cells were seeded on 15 cm plates at 3×10^6 cells per plate and cultivated until reaching 70–90% confluence. First, 1% formaldehyde was added to the medium, and the plates were slowly swayed for 10 min to fix the cells. The fixation was stopped by adding 125 mM glycine for 5 min. Subsequently, the cells were washed twice with PBS and scraped in 1 mL cold buffer B (10 mM HEPES/KOH, pH = 6.5; 10 mM EDTA; 0.5 mM EGTA; 0.25% Triton X-100) per 15 cm plate. All plates containing the same cell line were pooled in a 15 mL tube. The tubes were centrifuged for 5 min at 2000 rpm and 4 °C. The supernatant was removed, and the pellet was resuspended in 1 mL cold buffer C (10 mM HEPES/KOH, pH = 6.5 I 10 mM EDTA; 0.5 mM EGTA; 200 mM NaCl) per 15 cm plate followed by a 15 min incubation time on ice. Then, the tubes were centrifuged with the same settings as mentioned before. After removing the supernatant, the pellet was resuspended in 200 µL cold buffer D (50 mM Tris/HCl, pH = 8.0; 10 mM EDTA; 1% SDS; 1xPIC (cOmplete™, Protease Inhibitor Cocktail [Roche, Basel, Switzerland; 04693116001])) per 15 cm plate, vortexed, and incubated for 10–20 min on ice. To shear the chromatin, the samples were sonicated two times for 7 min each using a precooled Bioruptor® (Diagenode, Denville, NJ, USA). The samples were centrifuged for 10 min at 13,000 rpm and 4 °C. The supernatant contained the sheared chromatin.

2.13. Chromatin Immunoprecipitation

Chromatin immunoprecipitation (ChIP) for ChIP-qPCR was performed according to the one-day ChIP kit protocol (Diagenode, Denville, NJ, USA; C01010080). For each

ChIP, 3 μg of either IgG control antibody (Diagenode, Denville, NJ, USA; C15410206) or homemade SAMD1 antibody recognizing the SAM domain was applied [9]. For the ChIP of histone marks, 1 μg of H3K4me2 antibody (Diagenode, Denville, NJ, USA; C15310035) was used.

To prepare the samples for ChIP sequencing, the one-day ChIP kit protocol was used as described above, but the DNA purification was modified. For DNA elution, the beads were incubated with 230 μL elution buffer (100 mM NaHCO3; 1% SDS) for 30 min at room temperature while shaking. Afterward, the samples were centrifuged at 13,000 rpm for 1 min, and 200 μL of supernatant was transferred to a fresh tube. The input DNA was dissolved in 50 μL of dH2O, and 150 μL of elution buffer was added to obtain an equal volume in all samples. Eight microliters of 5 M NaCl were added to each sample, and the samples were incubated at 65 °C overnight to reverse the cross-linking.

On the next day, 8 μL of 1 M Tris/Cl pH 6.5, 4 μL 0.5 M EDTA, and 2 μL of Proteinase K (10 μg/μL) were added to each sample. All samples were incubated at 45 °C for 1 h while shaking. The DNA was purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany; 28104), whereby all samples prepared with the same antibody were pooled on the same column. To elute the DNA, the columns were incubated for 1 min with 30 μL of sterile 2 mM Tris/Cl, pH 8.5, and centrifuged at 13,000 rpm for 1 min.

The concentration of the samples was determined using the Quant-iT™ dsDNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA; Q33120) and the NanoDrop™ 3300 (Thermo Fisher Scientific, Waltham, MA, USA). At least 4 ng of DNA was used for library preparation.

2.14. Library Preparation and Next-Generation Sequencing

Next-generation sequencing was performed at the Genomics Core Facility Marburg (Center for Tumor Biology and Immunology, Hans-Meerwein-Str. 3, 35043 Marburg, Germany). For ChIP-Seq, the Microplex library preparation kit v2 (Diagenode, Denville, NJ, USA, C05010012) was used for indexed sequencing library preparation with chromatin immunoprecipitated DNA. Libraries were purified on AMPure magnetic beads (Beckman, Brea, CA, USA; A6388). RNA was prepared as described in "RNA preparation", and integrity was assessed on an Experion StdSens RNA Chip (Bio-Rad, Hercules, CA, USA; 7007103). RNA-Seq libraries were prepared using the TruSeq Stranded mRNA Library Prep kit (Illumina, San Diego, CA, USA, 2002059). RNA-Seq and ChIP-Seq libraries were quantified on a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Next-generation sequencing was performed on an Illumina NextSeq 550 (Illumina, San Diego, CA, USA).

2.15. Bioinformatics Analysis

ChIP-Seq data were aligned to the human genome hg38 using Bowtie [19]. Peak calling was performed with MACS2 with standard settings [20]. The genomic distribution of SAMD1 was determined using ChIPseeker (Galaxy Version 1.28.3) [21]. Gene ontology analysis of SAMD1 target loci was performed using GREAT [22]. Bigwig files, heatmaps, and binding profiles were created using Galaxy/DeepTools [23]. Enhancers were defined as H3K4me2 peaks that did not overlap with promoter sites. Enriched motifs at enhancers were identified using HOMER [24]. ChIP-Seq tracks were visualized using the UCSC browser [25]. Promoter definition and CpG islands were downloaded from the UCSC table browser. The transcription factor list was obtained from http://humantfs.ccbr.utoronto.ca/ (accessed on 11 January 2022) [26].

RNA-Seq samples were aligned to the human transcriptome GENCODE v32 using RNA-Star (2.7.2b) [27]. Reads per gene were calculated using feature counts (2.0.1). Differentially regulated genes and normalized read counts were determined using DESeq2 (2.11.40.6) [28]. Genes with at least 0.5-fold (log2) difference and a *p*-value below 0.01 were considered differentially expressed genes. For the correlation analysis of SAMD1 expression with proliferation, the average expression of 11 proliferation-associated genes

(*BIRC5*, *CCNB1*, *CDC20*, *NUF2*, *CEP55*, *NDC80*, *MKI67*, *PTTG1*, *RRM2*, *TYMS*, *UBE2C*) [29] was used. Gene set enrichment analysis (GSEA) [30] was performed with standard settings.

The following internet tools and databases were used: GePIA http://gepia.cancer-pku.cn/ (accessed on 11 January 2022) [22], GePIA2 http://gepia2.cancer-pku.cn/ (accessed on 11 January 2022) [31], ProteinAtlas https://www.proteinatlas.org/ (accessed on 11 January 2022) [32], UALCAN http://ualcan.path.uab.edu/ (accessed on 11 January 2022) [33], HCCDB http://lifeome.net/database/hccdb/home.html (accessed on 11 January 2022) [34], GREAT: http://great.stanford.edu/public/html/ (accessed on 11 January 2022) [35], KM-Plotter: https://kmplot.com/ (accessed on 11 January 2022) [36], GALAXY: https://usegalaxy.org/ (accessed on 11 January 2022) [37], and UCSC Browser: https://genome.ucsc.edu/ (accessed on 11 January 2022) [25].

The following public datasets were used: HepG2 WGBS: GSM3633977 [31], HepG2 MYC: GSM822291 [31], HepG2 HES1: GSM803448 [32], HepG2 ATF4: ENCSR669LCD_2, GSE96304 [31], HepG2 H3K4me1: GSM798321 [31], HepG2 H3K4me3: GSM733737 [31], HepG2 RNA Polymerase II: GSM1670896 [33], mESCs MeDIP-Seq: GSM881346 [34], and mESCs ChIP SAMD1: GSM4287311 [9].

2.16. Statistical Analysis

The significance of the biological experiments was determined with an unpaired Student's *t*-test. The significance of gene expression changes was determined by the DESeq2 or GSEA tool. The significance of the gene expression differences between two gene groups was determined via ANOVA. The significance of patient survival was determined by the GePIA tool [22].

3. Results

3.1. SAMD1 Is Highly Expressed in Liver Cancer and Associated with a Poor Prognosis

We previously characterized SAMD1 as a novel unmethylated CGI-binding protein in mouse ES cells [9], but its function in other biological contexts remained unclear. The investigation of public cancer gene expression data from TCGA showed that the *SAMD1* transcript is mostly upregulated in cancerous tissues compared to normal tissues (Figure 1a) [35]. Significant upregulation of *SAMD1* occurs in cholangiocarcinoma (CHOL), diffuse large B-cell lymphoma (DLBC), head and neck squamous cell carcinoma (HNSC), brain lower-grade glioma (LGG), liver hepatocellular carcinoma (LIHC), ovarian serous cystadenocarcinoma (OV), pancreatic adenocarcinoma (PAAD), sarcoma (SARC), and thymoma (THYM). One of the strongest increases in SAMD1 expression in tumor versus normal tissues was found in liver cancer (Figure 1a). This increased expression can also be observed in other HCC datasets (Supplementary Figure S1a) [36]. The data from TCGA further suggest that high *SAMD1* expression correlates with poor patient survival (Figure 1b) [37]. This association was independent of sex, alcohol consumption, or virus infection (Supplementary Figure S1b). Furthermore, we found that the expression level of *SAMD1* increases in advanced cancer stages and that *SAMD1* expression is highest in cancer samples that also bear a p53 mutation, which is a common liver cancer driver (Figure 1c,d) [38]. The expression of *SAMD1* positively correlates with the expression of common proliferation markers (Figure 1e), suggesting that high SAMD1 levels are associated with higher aggressiveness of cancer.
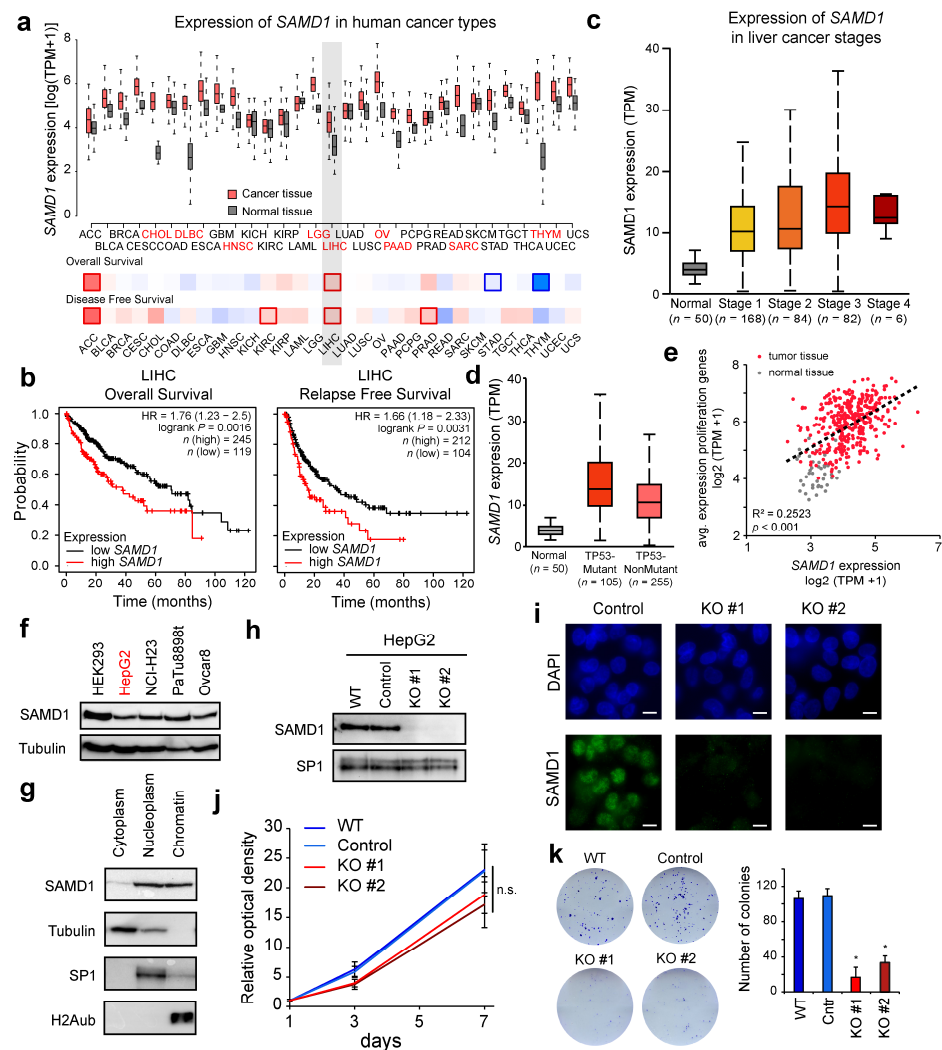
Figure 1. *SAMD1* is highly expressed in liver cancer and is associated with poor prognosis, and its deletion impairs the cellular properties of HepG2 cells. (**a**) Comparison of SAMD1 expression in normal and cancer tissue using data from TCGA, visualized by GePIA [35]. Red cancer names indicate significant upregulation. The lower boxes indicate the hazard ratio of patients' overall and disease-free survival. Red = high SAMD1 expression correlated with poor prognosis. Blue = high SAMD1 expression correlated with good prognosis. Bold frames indicate significance. (**b**) Kaplan–Meier survival curves show the correlation of SAMD1 expression with patient survival. Plots were made using KM plotter [37]. (**c**) Expression of SAMD1 in distinct liver cancer stages derived from UALCAN [38]. Whisker plots represent the upper and lower quartiles with 5 and 95% whiskers. (**d**) SAMD1 expression in patient samples with and without TP53 mutations compared to healthy liver tissue ("Normal"). Data derived from UALCAN [38]. Box plots represent upper and lower quartiles of the data with 5 and 95% whiskers. (**e**) Correlation of SAMD1 expression with the average expression of 11 proliferation-associated genes in liver tissues. The *p*-value was calculated using ANOVA. (**f**) Western blot of SAMD1 in various human cell lines. HEK293—human embryonic kidney, HepG2—liver cancer, NCI-H23—lung cancer, PaTu8988t—pancreatic cancer, Ovcar8—ovarian carcinoma. The used cell line in this study (HepG2) is marked red. (**g**) Cellular fractionation of HepG2 cells followed by Western blotting. (**h**) Western blot of two SAMD1 knockout clones. (**i**) SAMD1 immunofluorescence in control and SAMD1 KO cells. Bar = 10 μM. (**j**) Proliferation of wild-type, control, and SAMD1 KO HepG2 cells. Error bars indicate the SD of three biological replicates. (**k**) Representative picture of colony formation of wild-type, control, and SAMD1 KO HepG2 cells and quantification. Error bars indicate the SD of three biological replicates. * *p* < 0.05 (Student's *t*-test). In (**h**–**k**) "#" refers to two distinct SAMD1 KO clones.

Thus, these data support the idea that SAMD1 may have a pro-proliferative role in liver cancer cells, which could contribute to worse survival of the patients. Since the role of SAMD1 in cancer has not yet been explored, we decided to investigate the cellular function of SAMD1 in liver cancer cells in greater detail.

3.2. SAMD1 Deletion Impairs the Proliferation and Biological Properties of Liver Cancer Cells

Both public data (Supplementary Figure S1c) and Western blotting experiments (Figure 1f) suggest that the SAMD1 protein is abundantly expressed in most human cancer cell lines. To address the role of SAMD1 in liver cancer, we used the commonly applied HepG2 cell line [16], in which SAMD1 can easily be detected via Western blotting (Figure 1f). Notably, contradictory findings were reported about the cellular localization of SAMD1 [10]. In a previous publication, SAMD1 was described to be secreted [12], while our work in mouse ES cells suggested the presence of SAMD1 in the nucleus and bound to chromatin [9]. In HepG2 cells, we also found SAMD1 predominantly in the nucleoplasm and the chromatin fraction, while only a small part was present in the cytoplasmic fraction (Figure 1g). This finding suggests that SAMD1 plays a role in chromatin, not only in mouse stem cells but also in human cancer cells.

To address the potential biological role of SAMD1, we created SAMD1 knockout cells using CRISPR/Cas9 targeting with two different single guide RNAs (sgRNAs), leading to two different KO clones (KO#1 and KO#2). The successful knockout was validated by Western blot and immunofluorescence analysis (Figure 1h,i). SAMD1 knockout HepG2 cells showed a slightly reduced proliferation rate (Figure 1j), consistent with the hypothesis that SAMD1 contributes to liver cancer cell proliferation. To further address the consequence of SAMD1 deletion, we performed colony formation assays, which assessed the ability of the cells to proliferate from single cells. The ability of SAMD1 KO cells to form colonies was significantly reduced in comparison to that of WT and control cells (Figure 1k), suggesting that the absence of SAMD1 strongly impairs the clonogenic survival of HepG2 cells. Presumably, the SAMD1 KO cells require certain factors secreted by other cells for growth, which might explain the different impact of SAMD1 KO on proliferation (Figure 1j) and colony formation (Figure 1k).

Together, these experiments suggest a pivotal role of SAMD1 in maintaining the optimal growth of HepG2 cells.

3.3. SAMD1 Deletion Diminishes MYC Target Gene Expression and Stem Cell Signature in HepG2 Cells

To understand the reasons why SAMD1 deletion reduces cell proliferation and interferes with cellular properties, we investigated the consequences on gene expression using unbiased RNA-Seq experiments. We used three independent SAMD1 KO clones to avoid clone-specific effects. PCA (principal component analysis) demonstrated that the SAMD1 knockout cells were highly distinct from the control cells (Figure 2a). We identified 762 significantly upregulated and 359 significantly downregulated genes ($p < 0.01$, log2-fold change > 0.5) in the absence of SAMD1 (Figure 2b,c). Interestingly, the most upregulated gene was the *L3MBTL3* gene, which we have already observed in mouse ES cells to become strongly upregulated upon SAMD1 deletion [9]. The L3MBTL3 protein is a direct interacting partner of SAMD1 [9] and is involved in chromatin regulation and transcriptional repression [39,40]. Considering the observation that *L3MBTL3* gene expression is strongly affected in these two unrelated cell lines, we postulate that *L3MBTL3* upregulation could be a compensatory effect and is possibly a common consequence after SAMD1 deletion. We compared the gene expression changes upon SAMD1 deletion in HepG2 and mouse ES cells to address in further detail to what extent the transcriptional consequences upon SAMD1 deletion are similar in these distinct cell lines. We found more pronounced effects in HepG2 cells overall, but surprisingly little correlation between these two cell lines (Figure 2d), in addition to the upregulation of *L3MBTL3* and the downregulation of *SAMD1* itself.

These findings suggest that the consequence of SAMD1 deletion is possibly highly cell type-specific.

To investigate the role of SAMD1 in the context of HepG2 cells, we performed unbiased gene set enrichment analyses (GSEA). Initially focusing on common hallmarks, we observed a strong downregulation of MYC target genes (Figure 2e), which are of fundamental importance for cell growth [6] and the maintenance of stem cell characteristics in cancer cells [41]. Consistently, we also found reduced expression of genes related to G2/M checkpoints and ESC-like signatures (Figure 2e, upper right/lower left panel). On the other hand, genes related to the interferon signaling pathway were upregulated (Figure 2e, lower right panel), suggesting multiple opposing consequences upon SAMD1 deletion.
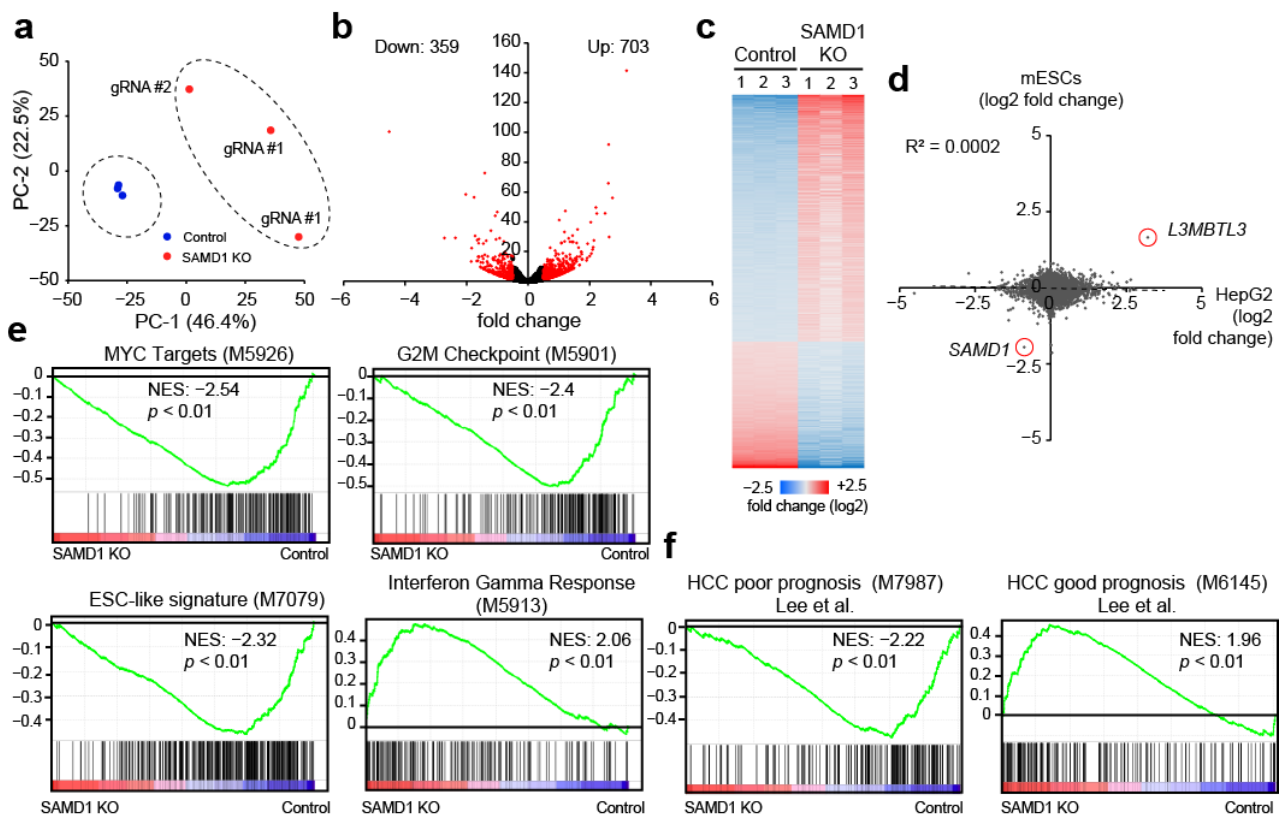
Figure 2. SAMD1 deletion leads to a more favorable transcriptional network. (**a**) Principal component analysis (PCA) of RNA-Seq data from three control samples and three independent SAMD1 KO clones (obtained with two distinct gRNAs, marked by #). (**b**) Volcano plot of RNA-Seq data. Red dots represent significantly differentially expressed genes ($p < 0.01$) with a fold change larger than 0.5 (log2). (**c**) Heatmap of differentially expressed genes from (**b**). (**d**) Comparison of gene expression changes upon SAMD1 deletion in HepG2 cells and mouse ES cells. The scale depicts log2 fold change. (**e**) Unbiased gene set enrichment analysis (GSEA) of RNA-Seq data from HepG2 cells. (**f**) GSEA of a predefined gene set associated with poor or good prognosis in HCC [42].

We then asked how the change in gene expression upon SAMD1 deletion may correlate with patient survival. Using previously curated datasets of liver cancer patients [42], we found that gene sets expressed in hepatocellular carcinoma patients with poorer prognosis become predominantly downregulated in SAMD1 KO HepG2 cells, while genes that are expressed in HCC patients with better survival become upregulated (Figure 2f).

Taken together, these data suggest that high SAMD1 expression in liver cancer cells is involved in establishing an unfavorable transcriptional network that contributes to worse prognosis for HCC patients.

3.4. SAMD1 Targets Active and Unmethylated CGIs in HepG2 Cells

Our previous work in mouse ES cells suggested that SAMD1 functions as a transcriptional repressor at actively transcribed and unmethylated CGIs [9]. To assess whether SAMD1 also binds to CGI-containing promoters in HepG2 cells, we performed ChIP-qPCR experiments and confirmed that SAMD1 binds to the *NANOS1* gene, which is one of its top targets in mouse ES cells (Figure 3a) [9]. In contrast, we did not find SAMD1 binding at *CBLN1*, another top target in mouse ES cells (Figure 3a). This inconsistent binding behavior implicates a dissimilar binding pattern of SAMD1 in HepG2 and mouse ES cells.
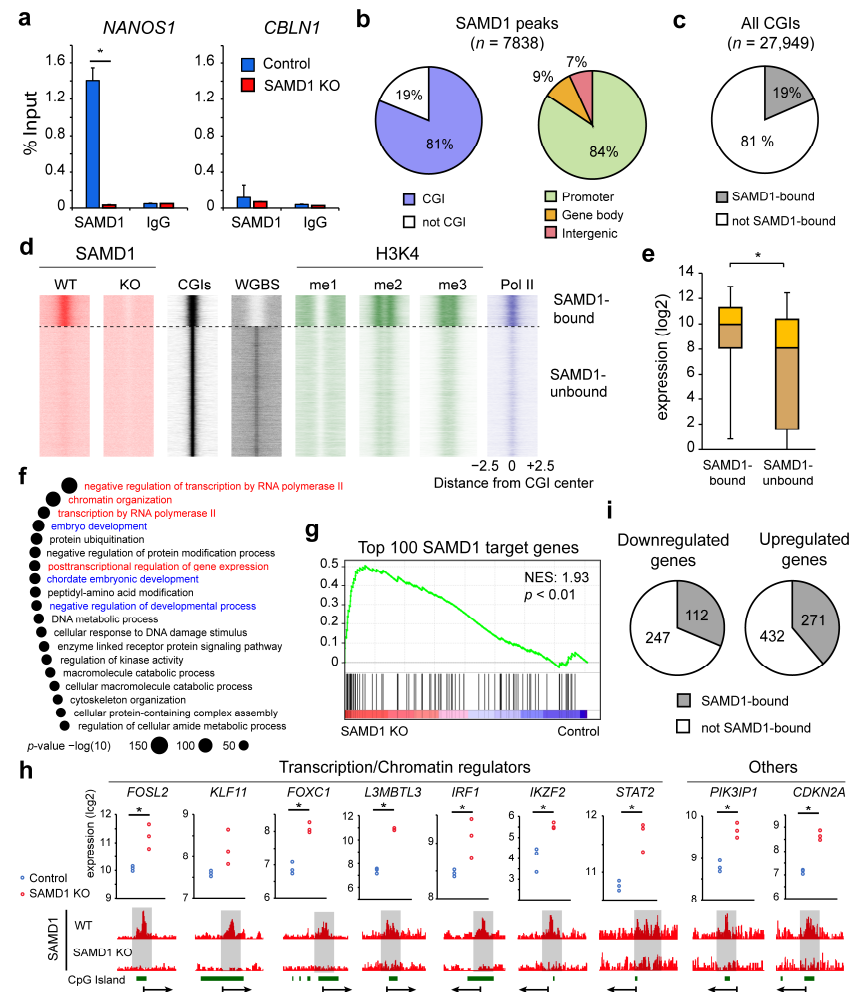
Figure 3. SAMD1 targets unmethylated CGIs in HepG2 cells, where it represses key transcription and signaling factors. (**a**) SAMD1 ChIP-qPCR of the *NANOS1* and *CBLN1* genes in HepG2 cells. Error bars indicate the SD of two technical replicates * $p < 0.05$ (Student's *t*-test). (**b**) Distribution of significant SAMD1 peaks at CpG islands and promoters, gene bodies, and intergenic regions. (**c**) Distribution of CGIs with and without bound SAMD1. (**d**) Heatmap of all CGIs clustered based on SAMD1 binding. A comparison with DNA methylation (WGBS = whole-genome bisulfite sequencing), active histone marks, and RNA polymerase II (Pol II) is shown [31]. (**e**) Expression of SAMD1-bound and SAMD1-unbound genes. Box plots represent upper and lower quartiles of the data with 5 and 95% whiskers. * $p < 0.05$. (**f**) Gene ontology of the SAMD1-bound genes using GREAT. Transcription-related ontologies are marked in red. Development-associated ontologies are marked in blue. (**g**) GSEA of the top 100 SAMD1 target genes. (**h**) Examples of direct SAMD1 target genes that become upregulated upon SAMD1 deletion. * $p < 0.01$, determined by DeSeq2. (**i**) Distribution of SAMD1-bound genes within the up- and downregulated genes.

To obtain a better picture of SAMD1 chromatin binding in HepG2 cells, we elucidated the genomic targets of SAMD1 via ChIP-Seq. Comparing the control and SAMD1 knockout cells, we identified 7838 regions with significantly enriched SAMD1 recruitment. Consistent with our results in mouse ES cells, SAMD1 was almost exclusively present at CpG islands and consequently at promoters in HepG2 cells (Figure 3b). Given that only a subset of CpG islands is targeted by SAMD1 (Figure 3c), we investigated the properties of SAMD1-bound CGIs in greater detail. Using public HepG2 DNA methylation data from the ENCODE consortium [31], we found that SAMD1-bound CGIs are mostly unmethylated (Figure 3d), consistent with our finding that SAMD1 is repelled by methylated DNA [9]. Further analysis showed that SAMD1-bound CGIs are decorated with active histone marks and show high levels of RNA polymerase II (Figure 3d), indicating that SAMD1 binds to transcriptionally active CGIs. Consistently, we found that the corresponding genes were highly expressed (Figure 3e). Gene ontology analysis using GREAT demonstrated that the SAMD1-targeted loci are involved in multiple cellular processes, including transcriptional regulation, chromatin organization, and developmental processes (Figure 3f). Thus, SAMD1 may play a pivotal role in establishing a specific chromatin landscape in HepG2 cells by modulating the expression of the highly active genes that contribute to chromatin organization.

Next, we addressed to what extent cell type-specific characteristics contribute to the chromatin binding of SAMD1. For this purpose, we compared the binding pattern of SAMD1 in HepG2 cells with that from our previous study in mouse ES cells [9]. SAMD1 levels were calculated for all gene promoters that existed in both organisms (Supplementary Figure S2a). We identified approximately 600 promoters, such as the *Nanos1/NANOS1* promoter, which are highly bound by SAMD1 in both cell types, but we also identified approximately 1000 promoters that are only bound by SAMD1 in one of the two cell lines (Supplementary Figure S2a). For example, the *Cbln1/CBLN1* promoter, which is the top SAMD1 target in mouse ES cells [9], was not bound in HepG2 cells (Supplementary Figure S2b), confirming our results from the ChIP-qPCR experiment (Figure 3a). In contrast, the promoters of *Prnp/PRNP* and *Ankrd36/ANKRD36* were SAMD1-bound in HepG2 cells, but not in mouse ES cells (Supplementary Figure S2b). This finding suggests that SAMD1 binding differs depending on the biological and genomic context. To address this in more detail, we investigated the DNA methylation levels and the presence of CGIs at each promoter (Supplementary Figure S2c,d). We found that genes that are bound in mouse ES cells, but not in HepG2 cells, differ mainly regarding their DNA methylation status. These genes are unmethylated in mouse ES cells but are mostly methylated in HepG2 cells. Conversely, genes that are bound in HepG2 cells but not in mouse ES cells are mostly unmethylated in both cell types but differ strongly regarding the CGI content at the promoters. They often have large CGIs in humans, but smaller or no CGIs in mice. The *Prnp/PRNP* and *Ankrd36/ANKRD36* promoters are examples of this phenomenon (Supplementary Figure S2a,b). Together, these results suggest that SAMD1 binds most efficiently to promoter CGIs that are large and unmethylated. Thus, dependent on the methylation status and CGI sizes, it is likely that the SAMD1 binding pattern differs from organism to organism and cell type to cell type. These differences may partially explain why the gene expression changes upon SAMD1 deletion in mouse ES cells and human HepG2 cells showed hardly any correlation (Figure 2d).

3.5. SAMD1 Directly Represses Key Transcription Factors and Signaling Regulators

To gain further insights into how SAMD1 impacts gene expression in HepG2 cells, we assessed the transcriptional changes of SAMD1-bound genes. Using our RNA-Seq data, we found that the top 100 SAMD1 target genes were preferentially upregulated upon SAMD1 deletion (Figure 3g), consistent with our previous observation that SAMD1 mainly functions as a transcriptional repressor at highly expressed genes [9]. Thus, SAMD1 may counteract activating regulators at its target genes in HepG2 cells.

Genes that are bound by SAMD1 and become derepressed upon SAMD1 deletion are likely direct downstream targets of SAMD1 in HepG2 cells. Consistent with the gene ontology analysis (Figure 3f), this group of genes includes many transcription factors (*IRF1*, *FOXC1*, *IKZF1*, *FOSL2*, *STAT2*) and chromatin regulators (*L3MBTL3*) (Figure 3h). Furthermore, we identified the *CDKN2A* gene, which encodes the tumor suppressors p16INK4a and p14ARF [43], and *PIK3IP1* as direct targets of SAMD1 (Figure 3h).

Interestingly, only approximately one-third of the genes affected in gene expression after SAMD1 KO in HepG2 cells were bound by SAMD1 itself, and up- or downregulated genes were similarly occupied by SAMD1 (Figure 3i). These findings suggest that many of the observed gene expression changes are not directly related to SAMD1 chromatin binding but are due to indirect effects. Therefore, we hypothesized that the derepression of direct SAMD1 target genes may influence signaling and transcriptional pathways, which, in turn, alter further downstream targets.

3.6. SAMD1 Deletion Increases Susceptibility of HepG2 Cells to Serum Starvation

As described above, SAMD1 deletion leads to the upregulation of its direct target *PIK3IP1*. We confirmed the results from the RNA-Seq analysis via RT-qPCR (Figure 4a). The PIK3IP1 protein acts as an inhibitor of phosphatidylinositide-3-kinases (PI3Ks) [44], which are key kinases of many pro-proliferative signaling pathways [45] and play a central role in the insulin signaling pathway in liver cells [46]. The inhibition of PI3Ks through PIK3IP1 has been shown to inhibit DNA synthesis and the survival of hepatocytes and to suppress the development of hepatocellular carcinoma [47].

We asked whether the increased expression of *PIK3IP1* upon SAMD1 deletion may impair the mTOR signaling pathway, which is downstream of PI3K and a key regulator of ribosome biogenesis and translation [48]. Increased mTOR signaling often occurs in liver cancer and is involved in drug resistance, which is why it has been proposed to be suitable for drug targeting approaches [49,50]. Gene set enrichment analysis demonstrated that mTORC1 signaling was indeed downregulated in SAMD1 KO cells (Figure 4b). Consistently, we found a strong downregulation of rapamycin-sensitive genes (Figure 4b), and the majority of ribosomal genes showed reduced expression in these cells (Figure 4c), supporting that SAMD1 knockout cells are characterized by impaired mTOR signaling and ribosome biogenesis.

Upon serum starvation, the mTOR signaling pathway is typically inactivated, which leads to autophagy to compensate for the reduced energy supply [51]. We hypothesized that impaired mTOR signaling in SAMD1 KO cells may lead to increased sensitivity to reduced serum levels. Consistent with this idea, we found that SAMD1 KO cells that were starved for 72 h showed significantly decreased cell viability compared to control cells (Figure 4d,e), suggesting that SAMD1 KO cells have a lower capacity to respond to starvation. Importantly, insulin supplementation during starvation almost completely abolished the negative effect of SAMD1 KO (Figure 4f,g). This finding supports the idea that reduced mTOR signaling in SAMD1 KO cells can be compensated by artificially activating the insulin pathway under starvation conditions.

Taken together, these experiments support the model in which SAMD1 acts as a negative regulator of *PIK3IP1* in wild-type HepG2 cells, which is a negative regulator of PI3K in hepatocellular carcinoma cells [47]. Thus, high SAMD1 expression could contribute to an augmentation of the mTOR signaling pathway in liver cancer cells (Figure 4h), which may lead to enhanced cell growth [52].
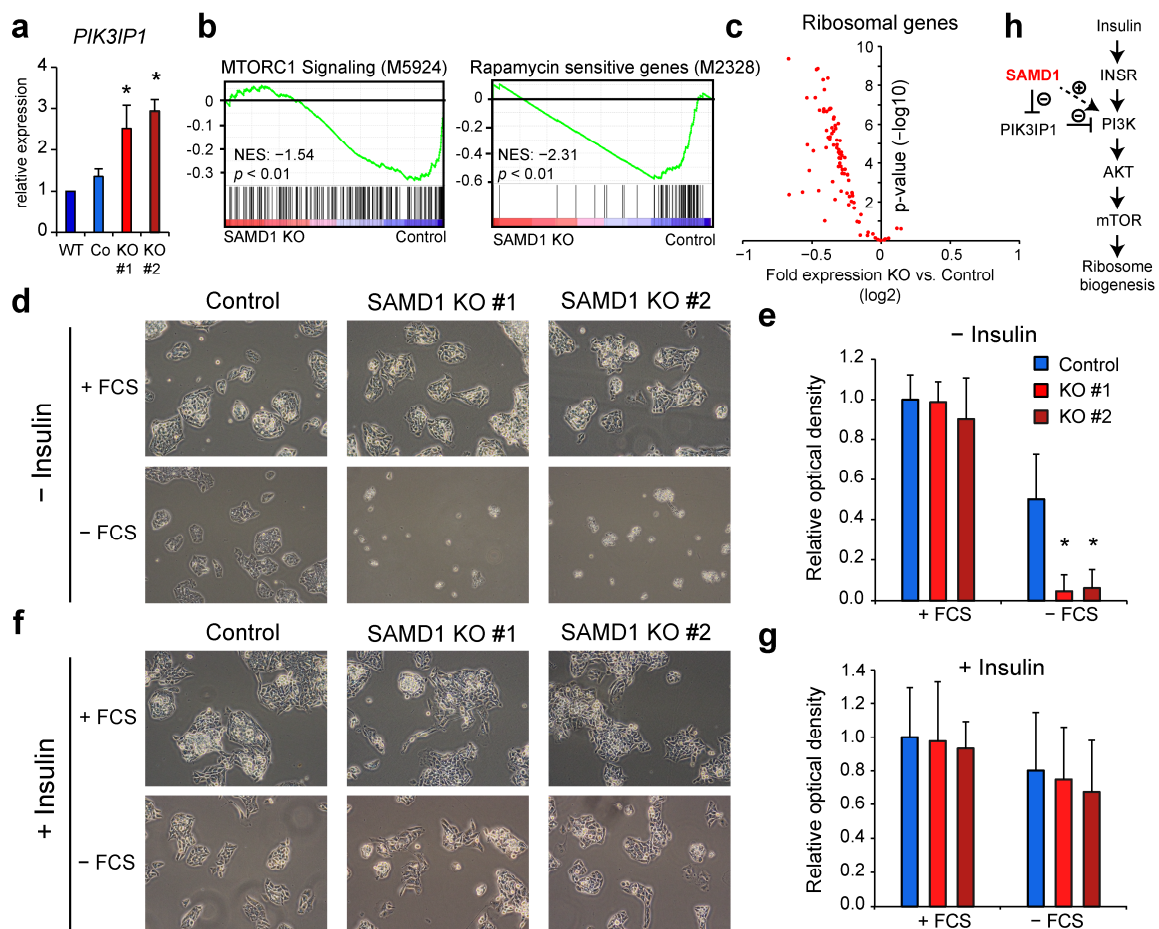
Figure 4. SAMD1 deletion impairs the mTOR signaling pathway and sensitizes cells to serum starvation. (**a**) Expression of *PIK3IP1* in wild-type, control, and SAMD1 KO cells, measured via RT-qPCR. Error bars indicate the SD of three biological replicates. * $p < 0.05$ (Student's *t*-test) (**b**) GSEA of mTOR signaling and rapamycin-sensitive genes in SAMD1 KO cells. (**c**) Volcano plot of ribosomal genes comparing expression in control and SAMD1 KO cells. (**d**) Representative bright-field microscopy of SAMD1 KO and control HepG2 cells upon serum starvation. (**e**) MTT viability assay of cells treated as in (**d**). Error bars indicate the SD of three biological replicates. * $p < 0.05$ (Student's *t*-test). (**f**) Representative bright-field microscopy of SAMD1 KO and control HepG2 cells upon serum starvation supplemented with insulin. (**g**) MTT viability assay of cells treated as in (**f**). Error bars indicate the SD of three biological replicates. * $p < 0.05$ (Student's *t*-test). (**h**) Model of SAMD1's influence on the mTOR signaling pathway. In (**a**,**d**–**g**) the "#" refers to two distinct SAMD1 KO clones.

3.7. SAMD1 Deletion Leads to Reduced H3K4me2 Levels at Most Promoters

Because SAMD1 also represses many transcription factors (Figure 3f,h), we next investigated the impact on the chromatin landscape. For this purpose, we assessed the genome-wide distribution of the active histone mark H3K4me2 in control and SAMD1-deleted cells. The H3K4me2 mark is present both at promoters and enhancers, allowing us to obtain a comprehensive picture of global changes in the chromatin landscape. Using the obtained data, we identified 65,288 significant H3K4me2 peaks in the control cells and 83,662 in the knockout cells (Figure 5a). Of those peaks, 58,798 were shared in both cell lines, while 24,864 peaks were specific for the knockout cells and 6490 for the control cells.
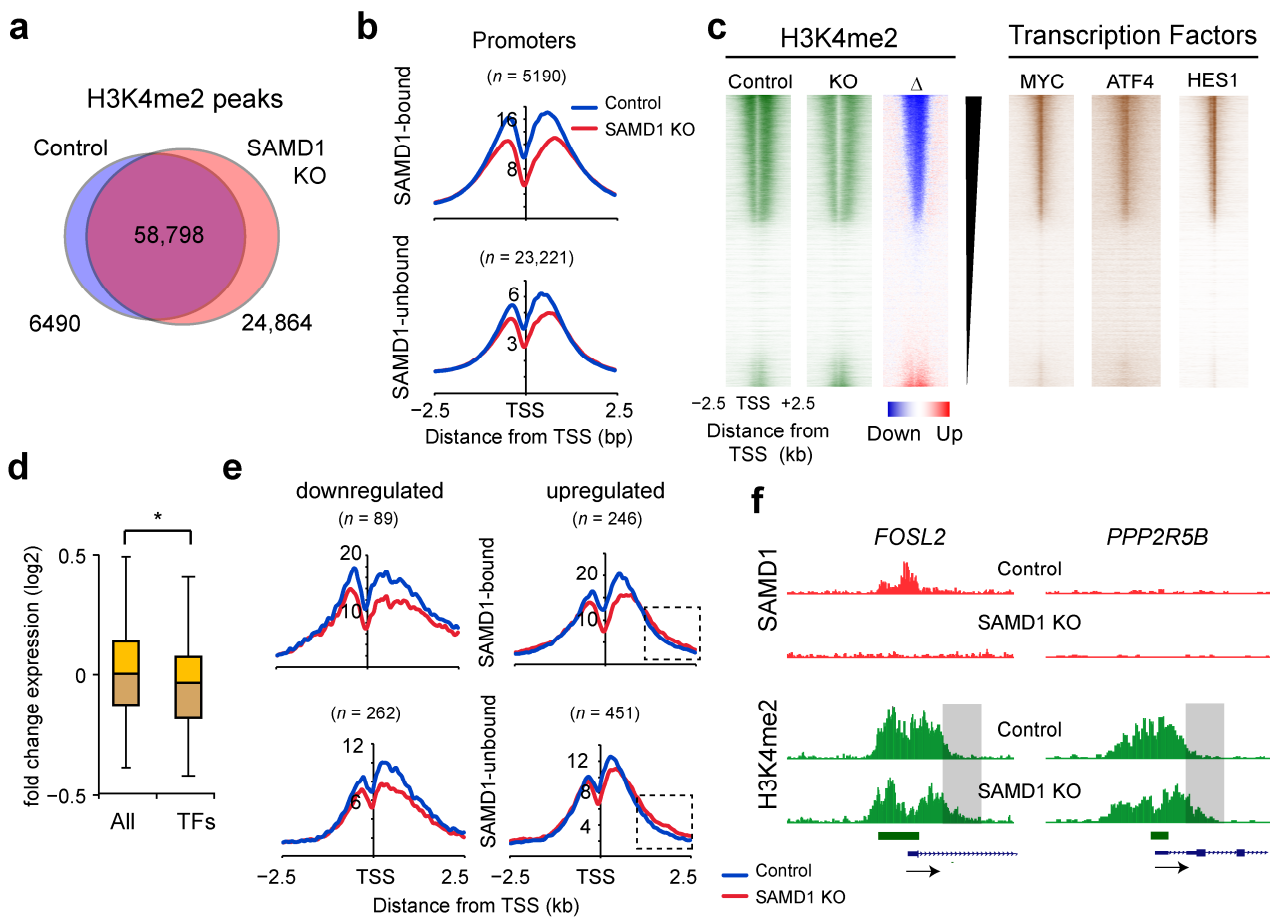
Figure 5. SAMD1 deletion leads to reduced H3K4me2 promoter levels independent of SAMD1 occupancy. (**a**) Venn diagram showing the overlap of significant H3K4me2 peaks in control and SAMD1 knockout cells. (**b**) H3K4me2 ChIP-Seq promoter profiles at SAMD1-bound and SAMD1-unbound promoters. (**c**) Heatmap showing the overlap of reduced H3K4me2 with three example transcription factors. [26] (n = 1386) in SAMD1 KO cells compared to all genes. Box plots represent upper and lower quartiles of the data with 5 and 95% whiskers. Significance was evaluated using ANOVA. * $p < 0.01$. (**e**) H3K4me2 ChIP-Seq promoter profiles of up- and downregulated genes. (**f**) Example ChIP-Seq data of upregulated genes that show increased H3K4me2 downstream of the TSS.

At the promoters, H3K4me2 levels were reduced, independent of SAMD1 binding (Figure 5b). These promoter-specific effects could be an indirect consequence of SAMD1 deletion and are probably not related to its chromatin regulatory role [9]. When classifying the promoters regarding their H3K4me2 changes, we found that the promoters with strongly reduced H3K4me2 are occupied by many transcription factors, such as MYC, HES1, and ATF4, while promoters that have increased H3K4me2 appear less targeted by transcription factors (Figure 5c). Although several transcription factor genes were repressed by SAMD1 (Figure 3h), they were, on average, downregulated in SAMD1 KO cells when all transcription factors were taken into account (Figure 5d). Among them are the genes for the abovementioned transcription factors MYC (-0.39-fold change (log2)), HES1 (-0.81), and ATF4 ($-0,46$). This suggests that SAMD1 deletion indirectly impairs many transcription factor-dependent processes, leading to reduced H3K4me2 at most promoters. Consistent with this global reduction of H3K4me2 at promoters, we also observed a reduced H3K4me2 level in the SAMD1 knockout cells by Western blotting, while other histone marks such as H3K4me3 and H3K14ac appear to be less affected (Supplementary Figures S3 and S4).

Interestingly, genes that have increased expression after SAMD1 deletion often show increased H3K4me2 levels downstream of the TSS (Figure 5e,f). This effect is independent of SAMD1 binding, as well. Thus, at upregulated genes, other chromatin regulatory mechanisms, such as chromatin remodeling processes, are possibly altered after SAMD1 deletion. This finding suggests that several independent chromatin regulatory processes at promoters are affected after SAMD1 deletion. Given the possibility that many of them are indirectly influenced by SAMD1, this cannot be easily dissected.

3.8. SAMD1 Deletion Activates Enhancers Enriched for Interferon-Stimulated Response Elements

At enhancers where SAMD1 is not binding, we also observed substantial changes. Looking at all enhancer sites (*n* = 79,253), we observed, on average, an increased level of H3K4me2 (Figure 6a), opposite to the consequences found at the promoters (Figure 5b). A large group of enhancers is H3K4me2 decorated in both cell lines (group 1, *n* = 49,133). Nevertheless, we also identified many enhancers with low H3K4me2 in the control cells but increased H3K4me2 in the SAMD1 knockout cells (group 2, *n* = 23,960) (Figure 6b). In contrast, we found many fewer enhancers that lost H3K4me2 upon SAMD1 deletion (group 3, *n* = 6160) (Figure 6b). Consistent with the known function of enhancers [53], the alteration of enhancer H3K4me2 levels also correlates with the gene expression changes of nearby genes, supporting that the regulation of the enhancer landscape plays a major role in the observed gene expression changes.

To gain insights into how SAMD1 deletion influences the enhancer landscape, we performed motif enrichment analysis (Figure 6d). At enhancers that are active in both cell lines, we found that the motif for HNF (hepatocyte nuclear factor) was most strongly enriched. This finding is not unexpected given that HNF4 is a major transcriptional regulator of liver cells [54,55]. Additionally, the interferon-stimulated response element (ISRE), as well as the motifs for PPAR and AP-1 transcription factors, were enriched. The enhancers that gain H3K4me2 upon SAMD1 deletion have a similar set of enriched motifs, but instead of the HNF motif, the ISRE motif is most strongly enriched (Figure 6d). We confirmed an increase in H3K4me2 at some of those enhancers via ChIP-qPCR (Figure 6e). The finding that these ISRE-possessing enhancers have increased H3K4me2 in SAMD1 KO cells supports our previous observation that augmented interferon signaling occurs in these cells (Figure 2e). The increased expression of IRF1 and STAT2 (Figure 3h), which bind to this motif [56], may contribute to this phenomenon.

In summary, our RNA-Seq and ChIP-Seq experiments suggest a global readjustment of the transcriptional and chromatin landscape in HepG2 cells after SAMD1 deletion, leading to an altogether more favorable gene signature (Figure 2f).
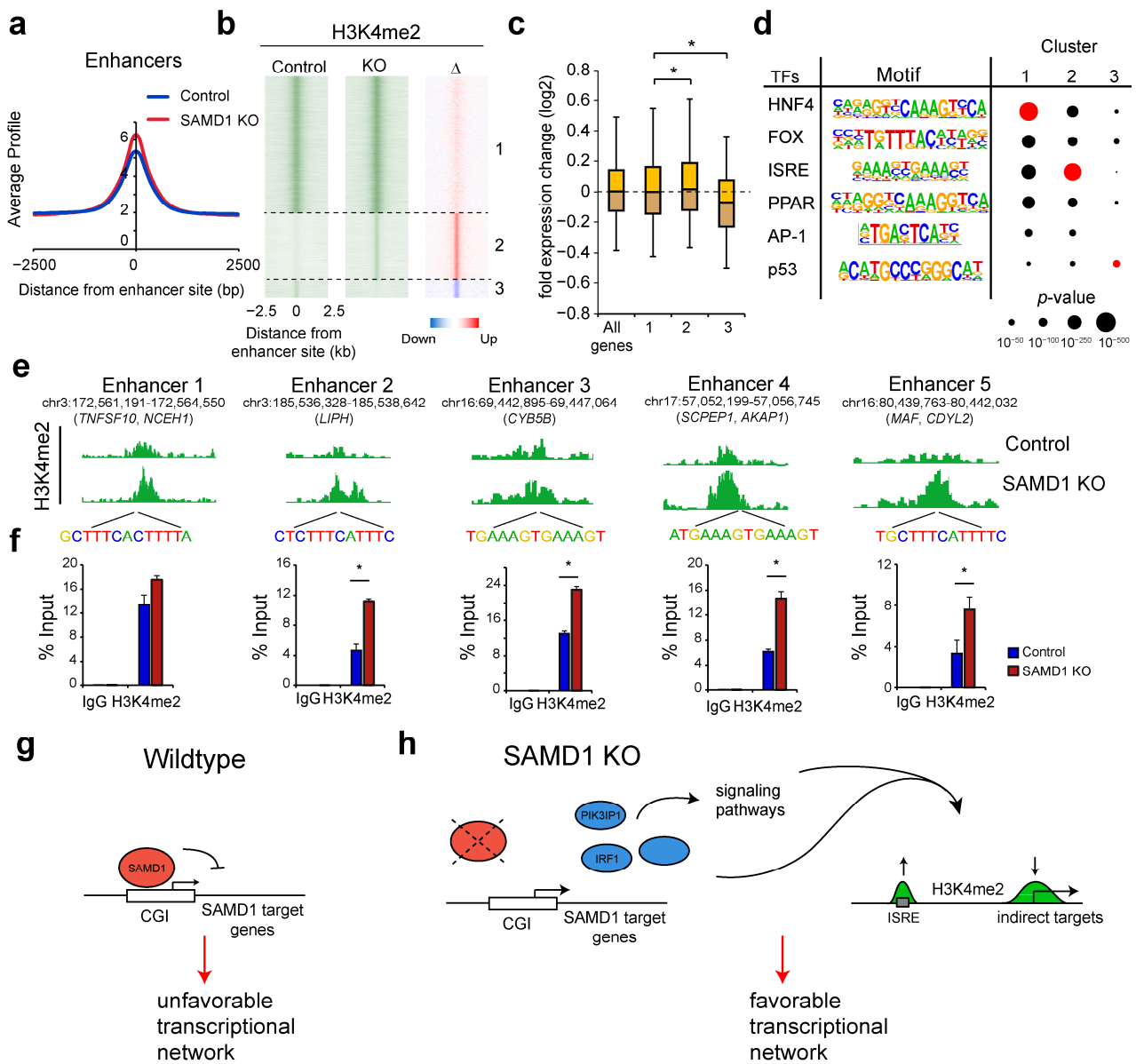
Figure 6. SAMD1 deletion leads to increased H3K4me2 levels on a subset of enhancers. (**a**) H3K4me2 ChIP-Seq profiles in control and SAMD1 KO cells at all enhancers. (**b**) Heatmap of H3K4me2 at enhancers. The difference is shown in the right plot. Three clusters were defined based on levels in control and knockout cells. (group 1: *n* = 49,133; 2: *n* = 23,960; 3: *n* = 6160) (**c**) Gene expression changes of the neighboring genes to the enhancers in each group, compared to all genes. Box plots represent upper and lower quartiles of the data with 5 and 95% whiskers. Significance was evaluated using ANOVA. * $p < 0.01$. (**d**) Motif analysis of the three clusters from B. The motif that is most strongly enriched in each cluster is marked red. (**e**) H3K4me2 ChIP-Seq data at example enhancers with increased H3K4me2 after SAMD1 deletion. (**f**) ChIP-qPCR of H3K4me2 at these enhancers in control and SAMD1 KO cells. Error bars indicate the SD of two technical replicates. * $p < 0.05$ (Student's *t*-test). (**g**) Model of SAMD1's role in hepatocellular carcinoma cells. In wild-type HepG2 cells, SAMD1 binds to unmethylated CpG islands functioning as a transcriptional repressor and contributing to an unfavorable transcriptional network. (**h**) Upon SAMD1 deletion, SAMD1 target genes such as *PIK3IP1* and *IRF1* become derepressed, which affects signaling pathways and further downstream targets. This leads to a reduced H3K4me2 level at most promoters and an increase in H3K4me2 levels at enhancers related to the interferon response. Together, this establishes a more favorable transcriptional setup.

4. Discussion

Our work demonstrated that SAMD1 is commonly upregulated in liver cancer and that its high expression is associated with poor prognosis in this cancer type (Figure 1a,b). Given the detectable expression of SAMD1 in HepG2 liver cancer cells (Figure 1f), these cells were selected as our experimental model for further investigations. The HepG2 cell line is widely used as a liver cancer model and shares common characteristics with patient liver cancers [16,17]. These cells were also included in comprehensive epigenome analyses by the ENCODE project [31], allowing us to compare our data with other genomic features. Using HepG2 cells, we addressed the question of how the upregulation of SAMD1 in liver cancer may contribute to liver cancer progression and poor prognosis.

Upon deletion of SAMD1, we observed a slightly reduced proliferation capacity and strongly impaired colony-forming ability (Figure 1j,k). This observation suggests that the absence of SAMD1 substantially impairs the biological function of these cells, particularly the ability to grow colonies out of one cell. Indeed, the analysis of the gene expression changes upon SAMD1 deletion suggests that several key processes are affected (Figure 2e). Specifically, we observed a downregulation of common cancer-related gene sets, such as MYC target genes and genes involved in an ESC-like signature. In contrast, we observed an increase in interferon response genes. Importantly, the investigation of common signatures that have been associated with good or poor prognosis [42] suggests that SAMD1 deletion alters the transcriptional network towards a signature that would be preferable for liver cancer patients (Figure 2f).

Unexpectedly, the gene expression changes that we observed upon SAMD1 deletion did not correlate well with the changes that we had observed before in mouse ES cells (Figure 2d). In addition to SAMD1 itself, the only gene that was clearly affected in both cell lines was *L3MBTL3*. Given that the L3MBTL3 protein is a direct interaction partner of SAMD1 [9], the deletion of SAMD1 may be compensated by the increased expression of L3MBTL3 via an unknown feedback mechanism. The otherwise low correlation in the gene expression changes suggests that the gene regulatory function of SAMD1 is possibly strongly dependent on the cellular context, and that the results from one cell type cannot be easily transferred to other biological systems.

The genome-wide analysis of the SAMD1 binding pattern provided us with further details on how SAMD1 regulates gene expression in HepG2 cells. Consistent with our biochemical studies indicating that the DNA-binding winged-helix domain of SAMD1 prefers unmethylated CpG motifs [9], SAMD1 exclusively binds to unmethylated CGIs in HepG2 cells. These CGIs are typically highly active, meaning they are decorated by active histone marks, and the associated genes are highly expressed (Figure 3d,e). The investigation of the SAMD1 binding patterns in mouse ES cells and human HepG2 cells further suggests that SAMD1 preferentially binds to large, unmethylated CpG islands (Supplementary Figure S2). The self-association ability of SAMD1 via its SAM domain [9] may allow the cooperative binding of SAMD1 to several unmethylated CpG motifs, which might work best at larger CGIs.

We found that SAMD1-bound genes were preferentially upregulated upon SAMD1 deletion (Figure 3g), supporting the previously observed repressive role of SAMD1 [9]. The repressive function of SAMD1 may include the function of the KDM1A histone demethylase [9], but other mechanisms, such as the function of SAMD1 interacting with the MBT domain-containing proteins L3MBTL3, SFMBT1, and SFMBT2, may also be involved [9,39,57]. Additionally, the potential SAM polymerization ability of SAMD1 could play a role in the function of SAMD1 as a transcriptional repressor at CGIs [9]. Thus, it will be of interest to further decipher the molecular mechanisms of SAMD1's repressive function in human cancer cells in future studies.

One of the major direct target genes of SAMD1 appears to be the gene *PIK3IP1*, where SAMD1 may cooperate with KDM1A to regulate gene transcription [9,58]. *PIK3IP1* functions as an inhibitor of PI3 kinases [44,47], thus acting upstream of the PI3K/Akt/mTOR pathway. This pathway is commonly dysregulated in HCC and has been a major target

for therapeutic interventions in HCC [50,52]. High *PIK3IP1* expression is associated with better prognosis [47], suggesting that *PIK3IP1* upregulation may contribute to the shift of the transcriptional landscape towards a more favorable setting. Our data suggest that impaired mTOR signaling and reduced expression of ribosome biogenesis genes are important features of HepG2 SAMD1 knockout cells.

In addition to regulating signaling pathways, SAMD1 deletion appears to strongly influence the chromatin landscape. For an initial characterization, we focused on the active H3K4me2 histone mark. Surprisingly, we observed opposite effects at promoters and enhancers. We found a decrease in H3K4me2 at promoters, while we mostly observed an increase at enhancers. In both cases, these changes appear to be largely independent of SAMD1 presence, since most changes occur at SAMD1-bound and SAMD1-unbound loci. This observation implies that most of the alterations in the chromatin landscape are due to the indirect modulations of chromatin regulatory mechanisms. Consistently, only approximately one-third of dysregulated genes are bound by SAMD1.

One of these indirect effects may involve the IRF1 protein, which is a direct target of SAMD1 and is upregulated in SAMD1 KO cells (Figure 3h). IRF1 has been described as a tumor suppressor in many cancer types and suppresses MYC-driven oncogenesis [59,60]. The precise mechanism of action of IRF1 has not yet been fully explored, but likely includes both repressive and activating functions [61,62]. IRF1 also plays an important role in the interferon response [63], which may explain the increased expression of genes related to interferon signaling (Figure 2e) and the increased level of H3K4me2 at enhancers with interferon response elements (Figure 6b,d). IRF1 has also been linked to a better prognosis in HCC [64], supporting a relevant role of IRF1 in liver cancer cells.

Another interesting candidate found to be dysregulated was the gene *CDKN2A*, encoding the tumor suppressors p14ARF and p16INK4a. The *CDKN2A* gene is often dysregulated in hepatocellular carcinomas by promoter methylation [65]. In addition to IRF1, PIK3IP1, and CDKN2A, SAMD1 likely affects many other critical factors that further influence downstream pathways (Figure 6h). To gain an understanding of the direct effects of SAMD1 deletion on the chromatin landscape, it will be necessary to track immediate changes upon SAMD1 removal.

Nonetheless, our work suggests that the absence of SAMD1 has a major impact on the transcriptional and chromatin landscape in HepG2 cells. The high expression of SAMD1 in these cells contributes to a pro-proliferative setting that is required for optimal growth. The absence of SAMD1 leads to a more favorable transcriptional signature, which is linked to better prognosis (Figure 2f) [42]. Whether this signature could also improve the response to chemotherapeutic drugs remains to be determined.

In addition to liver cancer, SAMD1 is upregulated in many other human cancer types, suggesting that SAMD1 could play a pivotal role in various cancers (Figure 1a). Notably, given conflicting gene annotations [10], the human SAMD1 gene has not been included in comprehensive CRISPR screening experiments that investigated the role of most human genes in human cancer cell lines [66]. However, in the limited number of CRISPR screens that included SAMD1, it often scores highly [13,67], suggesting a functional role of SAMD1 in multiple biological processes. In the future, it will also be of interest to address whether SAMD1 plays a role in other human cancer cell types and to analyze its involvement during tumor onset.

The work presented here has several limitations. First, all experiments were performed in one cell line: HepG2 cells. Consequently, we cannot exclude the possibility that the observed biological effects and gene expression changes are different in other liver cancer cell lines or in liver cancer tissues. Given that we observed almost no correlation between the gene expression changes in mouse ES cells and HepG2 cells (Figure 2d), it is likely that the role of SAMD1 is highly cell type-specific. Therefore, it is not possible to extrapolate our findings to other cells until further data are collected. Our characterization of the chromatin landscape is restricted to the H3K4me2 mark, thus providing only a limited picture of changes in the chromatin state. It is possible that further chromatin marks, including

repressive histone modifications, are globally altered upon SAMD1 removal, which will be of interest to be addressed in future studies. Furthermore, it is currently unclear which of the direct targets of SAMD1 are most relevant for the observed phenotypes. Additionally, we cannot exclude the possibility that some of the observed effects are due to the chromatin-independent roles of SAMD1. Further work is required to fully comprehend SAMD1's role in liver and other cancer types.

5. Conclusions

Collectively, our work demonstrates that SAMD1 modulates the transcription of unmethylated CGI-containing genes in HepG2 cells, which contributes to the establishment of an unfavorable transcriptional network (Figure 6g). This finding may explain why high SAMD1 expression is associated with worse prognosis. Interfering with SAMD1's function may be suitable to shift liver cancer cells towards a more favorable setup (Figure 6h), which could provide a new strategy for the treatment of liver cancer patients.

References

1. Llovet, J.M.; Kelley, R.K.; Villanueva, A.; Singal, A.G.; Pikarsky, E.; Roayaie, S.; Lencioni, R.; Koike, K.; Zucman-Rossi, J.; Finn, R.S. Hepatocellular carcinoma. *Nat. Rev. Dis. Primers* **2021**, *7*, 6. [CrossRef] [PubMed]
2. Chen, S.; Cao, Q.; Wen, W.; Wang, H. Targeted therapy for hepatocellular carcinoma: Challenges and opportunities. *Cancer Lett.* **2019**, *460*, 1–9. [CrossRef] [PubMed]
3. Dimri, M.; Satyanarayana, A. Molecular Signaling Pathways and Therapeutic Targets in Hepatocellular Carcinoma. *Cancers* **2020**, *12*, 491. [CrossRef] [PubMed]
4. Whittaker, S.; Marais, R.; Zhu, A.X. The role of signaling pathways in the development and treatment of hepatocellular carcinoma. *Oncogene* **2010**, *29*, 4989–5005. [CrossRef]
5. Aliya, S. Targeting Key Transcription Factors in Hepatocellular Carcinoma. *Crit. Rev. Oncog.* **2021**, *26*, 51–60. [CrossRef]
6. Zheng, K.; Cubero, F.J.; Nevzorova, Y.A. c-MYC-Making Liver Sick: Role of c-MYC in Hepatic Cell Function, Homeostasis and Disease. *Genes* **2017**, *8*, 123. [CrossRef]
7. Pelengaris, S.; Khan, M.; Evan, G. c-MYC: More than just a matter of life and death. *Nat. Rev. Cancer* **2002**, *2*, 764–776. [CrossRef]
8. Chen, H.; Liu, H.; Qing, G. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct. Target. Ther.* **2018**, *3*, 5. [CrossRef]

9.	Stielow, B.; Zhou, Y.; Cao, Y.; Simon, C.; Pogoda, H.M.; Jiang, J.; Ren, Y.; Phanor, S.K.; Rohner, I.; Nist, A.; et al. The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Sci. Adv.* **2021**, *7*, eabf2229. [CrossRef]

10.	Stielow, B.; Simon, C.; Liefke, R. Making fundamental scientific discoveries by combining information from literature, databases, and computational tools—An example. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3027–3033. [CrossRef]

11.	Deaton, A.M.; Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **2011**, *25*, 1010–1022. [CrossRef] [PubMed]

12.	Lees, A.M.; Deconinck, A.E.; Campbell, B.D.; Lees, R.S. Atherin: A newly identified, lesion-specific, LDL-binding protein in human atherosclerosis. *Atherosclerosis* **2005**, *182*, 219–230. [CrossRef] [PubMed]

13.	Norman, T.M.; Horlbeck, M.A.; Replogle, J.M.; Ge, A.Y.; Xu, A.; Jost, M.; Gilbert, L.A.; Weissman, J.S. Exploring genetic interaction manifolds constructed from rich single-cell phenotypes. *Science* **2019**, *365*, 786–793. [CrossRef] [PubMed]

14.	Yamagishi, J.; Natori, A.; Tolba, M.E.; Mongan, A.E.; Sugimoto, C.; Katayama, T.; Kawashima, S.; Makalowski, W.; Maeda, R.; Eshita, Y.; et al. Interactive transcriptome analysis of malaria patients and infecting Plasmodium falciparum. *Genome Res.* **2014**, *24*, 1433–1444. [CrossRef]

15.	Campbell, B.; Engle, S.; Ozolins, T.; Bourassa, P.; Aiello, R. Ablation of SAMD1 in Mice Causes Failure of Angiogenesis, Embryonic Lethality. *bioRxiv* **2022**. Available online: https://www.biorxiv.org/content/10.1101/2022.01.11.473462v2.full (accessed on 11 January 2022). [CrossRef]

16.	Donato, M.T.; Tolosa, L.; Gomez-Lechon, M.J. Culture and Functional Characterization of Human Hepatoma HepG2 Cells. *Methods Mol. Biol.* **2015**, *1250*, 77–93.

17.	Sefried, S.; Haring, H.U.; Weigert, C.; Eckstein, S.S. Suitability of hepatocyte cell lines HepG2, AML12 and THLE-2 for investigation of insulin signalling and hepatokine gene expression. *Open Biol.* **2018**, *8*, 180147. [CrossRef]

18.	Volkel, S.; Stielow, B.; Finkernagel, F.; Stiewe, T.; Nist, A.; Suske, G. Zinc finger independent genome-wide binding of Sp2 potentiates recruitment of histone-fold protein Nf-y distinguishing it from Sp1 and Sp3. *PLoS Genet.* **2015**, *11*, e1005102. [CrossRef]

19.	Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **2009**, *10*, R25. [CrossRef]

20.	Zhang, Y.; Liu, T.; Meyer, C.A.; Eeckhoute, J.; Johnson, D.S.; Bernstein, B.E.; Nusbaum, C.; Myers, R.M.; Brown, M.; Li, W.; et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **2008**, *9*, R137. [CrossRef]

21.	Yu, G.; Wang, L.G.; He, Q.Y. ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **2015**, *31*, 2382–2383. [CrossRef] [PubMed]

22.	Tang, Z.; Li, C.; Kang, B.; Gao, G.; Zhang, Z. GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **2017**, *45*, W98–W102. [CrossRef] [PubMed]

23.	Ramirez, F.; Dundar, F.; Diehl, S.; Gruning, B.A.; Manke, T. deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **2014**, *42*, W187–W191. [CrossRef] [PubMed]

24.	Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y.C.; Laslo, P.; Cheng, J.X.; Murre, C.; Singh, H.; Glass, C.K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **2010**, *38*, 576–589. [CrossRef] [PubMed]

25.	Kent, W.J.; Sugnet, C.W.; Furey, T.S.; Roskin, K.M.; Pringle, T.H.; Zahler, A.M.; Haussler, D. The human genome browser at UCSC. *Genome Res.* **2002**, *12*, 996–1006. [CrossRef]

26.	Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2108**, *172*, 650–665. [CrossRef]

27.	Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]

28.	Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef]

29.	Nielsen, T.O.; Parker, J.S.; Leung, S.; Voduc, D.; Ebbert, M.; Vickery, T.; Davies, S.R.; Snider, J.; Stijleman, I.J.; Reed, J.; et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **2010**, *16*, 5222–5232. [CrossRef]

30.	Subramanian, A.; Tamayo, P.; Mootha, V.K.; Mukherjee, S.; Ebert, B.L.; Gillette, M.A.; Paulovich, A.; Pomeroy, S.L.; Golub, T.R.; Lander, E.S.; et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15545–15550. [CrossRef]

31.	ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [CrossRef] [PubMed]

32.	Gertz, J.; Savic, D.; Varley, K.E.; Partridge, E.C.; Safi, A.; Jain, P.; Cooper, G.M.; Reddy, T.E.; Crawford, G.E.; Myers, R.M. Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol. Cell* **2013**, *52*, 25–36. [CrossRef] [PubMed]

33.	Tropberger, P.; Mercier, A.; Robinson, M.; Zhong, W.; Ganem, D.E.; Holdorf, M. Mapping of histone modifications in episomal HBV cccDNA uncovers an unusual chromatin organization amenable to epigenetic manipulation. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E5715–E5724. [CrossRef] [PubMed]

34.	Xiao, S.; Xie, D.; Cao, X.; Yu, P.; Xing, X.; Chen, C.C.; Musselman, M.; Xie, M.; West, F.D.; Lewin, H.A.; et al. Comparative epigenomic annotation of regulatory DNA. *Cell* **2012**, *149*, 1381–1392. [CrossRef] [PubMed]

35. Tang, Z.; Kang, B.; Li, C.; Chen, T.; Zhang, Z. GEPIA2: An enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* **2019**, *47*, W556–W560. [CrossRef]

36. Lian, Q.; Wang, S.; Zhang, G.; Wang, D.; Luo, G.; Tang, J.; Chen, L.; Gu, J. HCCDB: A Database of Hepatocellular Carcinoma Expression Atlas. *Genom. Proteom. Bioinform.* **2018**, *16*, 269–275. [CrossRef]

37. Menyhart, O.; Nagy, A.; Gyorffy, B. Determining consistent prognostic biomarkers of overall survival and vascular invasion in hepatocellular carcinoma. *R Soc. Open Sci.* **2018**, *5*, 181006. [CrossRef]

38. Chandrashekar, D.S.; Bashel, B.; Balasubramanya, S.A.H.; Creighton, C.J.; Ponce-Rodriguez, I.; Chakravarthi, B.; Varambally, S. UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses. *Neoplasia* **2017**, *19*, 649–658. [CrossRef]

39. Bonasio, R.; Lecona, E.; Reinberg, D. MBT domain proteins in development and disease. *Semin. Cell Dev. Biol.* **2010**, *21*, 221–230. [CrossRef]

40. Xu, T.; Park, S.S.; Giaimo, B.D.; Hall, D.; Ferrante, F.; Ho, D.M.; Hori, K.; Anhezini, L.; Ertl, I.; Bartkuhn, M.; et al. RBPJ/CBF1 interacts with L3MBTL3/MBT1 to promote repression of Notch signaling via histone demethylase KDM1A/LSD1. *EMBO J.* **2017**, *36*, 3232–3249. [CrossRef]

41. Wong, D.J.; Liu, H.; Ridky, T.W.; Cassarino, D.; Segal, E.; Chang, H.Y. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2008**, *2*, 333–344. [CrossRef] [PubMed]

42. Lee, J.S.; Chu, I.S.; Heo, J.; Calvisi, D.F.; Sun, Z.; Roskams, T.; Durnez, A.; Demetris, A.J.; Thorgeirsson, S.S. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* **2004**, *40*, 667–676. [CrossRef] [PubMed]

43. Zhao, R.; Choi, B.Y.; Lee, M.H.; Bode, A.M.; Dong, Z. Implications of Genetic and Epigenetic Alterations of CDKN2A (p16(INK4a)) in Cancer. *EBioMedicine* **2016**, *8*, 30–39. [CrossRef] [PubMed]

44. Zhu, Z.; He, X.; Johnson, C.; Stoops, J.; Eaker, A.E.; Stoffer, D.S.; Bell, A.; Zarnegar, R.; DeFrances, M.C. PI3K is negatively regulated by PIK3IP1, a novel p110 interacting protein. *Biochem. Biophys. Res. Commun.* **2007**, *358*, 66–72. [CrossRef]

45. Chalhoub, N.; Baker, S.J. PTEN and the PI3-kinase pathway in cancer. *Annu. Rev. Pathol.* **2009**, *4*, 127–150. [CrossRef]

46. Huang, X.; Liu, G.; Guo, J.; Su, Z. The PI3K/AKT pathway in obesity and type 2 diabetes. *Int. J. Biol. Sci.* **2018**, *14*, 1483–1496. [CrossRef]

47. He, X.; Zhu, Z.; Johnson, C.; Stoops, J.; Eaker, A.E.; Bowen, W.; DeFrances, M.C. PIK3IP1, a negative regulator of PI3K, suppresses the development of hepatocellular carcinoma. *Cancer Res.* **2008**, *68*, 5591–5598. [CrossRef]

48. Mayer, C.; Grummt, I. Ribosome biogenesis and cell growth: mTOR coordinates transcription by all three classes of nuclear RNA polymerases. *Oncogene* **2006**, *25*, 6384–6391. [CrossRef]

49. Matter, M.S.; Decaens, T.; Andersen, J.B.; Thorgeirsson, S.S. Targeting the mTOR pathway in hepatocellular carcinoma: Current state and future trends. *J. Hepatol.* **2014**, *60*, 855–865. [CrossRef]

50. Gremke, N.; Polo, P.; Dort, A.; Schneikert, J.; Elmshauser, S.; Brehm, C.; Klingmuller, U.; Schmitt, A.; Reinhardt, H.C.; Timofeev, O.; et al. mTOR-mediated cancer drug resistance suppresses autophagy and generates a druggable metabolic vulnerability. *Nat. Commun.* **2020**, *11*, 4684. [CrossRef]

51. Jung, C.H.; Ro, S.H.; Cao, J.; Otto, N.M.; Kim, D.H. mTOR regulation of autophagy. *FEBS Lett.* **2010**, *584*, 1287–1295. [CrossRef] [PubMed]

52. Ferrin, G.; Guerrero, M.; Amado, V.; Rodriguez-Peralvarez, M.; De la Mata, M. Activation of mTOR Signaling Pathway in Hepatocellular Carcinoma. *Int. J. Mol. Sci.* **2020**, *21*, 1266. [CrossRef] [PubMed]

53. Shlyueva, D.; Stampfel, G.; Stark, A. Transcriptional enhancers: From properties to genome-wide predictions. *Nat. Rev. Genet.* **2014**, *15*, 272–286. [CrossRef] [PubMed]

54. Sladek, F.M.; Zhong, W.M.; Lai, E.; Darnell, J.E., Jr. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes Dev.* **1990**, *4*, 2353–2365. [CrossRef] [PubMed]

55. Hayashi, Y.; Wang, W.; Ninomiya, T.; Nagano, H.; Ohta, K.; Itoh, H. Liver enriched transcription factors and differentiation of hepatocellular carcinoma. *Mol. Pathol.* **1999**, *52*, 19–24. [CrossRef] [PubMed]

56. Blaszczyk, K.; Nowicka, H.; Kostyrko, K.; Antonczyk, A.; Wesoly, J.; Bluyssen, H.A. The unique role of STAT2 in constitutive and IFN-induced transcription and antiviral responses. *Cytokine Growth Factor Rev.* **2016**, *29*, 71–81. [CrossRef] [PubMed]

57. Zhang, J.; Bonasio, R.; Strino, F.; Kluger, Y.; Holloway, J.K.; Modzelewski, A.J.; Cohen, P.E.; Reinberg, D. SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes Dev.* **2013**, *27*, 749–766. [CrossRef]

58. Lee, K.; Kitagawa, M.; Liao, P.J.; Virshup, D.M.; Lee, S.H. A Ras-LSD1 axis activates PI3K signaling through PIK3IP1 suppression. *Oncogenesis* **2020**, *9*, 2. [CrossRef]

59. Tanaka, N.; Ishihara, M.; Taniguchi, T. Suppression of c-myc or fosB-induced cell transformation by the transcription factor IRF-1. *Cancer Lett.* **1994**, *83*, 191–196. [CrossRef]

60. Nozawa, H.; Oda, E.; Nakao, K.; Ishihara, M.; Ueda, S.; Yokochi, T.; Ogasawara, K.; Nakatsuru, Y.; Shimizu, S.; Ohira, Y.; et al. Loss of transcription factor IRF-1 affects tumor susceptibility in mice carrying the Ha-ras transgene or nullizygosity for p53. *Genes Dev.* **1999**, *13*, 1240–1245. [CrossRef]

61. Xie, R.L.; Gupta, S.; Miele, A.; Shiffman, D.; Stein, J.L.; Stein, G.S.; van Wijnen, A.J. The tumor suppressor interferon regulatory factor 1 interferes with SP1 activation to repress the human CDK2 promoter. *J. Biol. Chem.* **2003**, *278*, 26589–26596. [CrossRef]

62. Kimura, T.; Nakayama, K.; Penninger, J.; Kitagawa, M.; Harada, H.; Matsuyama, T.; Tanaka, N.; Kamijo, R.; Vilcek, J.; Mak, T.W.; et al. Involvement of the IRF-1 transcription factor in antiviral responses to interferons. *Science* **1994**, *264*, 1921–1924. [CrossRef] [PubMed]

63. Feng, H.; Zhang, Y.B.; Gui, J.F.; Lemon, S.M.; Yamane, D. Interferon regulatory factor 1 (IRF1) and anti-pathogen innate immune responses. *PLoS Pathog* **2021**, *17*, e1009220. [CrossRef] [PubMed]

64. Yi, Y.; Wu, H.; Gao, Q.; He, H.W.; Li, Y.W.; Cai, X.Y.; Wang, J.X.; Zhou, J.; Cheng, Y.F.; Jin, J.J.; et al. Interferon regulatory factor (IRF)-1 and IRF-2 are associated with prognosis and tumor invasion in HCC. *Ann. Surg. Oncol.* **2013**, *20*, 267–276. [CrossRef] [PubMed]

65. Tannapfel, A.; Busse, C.; Weinans, L.; Benicke, M.; Katalinic, A.; Geissler, F.; Hauss, J.; Wittekind, C. INK4a-ARF alterations and p53 mutations in hepatocellular carcinomas. *Oncogene* **2001**, *20*, 7104–7109. [CrossRef] [PubMed]

66. Meyers, R.M.; Bryan, J.G.; McFarland, J.M.; Weir, B.A.; Sizemore, A.E.; Xu, H.; Dharia, N.V.; Montgomery, P.G.; Cowley, G.S.; Pantel, S.; et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **2017**, *49*, 1779–1784. [CrossRef] [PubMed]

67. Wang, T.; Yu, H.; Hughes, N.W.; Liu, B.; Kendirli, A.; Klein, K.; Chen, W.W.; Lander, E.S.; Sabatini, D.M. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **2017**, *168*, 890.e815–903.e815. [CrossRef]
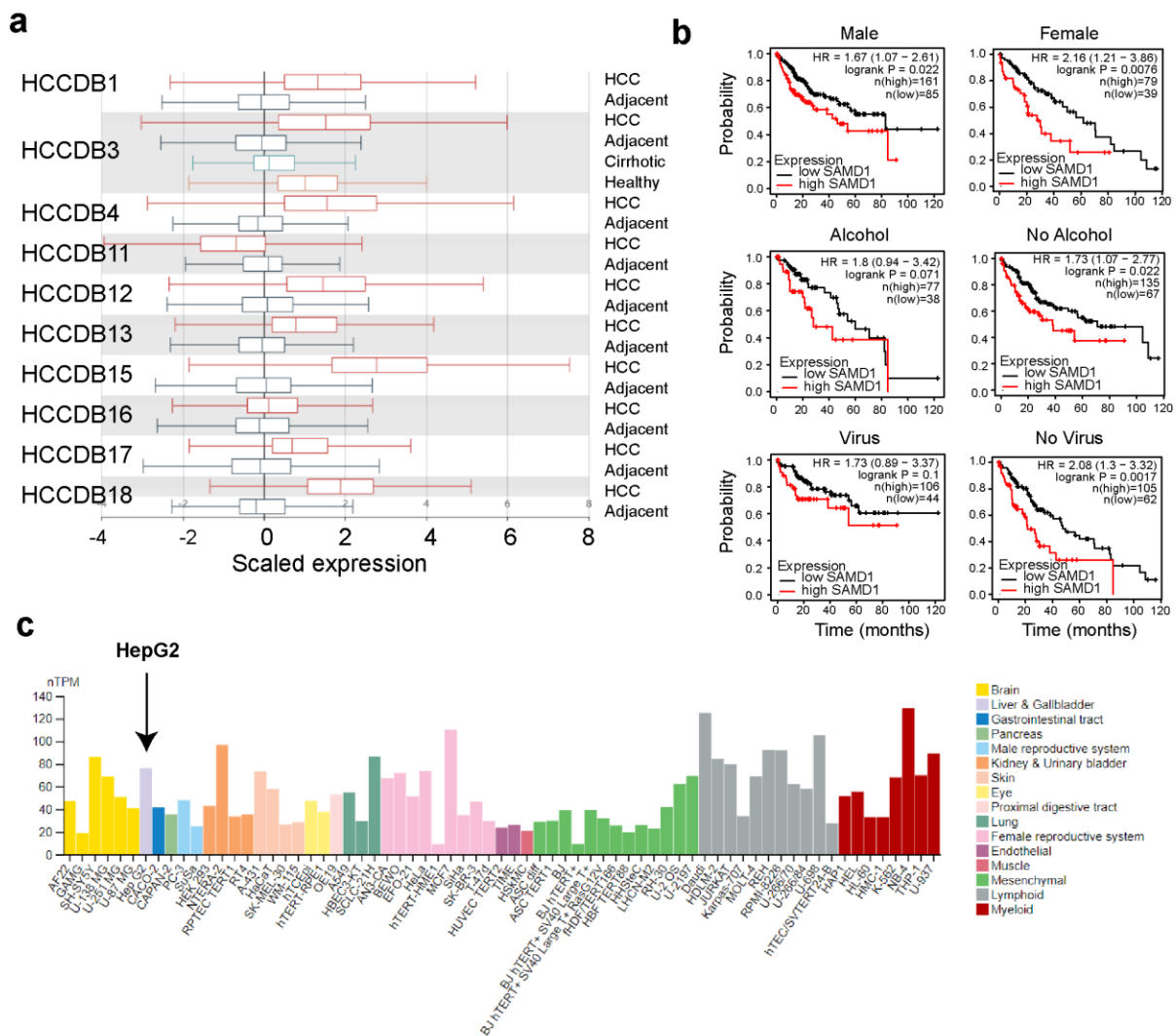
Figure S1. Further investigation of SAMD1 in liver cancer. **a** SAMD1 expression in liver cancer samples compared to healthy tissue in several independent HCC cohorts. Data derived from HCCDB[16]. **b** Patient survival from distinct cohorts with low and high SAMD1 expression. Plots were created using KM plotter[17]. **c** Expression of SAMD1 in human cancer cell lines. Date derived from Proteinatlas[19].
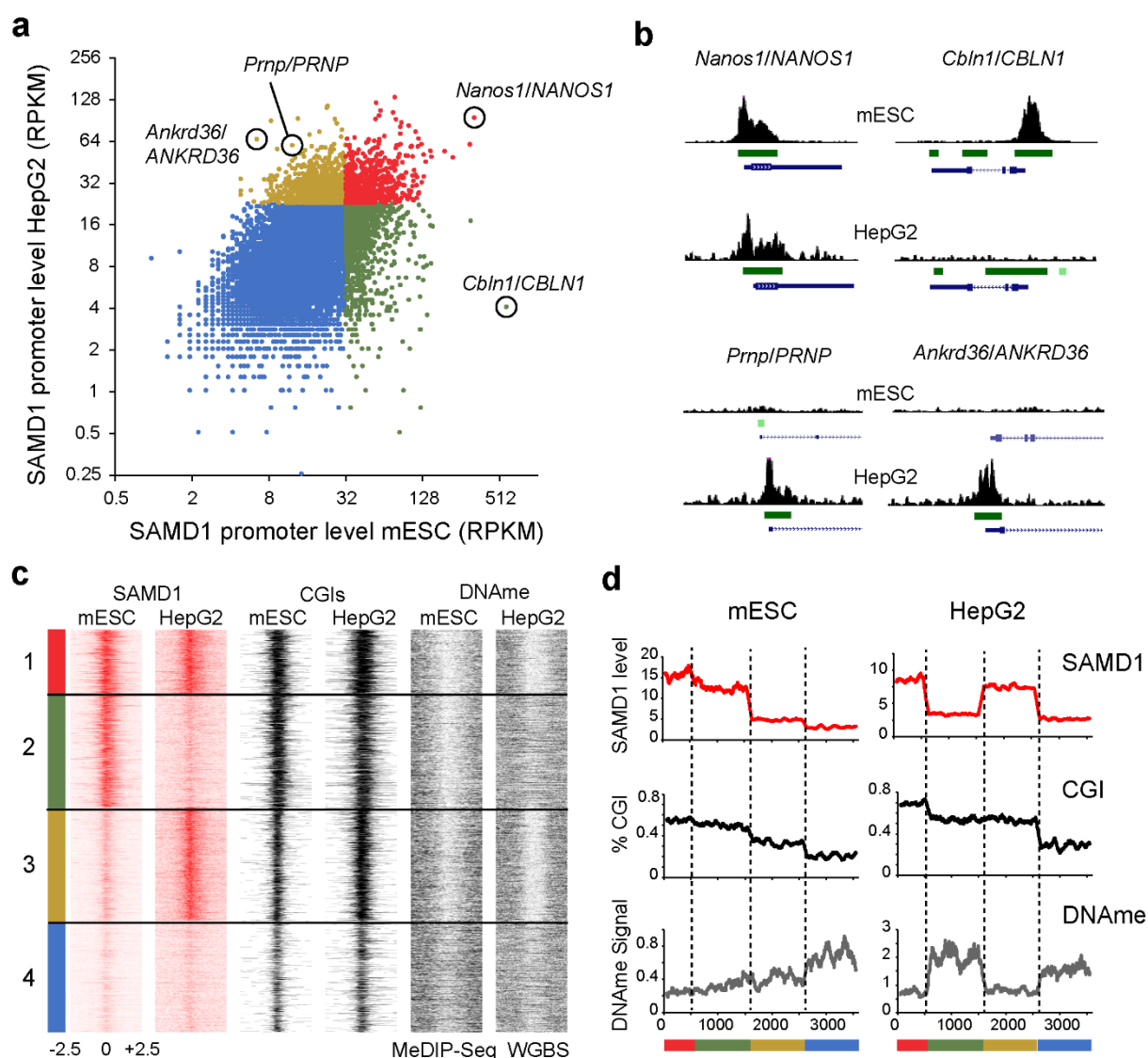
Figure S2. Comparison of SAMD1 chromatin binding in human HepG2 and mouse ES cells. **a** Correlation of SAMD1 promoter levels in mouse ES cells (mESCs) and human HepG2 cells. Only promoters that existed in both organisms were included. Four promoter groups were defined based on their SAMD1 levels in both cell lines. Group 1 (n= 586, red): bound in both cell types. Group 2 (n = 1002, green) bound only in mouse ES cells. Group 3 (n = 1016, yellow): bound only in HepG2 cells. Group 4 (n = 12,946, blue): not strongly bound in either cell type. **b** Examples of promoters that are bound by SAMD1 in both cell types, only in mouse ES cells or only in HepG2 cells. **c** Heatmap of the four groups from A in comparison with the presence of CGIs and DNA methylation. MeDIP-Seq = Methylated DNA immunoprecipitation sequencing. For group 4, 1000 randomly selected promoters were used. **d** Signal strength at each promoter group as in c, presented as a sliding window average (100 promoters).
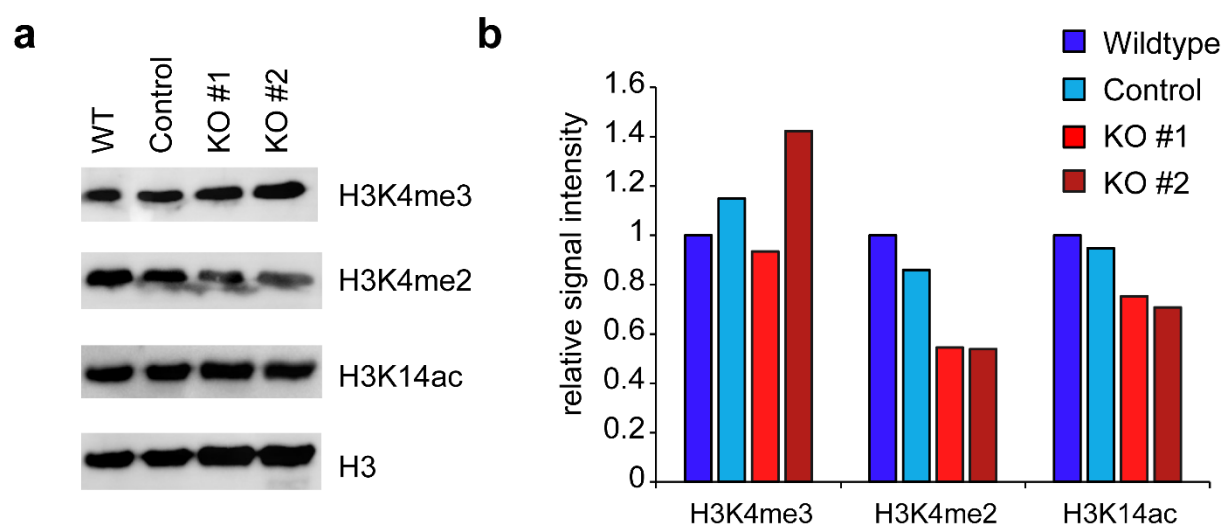
Figure S3. Global levels of histone modifications in SAMD1 KO HepG2 cells. **a** Western blotting of histone modifications in wildtype, control and SAMD1 KO cells. **b** Relative signal intensity, normalized to H3. .

SAMD1 suppresses epithelial-mesenchymal transition (EMT) pathways in pancreatic ductal adenocarcinoma

Clara Simon[1], Inka D. Brunke[1], Bastian Stielow[1], Ignasi Forné[2], Anna Mary Steitz[3], Merle Geller[1], Iris Rohner[1], Lisa Marie Weber[1], Sabrina Fischer[1], Lea Marie Jeude[1], Andrea Nist[4], Thorsten Stiewe[4], Magdalena Huber[5], Malte Buchholz[6], Robert Liefke[1,7]

Affiliations

[1] Institute of Molecular Biology and Tumor Research (IMT), Philipps University of Marburg, Marburg 35043, Germany

[2] Protein Analysis Unit, Biomedical Center (BMC), Faculty of Medicine, Ludwig-Maximilians-University (LMU) Munich, Martinsried 82152, Germany

[3] Translational Oncology Group, Center for Tumor Biology and Immunology, Philipps University of Marburg, Marburg 35043, Germany

[4] Genomics Core Facility, Institute of Molecular Oncology, Member of the German Center for Lung Research (DZL), Philipps University of Marburg, Marburg 35043, Germany

[5] Institute of Systems Immunology, Philipps University of Marburg, Marburg 35043, Germany

[6] Department of Gastroenterology, Endocrinology, Metabolism and Infection, Center for Tumor and Immunology (ZTI), Philipps University of Marburg, 35043 Marburg, Germany

[7] Department of Hematology, Oncology, and Immunology, University Hospital Giessen and Marburg, Marburg 35043, Germany

Correspondence:

Robert Liefke, robert.liefke@imt.uni-marburg.de

Abstract

SAMD1 (SAM domain-containing protein 1), a CpG island-binding protein, plays a pivotal role in the repression of its target genes. Despite its significant correlation with outcomes in various tumor types, the role of SAMD1 in cancer has remained largely unexplored. In this study we focused on pancreatic ductal adenocarcinoma (PDAC) and revealed that SAMD1 acts as a repressor of genes associated with epithelial-mesenchymal transition (EMT). Upon deletion of SAMD1 in PDAC cells, we observed significantly increased migration rates. SAMD1 exerts its effects by binding to specific genomic targets, including *CDH2*, encoding N-cadherin, which emerged as a driver of enhanced migration upon SAMD1 knockout. Furthermore, we discovered the FBXO11-containing E3 ubiquitin ligase complex as an interactor of SAMD1. FBXO11 ubiquitinates SAMD1 within its DNA-binding winged helix domain and inhibits SAMD1 chromatin binding genome-wide. High *FBXO11* expression in PDAC is associated with poor prognosis and increased expression of EMT-related genes, underlining an antagonistic relationship between SAMD1 and FBXO11. In summary, our findings provide new insights into the regulation of EMT-related genes in PDAC, shedding light on the intricate role of SAMD1 and its interplay with FBXO11 in this cancer type.

Introduction

Pancreatic cancer is a highly lethal form of cancer, accounting for 2.8% of newly diagnosed cancer cases but contributing to 4.7% of cancer-related deaths [1]. Unlike many other cancer types, the incidence and mortality rates have steadily increased in recent years [2]. The most prevalent and severe subtype of pancreatic cancer is pancreatic ductal adenocarcinoma (PDAC), with a 5-year survival rate of only 9% in the US [2].

Due to unspecific and late-occurring symptoms, PDAC is usually diagnosed in advanced stages [3]. Treatment options mainly include surgery and chemotherapy; however, most tumors are already deemed inoperable at diagnosis [4]. Epithelial-mesenchymal transition (EMT) is a crucial process in pancreatic cancer and is involved in early metastasis [5]. Genetically, PDAC is characterized by diverse mutations, with commonly affected genes including *TP53*, *CDKN2A*, *SMAD4*, and *KRAS* [6]. These genetic alterations contribute to the heterogeneity of tumors, which can vary substantially from patient to patient [6]. Recent advancements in PDAC research have focused on targeting the tumor microenvironment (TME). The TME in PDAC is characterized by its dense and desmoplastic nature, thereby influencing the druggability and chemoresistance of the tumor [7]. However, the identification of novel biomarkers for the early diagnosis of PDAC and the investigation of new druggable proteins are needed.

Sterile alpha motif domain-containing protein 1 (SAMD1) belongs to a novel class of CpG island-binding proteins alongside the histone acetyltransferases KAT6A and KAT6B [8-10]. These proteins share a winged helix (WH) domain that enables direct interaction with unmethylated CpG-rich DNA [8,10]. CpG islands (CGIs) are regulatory elements commonly found at promoter regions and play a critical regulatory role. Methylation of CGIs typically results in transcriptional silencing, while unmethylated CGIs are associated with active gene transcription [11]. CXXC domain-containing proteins and Polycomb-like proteins (PCLs) have also been identified to interact specifically with unmethylated CGIs [12-15].

In mouse embryonic stem (ES) cells, SAMD1 was found to be present at thousands of unmethylated CGIs and to recruit the chromatin regulator L3MBTL3 and the histone demethylase KDM1A to its genomic targets, thereby acting as a transcriptional repressor [8]. Deletion of SAMD1 in mouse embryonic stem cells leads to the dysregulation of multiple cellular pathways, including neuronal, developmental, and immune response pathways [8]. The absence of SAMD1 during mouse embryogenesis primarily impairs brain development and angiogenesis and leads to embryonic lethality [16]. SAMD1 function is also linked to muscle adaptation after exercise [17] and autism spectrum disorders [18].

In multiple tumor types, *SAMD1* expression is upregulated [9], and in liver cancer cells, knockout of SAMD1 has been shown to reduce proliferation and clonogenicity [19]. Moreover, liver cancer patients with high levels of SAMD1 exhibit a more unfavorable transcriptional network [19]. In the context of other cancer types, the role of SAMD1 remains largely unexplored.

Here, we show that in PDAC, SAMD1 acts as a repressor of EMT-related genes. After deleting SAMD1 in PDAC cell lines, we observed increased migration rates and upregulation of cancer-associated pathways, including the EMT pathway. We identified *CDH2* as a key downstream target of SAMD1 that is important for the migration phenotype. Furthermore, we identified the E3 ubiquitin ligase F-box only protein 11 (FBXO11) as an interactor of SAMD1 in PDAC, which inhibits the chromatin association of SAMD1. Together, our study offers novel insights into the control of EMT-related genes in PDAC, revealing the intricate involvement of SAMD1 and its interplay with FBXO11 in this cancer type.

Results

SAMD1 regulates EMT pathways in PDAC

Investigation of public cancer gene expression data from TCGA showed that *SAMD1* is commonly upregulated in cancer [20,21] (**Supplementary Figure 1a**). In some cancer types, high *SAMD1* expression correlates with poor prognoses, such as in liver cancer (LIHC) and kidney renal clear cell carcinoma (KIRC), indicating a more oncogenic role (**Supplementary Figure 1b**). In some other cancer types, such as cervical cancer (CESC) and thymoma (THYM), high *SAMD1* expression correlates with a better prognosis (**Supplementary Figure 1c**), suggesting a more tumor-suppressive role. Interestingly, the distinct relationships to patient survival do not correlate with changes in *SAMD1* gene expression upon tumorigenesis, given that *SAMD1* has increased expression in most cancer tissues compared to normal tissues (**Supplementary Figure 1a**). Thus, it is currently unknown why *SAMD1* has these potentially opposing roles in different cancer types.

A vital cancer type in which high *SAMD1* expression correlates with a better prognosis is pancreatic ductal adenocarcinoma (PDAC) (**Figure 1a**). We hypothesized that gaining a deeper understanding of the potential tumor-suppressive role of SAMD1 in PDAC may allow us to employ this function to limit cancer growth. To acquire insights into the role of SAMD1 in PDAC, we investigated TCGA data using gene set enrichment analysis (GSEA) [22] and compared PDAC samples with high and low *SAMD1* expression. In the samples with high *SAMD1* expression, we found a strongly decreased expression of genes related to epithelial-mesenchymal transition (EMT) (**Figure 1b**). Similar results were also obtained for thymoma

and cervical cancer (**Supplementary Figure 1d**), where high *SAMD1* expression also correlates with a better prognosis (**Supplementary Figure 1b**).

In contrast, the opposite is the case in cancer types where high *SAMD1* correlates with a worse prognosis, such as kidney cancer. Here, the expression of EMT pathway genes positively correlates with *SAMD1* expression (**Supplementary Figure 1b, d**). Notably, in all investigated cancer types, high *SAMD1* expression correlates with high expression of MYC target genes (**Supplementary Figure 1e**), independent of whether high *SAMD1* expression is favorable or unfavorable, suggesting that this feature is not predictive. Together, these observations lead to the hypothesis that in some cancer types, such as PDAC, SAMD1 may be involved in repressing EMT pathways, thereby inhibiting metastasis, which in turn may contribute to a better outcome. In PDAC, EMT is particularly relevant because it strongly correlates with the occurrence of metastasis, which massively reduces the chance of survival [23].
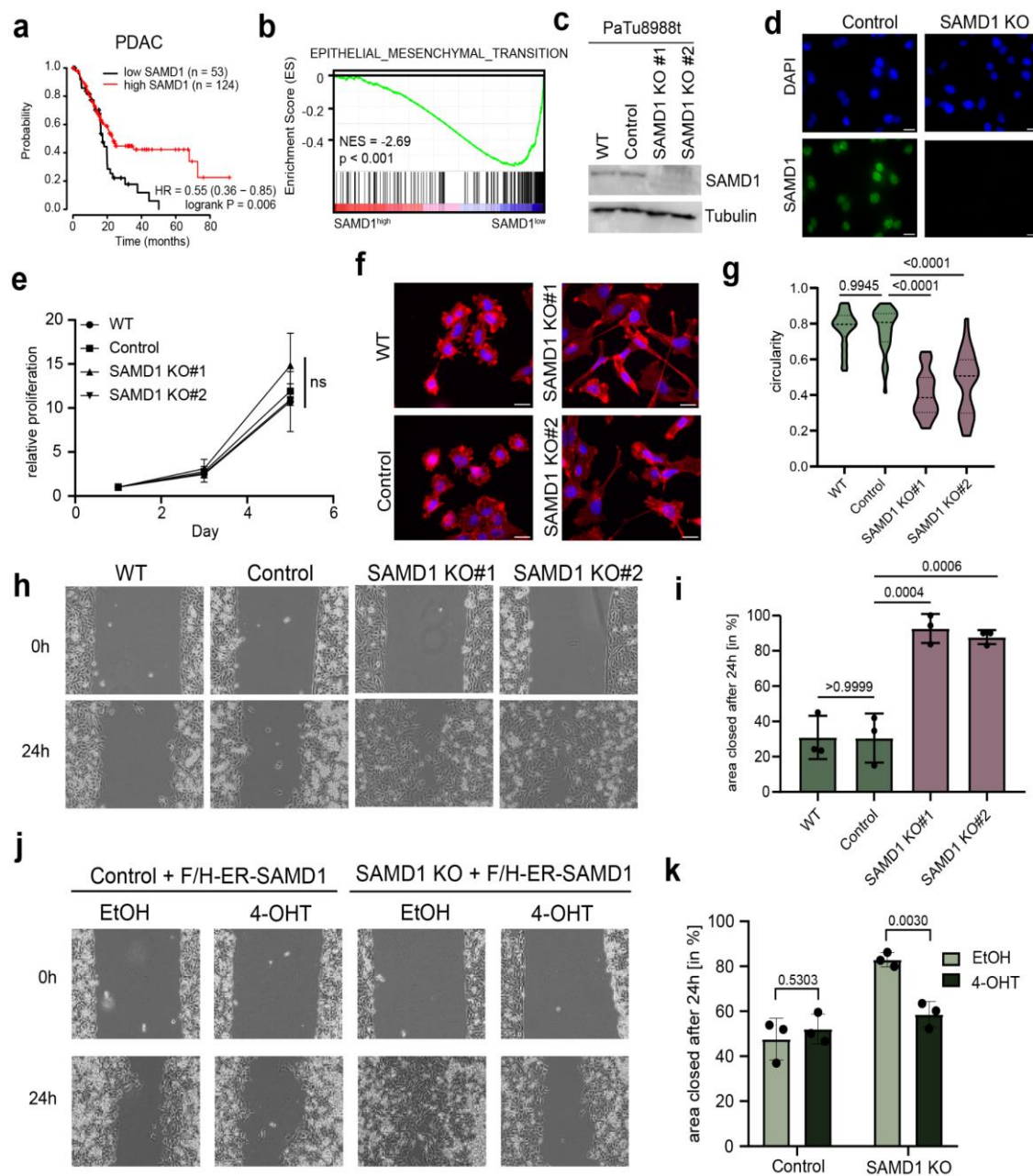
Figure 1: SAMD1 inhibits EMT pathways and cell migration.

a) Kaplan Meier survival curves showing the correlation of *SAMD1* expression with patient survival. Graphs were visualized via KM-Plotter [24].

b) GSEA for epithelial-mesenchymal transition using TCGA data analyzed for high and low *SAMD1* expression.

c) Western blot showing PaTu8988t wild-type cells, control cells, and two different SAMD1 knockout clones.

d) Immunofluorescence of PaTu8988t wild-type, and SAMD1 knockout cells, Bar=20 μM.

e) Proliferation assay of PaTu8988t wild-type, control cells and two different SAMD1 knockout clones, Data represent the mean ± SD of three biological replicates. Significance was analyzed using one-way ANOVA.

f) Representative phalloidin staining of PaTu8988t wild-type cells, control cells and two different SAMD1 knockout clones, Bar=20 μM.

g) Cell shape of PaTu8988t wild-type cells, control cells, and two different SAMD1 knockout clones. Circularity was determined using ImageJ Fiji. Significance was analyzed using one-way ANOVA.

h) Representative picture of one wound healing assay of PaTu8988t wild-type cells, control cells, and two different SAMD1 knockout clones.

i) Quantification of the wound healing assay from h). Data represent the mean ± SD of three biological replicates. Significance was analyzed using one-way ANOVA.

j) Representative picture of one wound healing assay of PaTu8988t control or SAMD1 KO cells with or without induction of SAMD1 rescue.

k) Quantification of the wound healing assay from m). Data represent the mean ± SD of three biological replicates. Significance was analyzed using Student's t-test.

To address the role of SAMD1 in the PDAC cells in more detail, we performed CRISPR/Cas9-mediated knockout approaches in the PDAC cell line PaTu8988t, validated via Western and immunofluorescence (**Figure 1c, d**). We did not observe a change in the proliferation rate upon SAMD1 deletion (**Figure 1e**). Still, investigation of the cells under the microscope showed that SAMD1 knockout led to a more elongated cell shape and more pronounced protrusions (**Figure 1f, g**). This phenotype suggested an increased mobility of the knockout cells. By analyzing the migration rates of PaTu8988t cells by wound healing (**Figure 1h, i**), we confirmed the higher mobility of the SAMD1 KO cells. Furthermore, transwell migration through 8 μm pores demonstrated increased migration rates after SAMD1 deletion in PaTu8988t cells (**Supplementary Figure 2a**), which could be visualized via crystal violet staining (**Supplementary Figure 2b**). We confirmed these results via an unbiased approach by tracking PaTu8988t cells for 24 h using time-lapse analysis (**Supplementary Figure 2c**, **Supplementary Video 1 and 2**).

Enhanced cellular migration, but no chance in proliferation, could also be found in BxPC3 PDAC cancer cells upon SAMD1 deletion (**Supplementary Figure 2d-g**). These results suggest that the influence on the cellular properties by SAMD1 is not restricted to a single cell line but is a more general theme in PDAC. This is further supported by the anticorrelation of *SAMD1* expression and EMT pathways in PDAC patients (**Figure 1b**).

To investigate whether the observed phenotype depends on the nuclear function of SAMD1, we made use of an estrogen-receptor (ER)-SAMD1 fusion protein, whose nuclear localization can be induced by 4-hydroxy-tamoxifen (4-OHT) (**Supplementary Figure 3a, b**). Using this approach, we demonstrated that the migration phenotype in PaTu8998t SAMD1 KO cells can be rescued upon translocation of SAMD1 into the nucleus (**Figure 1j, k**), suggesting that the observed phenotype is linked to the chromatin regulatory role of SAMD1.

SAMD1 directly represses CDH2, a key regulator of EMT.

To address which SAMD1 target genes participate in this phenotype, we performed gene expression analysis via RNA-seq upon SAMD1 knockout and analyzed the genomic distribution of SAMD1 via ChIP-seq in PaTu8998t cells. Principal component analysis (PCA) of the RNA-seq data demonstrated that the knockout led to a substantial shift in the transcriptional landscape (**Supplementary Figure 4a**). The knockout of SAMD1 led to significant deregulation of 854 genes, with 642 upregulated and 212 downregulated genes (cut-off: log2-fold-change > 0.5; p-value < 0.01) (**Figure 2a, Supplementary Figure 4b**). GSEA analysis of the RNA-seq data demonstrated that the deletion of SAMD1 leads to the dysregulation of multiple cancer-related pathways (**Figure 2b**). Specifically, we observed an upregulation of signaling pathways, including Hedgehog, KRAS, and WNT, and a downregulation of MYC and E2F target genes. Additionally, many transcription factors, including HOXB cluster genes, become dysregulated upon SAMD1 deletion (**Supplementary Figure 4b**). The EMT pathway genes were upregulated in the knockout cells as well (**Figure 2c**), consistent with the observed phenotype and in line with our initial hypothesis that SAMD1 may be involved in regulating this pathway.

Genome-wide analysis of SAMD1 chromatin binding showed that SAMD1 mainly binds to CpG island-containing gene promoters (**Figure 2d**), predominantly linked to chromatin and transcriptional regulation (**Supplementary Figure 4c**). The top SAMD1 target genes were significantly upregulated upon SAMD1 deletion, as assessed by GSEA (**Figure 2e**). These results are consistent with our previous findings from mouse ES [8] and HepG2 cells [19] and support that SAMD1 acts as a repressor at CGIs. Consequently, we hypothesized that the increased migratory ability of the knockout cells may be established by the derepression of one or several SAMD1 target genes.

One of the top upregulated EMT genes in the SAMD1 KO cells was *CDH2* (**Figure 2a, f**), whose promoter also showed high SAMD1 enrichment (**Figure 2g**). *CDH2* encodes for N-cadherin, a crucial regulator of cell adhesion and consequently for EMT in PDAC [25,26]. We also confirmed the upregulation of N-cadherin via immunofluorescence (**Figure 2h**). Therefore, *CDH2* could be a key downstream target of SAMD1 in PaTu8988t cells to influence cellular migration. Indeed, the cells possessing a double knockout of CDH2 and SAMD1 had similar characteristics to the wild-type cells regarding their migratory ability (**Figure 2i, j**) and cellular shape (**Supplementary Figure 5a, b**). Additionally, short-term inhibition of N-cadherin-mediated cell adhesion by ADH-1 (Exherin) [27] reduced the elongated phenotype of SAMD1 KO cells, making them more similar to wild-type cells (**Supplementary Figure 5c, d**). These findings support that *CDH2* is a critical downstream factor of SAMD1 that regulates the migration properties of the PaTu8988t cells.

To assess whether SAMD1 directly regulates *CDH2*, we used the ER-SAMD1 fusion protein described above. In SAMD1 KO cells expressing this fusion protein, the expression of *CDH2* was rescued upon 4-OHT treatment (**Figure 2k**). This rescue does not work with a winged helix domain mutant of SAMD1, indicating that the chromatin binding of SAMD1 is essential for the repression of *CDH2*. Furthermore, ChIP-qPCR confirmed that SAMD1 chromatin binding to the *CDH2* promoter can be rescued with the ER-SAMD1 fusion protein (**Figure 2l**). Similar results were also obtained for the *L3MBTL3* gene, a known SAMD1 target gene that becomes consistently upregulated upon SAMD1 deletion in several distinct cell types [8,19] (**Supplementary Figure 3c, d**). The regulatory effect of SAMD1 on *CDH2* and *L3MBTL3* was also be confirmed in BxPC3 cells (**Supplementary Figure 2h, i**).

Together, these results suggest that SAMD1 is directly involved in repressing *CDH2* in PDAC cells, an essential regulator of EMT [26].
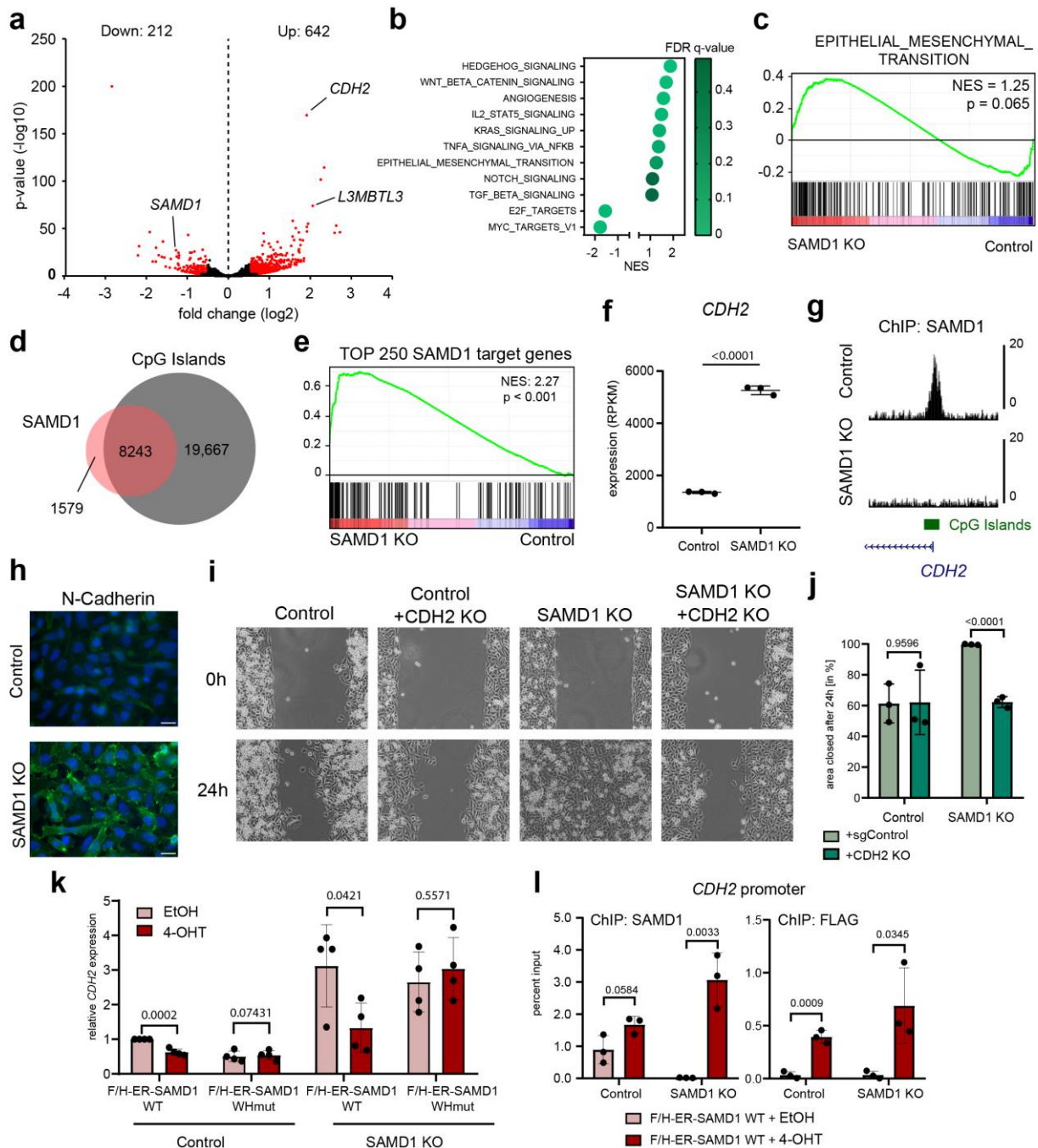
Figure 2: SAMD1 directly regulates EMT pathway genes in PaTu8988t cells.

a) Volcano plot of RNA-seq data comparing the results from three replicates of PaTu8988t control cells with three clonally independent SAMD1 KO cells.

b) GSEA for several pathways comparing the results from three replicates of PaTu8988t control cells with three clonally independent SAMD1 KO cells.

c) GSEA of epithelial-mesenchymal transition from b).

d) Venn diagram showing the overlap of SAMD1 peaks with all CpG islands in PaTu8988t cells.

e) GSEA of the top 250 SAMD1 targets in PaTu8988t cells, comparing the results from three replicates of control cells with three clonally independent SAMD1 KO cells.

f) RNA-seq results for *CDH2* expression (RPKM) comparing the results from three replicates of PaTu8988t control cells with three clonally independent SAMD1 KO cells.

g) Snapshot of the USCS browser showing a SAMD1 peak at the *CDH2* promoter in PaTu8988t control and SAMD1 KO cells.

h) Immunofluorescence of N-cadherin in PaTu8988t control and SAMD1 KO cells, Bar=20 μM.

i) Representative picture of one wound healing assay of PaTu8988t control, CDH2 KO, SAMD1 KO and CDH2/SAMD1 double KO cells.

j) Quantification of the wound healing assay from i). Data represent the mean ± SD of three biological replicates. Significance was analyzed using Student's t-test.

k) RT-qPCR showing *CDH2* expression with or without induction of SAMD1 rescue in PaTu8988t Control and SAMD1 KO cells. WHmut=RK-45/46-AA mutation of *SAMD1*. Data represent the mean ± SD of four biological replicates. Significance was analyzed using Student's t-test.

l) SAMD1 ChIP-qPCR at the *CDH2* promoter with or without induction of SAMD1 rescue in PaTu8988t Control and SAMD1 KO cells. Data represent the mean ± SD of three biological replicates. Significance was analyzed using Student's t-test.

The repressive activity of SAMD1 likely involves KDM1A.

The molecular details of the repressive function of SAMD1 are currently not fully understood. Previously, we showed that SAMD1 interacts with the KDM1A complex, which demethylates the active H3K4me2 histone mark, and the SAM- and MBT-domain proteins L3MBTL3 and SFMBT1 [8]. In the context of PaTu8988t cells, we confirmed that in the absence of SAMD1, the levels of KDM1A and L3MBTL3 are reduced at the *CDH2* gene, which can be rescued upon induction of the ER-SAMD1 with 4-OHT (**Figure 3a**). Consistently, via ChIP-qPCR, we observed an increase in H3K4me2 and H3K4me3 at the *L3MBTL3* and *CDH2* gene promoters upon SAMD1 deletion (**Figure 3b**). We also validated the interaction between SAMD1 and KDM1A in PaTu8988t cells via endogenous co-immunoprecipitation (**Figure 3c**). These results support that in PaTu8988t cells, KDM1A is likely involved in the repressive function of SAMD1.

To gain further insight into the interplay of SAMD1 with the KDM1A complex (**Figure 3d**), we performed mapping experiments that went beyond our previous experiments [8]. First, we confirmed that SAMD1 preferentially interacts with KDM1A via its winged helix domain and that deleting the SAM domain, which is essential for the interaction with L3MBTL3 [8], increases the interaction with KDM1A (**Figure 3e**). This phenomenon is also observable with other members of the KDM1A complex (**Supplementary Figure 6a**). Reverse mapping experiments suggested that SAMD1 preferentially interacts with the N-terminal part of the catalytic amino oxidase domain (AOD) of KDM1A (**Figure 3f**). This result is further supported by the observation that the KDM1A inhibitor ORY-1001 (Iadademstat), which covalently binds to the FAD cofactor within KDM1A [28], interferes with the interaction of KDM1A with SAMD1 (**Supplementary Figure 6b**). This effect was not observed for other KDM1A complex members, such as RCOR1 and PHF21A (**Supplementary Figure 6c, d**). This finding raises the possibility that SAMD1 may bind near the catalytic cleft of KDM1A, potentially causing an increased sensitivity of the interaction to the inhibitor treatment

Based on these results, we speculated that SAMD1 could potentially influence the demethylase activity of KDM1A, similar to other factors associated with the KDM1A complex [29]. To address this question, we immunoprecipitated KDM1A in either wild-type or SAMD1 KO HEK293 cells and used the obtained precipitate for demethylase assays, adding calf histones as a substrate. We found that the absence of SAMD1 reduced the ability of KDM1A to remove H3K4me2 efficiently (**Figure 3g, h**). This result suggests that SAMD1 modulates the function of KDM1A, possibly not just by influencing its recruitment to chromatin but also by influencing the catalytic activity of KDM1A. Both processes together may contribute to the repressive role of SAMD1. However, the molecular details of how SAMD1 affects the

enzymatic activity of KDM1A require further research. In addition, we cannot exclude the possibility that other mechanisms, such as the recruitment of L3MBTL3, are also crucial for the repressive function of SAMD1.
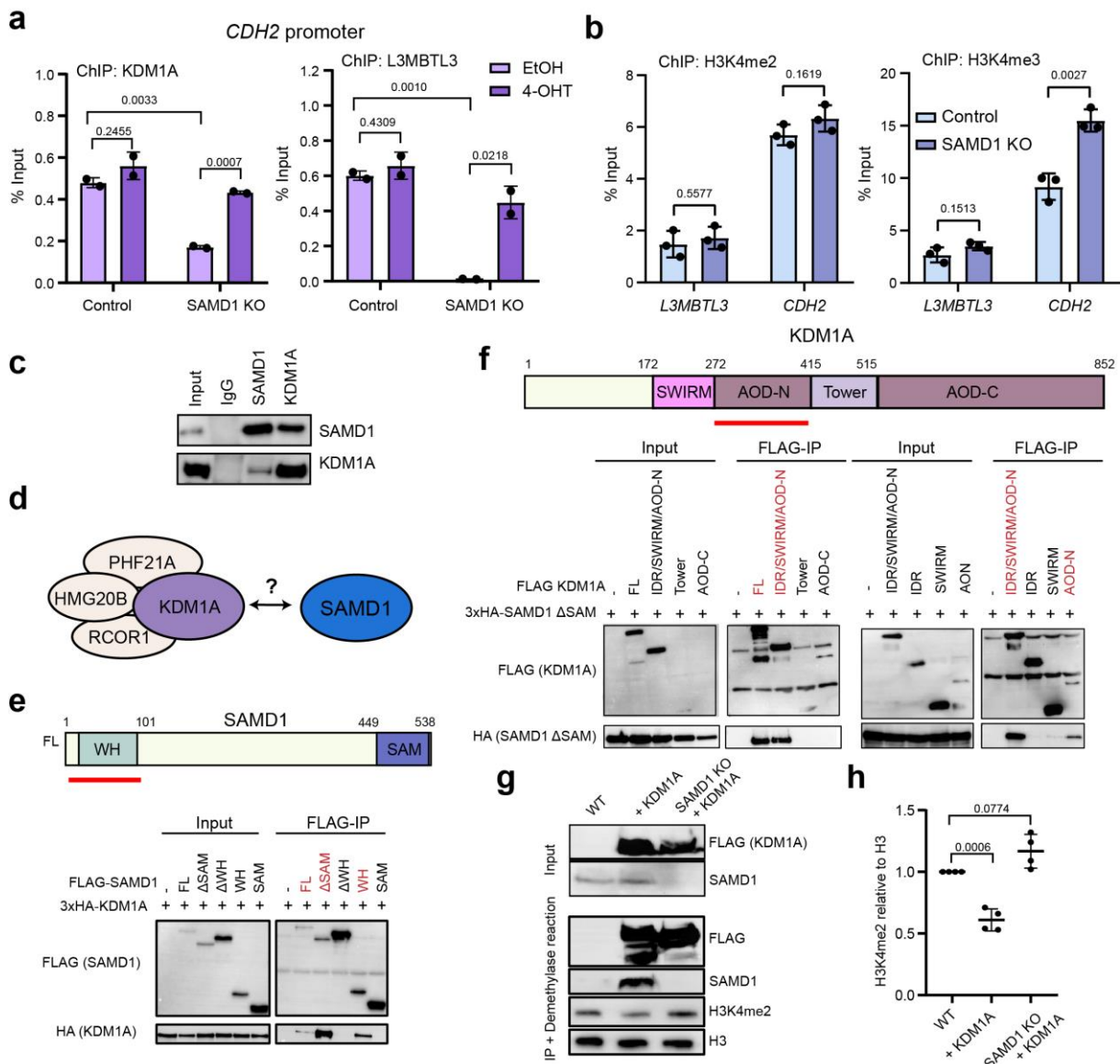

Figure 3: SAMD1 is required for full activity of KDM1A.

a) ChIP-qPCR at the *CDH2* promoter with or without induction of SAMD1 rescue in PaTu8988t Control and SAMD1 KO cells using KDM1A and L3MBTL3 antibodies. Data represent the mean ± SD of two biological replicates. Significance was analyzed using Student's t-test.

b) ChIP-qPCR of *CDH2* and *L3MBTL3* promoter in PaTu8988t Control and SAMD1 KO cells using H3K4me2 and H3K4me3 antibodies. Data represent the mean ± SD of three biological replicates. Significance was analyzed using Student's t-test.

c) Western blot of an endogenous Co-IP between SAMD1 and KDM1A in PaTu8988t cells.

d) Model of the interaction between SAMD1 and the KDM1A complex.

e) Structure of SAMD1; co-immunoprecipitation in HEK293 cells showing the interaction between different SAMD1 deletion mutants and KDM1A. Regions identified to interact with KDM1A are marked red.

f) Structure of KDM1A; co-immunoprecipitation in HEK293 cells showing the interaction between different KDM1A deletion mutants and SAMD1. Regions identified to interact with SAMD1 are marked red.

g) Representative western blot of KDM1A IP in HEK293 cells, followed by histone demethylase assay.

h) Quantification of four biological replicates of g). Significance was analyzed using one-way ANOVA.

SAMD1 interacts with the FBXO11 E3 ubiquitin ligase complex

Upon investigating the cellular localization of SAMD1 in PaTu8988t cells and further PDAC cell lines, we observed that compared to other human cell lines, SAMD1 is less present in the chromatin fraction in PDAC cells (**Figure 4a**). This finding raises the possibility that a certain molecular mechanism regulates the chromatin binding of SAMD1. In the context of pancreatic cancer cells, such a mechanism may be essential to overcome the tumor-suppressive function of SAMD1. To date, no process has been described that regulates the chromatin association of SAMD1.

To address whether SAMD1 may interact with additional proteins that could be involved in such a regulatory process, we performed unbiased IP-MS experiments in PaTu8988t cells. For this, we used cells expressing the ER-SAMD1 fusion protein (**Supplementary Figure 3a, b**). After inducing the nuclear localization of the protein via 4-OHT, we collected the cells and immunoprecipitated the SAMD1 protein. The cobound proteins were analyzed by LC-MS (**Figure 4b**). This experiment confirmed that SAMD1 interacts with L3MBTL3 and the KDM1A histone demethylase complex. We also identified L3MBTL4, consistent with our finding that the SAM domain of L3MBTL4 can interact with the SAM domain of SAMD1, similar to L3MBTL3 [8]. In addition to these expected interactions, we identified members of the FBXO11 complex as putative novel interaction partners of SAMD1. The FBXO11 complex consists of FBXO11 itself, RBX1, Cullin 1, and SKP1, all of which are enriched in the SAMD1 IP (**Figure 4b, c**). Additionally, we found enrichment of NEDD8, which is typically covalently associated with Cullin 1 and is required for the F-box protein-associated E3 ubiquitin ligase complexes to be active [30].

Via co-immunoprecipitation experiments in HEK293 cells, we validated that SAMD1 can interact with FBXO11 (**Figure 4d**). Additional mapping experiments suggested that several regions of SAMD1 are relevant for this interaction (**Figure 4e**). Only the SAM domain appears dispensable for the interaction with FBXO11 (**Figure 4e**). Interestingly, we found that FBXO11 can also be co-immunoprecipitated with KDM1A (**Figure 4f**). This interaction is facilitated by the N-terminal part of the AOD domain of KDM1A, which is the same region as for the interaction with SAMD1 (**Figure 4f, 3f**). Given that the interaction between KDM1A and FBXO11 is also sensitive to the KDM1A inhibitor ORY-1001 (**Supplementary Figure 6e**), similar to the SAMD1/KDM1A interaction (**Supplementary Figure 6b**), it suggests that KDM1A interacts with SAMD1 and FBXO11 simultaneously.

FBXO11 is an E3 ubiquitin ligase that regulates the ubiquitination of various target proteins, including BCL6, CDT2, and BAHD1 [31-33]. We hypothesized that FBXO11 may facilitate the ubiquitination of SAMD1. To investigate whether SAMD1 is ubiquitinated, we co-immunoprecipitated FLAG-tagged SAMD1 in HEK293 cells expressing HA-tagged ubiquitin. We observed that in the presence of the proteasome inhibitor MG-132, the immunoprecipitated FLAG-SAMD1 but not the control precipitate showed an HA-signal in Western blotting experiments (**Figure 4g**), demonstrating that SAMD1 becomes

ubiquitinated when the proteasome is inhibited. Surprisingly, however, cycloheximide chase experiments suggested that SAMD1 is highly stable, with no obvious turn-over within 6 hours (**Figure 4h**). To assess which region of SAMD1 is mostly ubiquitinated, we used SAMD1 deletion mutants. We found that deleting the WH domain decreased the ubiquitination level, while deletion of the SAM domain, or using the WH domain alone led to an increased ubiquitination level (**Figure 4i**). This result suggests that the WH domain is the primary ubiquitination site of SAMD1.
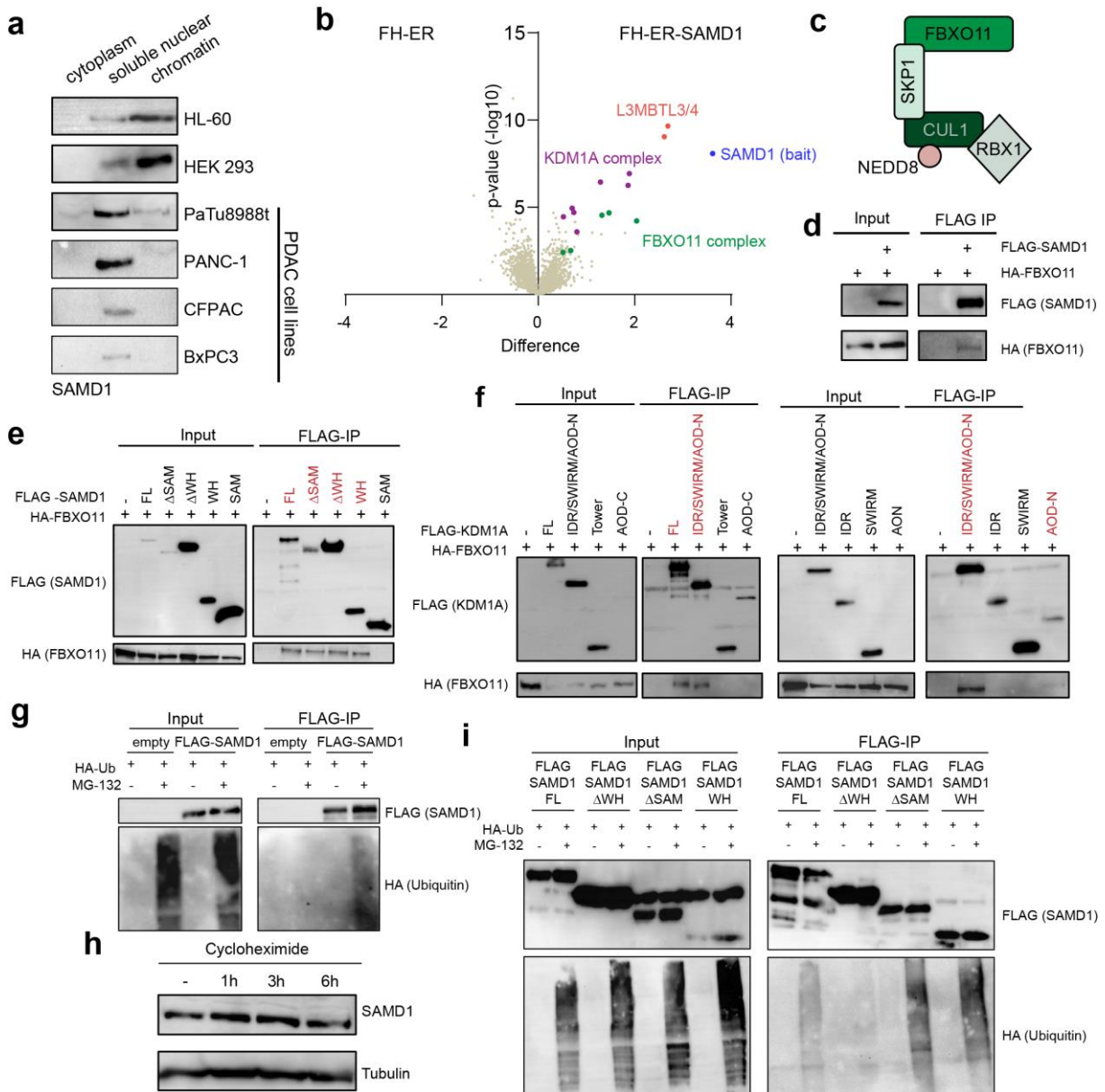

Figure 4: SAMD1 interacts with FBXO11 in PaTu8988t cells and is ubiquitinated.

a) Fractionation of different cell lines followed by SAMD1 western blotting.
b) Volcano plot of proteins identified by mass spectrometry after IP of FH-ER and FH-ER-SAMD1.
c) Model of the FBXO11 E3-ubiquitin ligase complex.
d) Co-immunoprecipitation in HEK293 cells showing the interaction between SAMD1 and FBXO11.
e) Co-immunoprecipitation in HEK293 cells showing the interaction between different SAMD1 deletion mutants and FBXO11.

f) Co-immunoprecipitation in HEK293 cells showing the interaction between different KDM1A deletion mutants and FBXO11.
g) Ubiquitination assay after empty vector or SAMD1 transfection in HEK293 cells.
h) Cycloheximide chase analysis of SAMD1 in PaTu8988t cells.
i) Ubiquitination assay in HEK293 cells using different SAMD1 deletion constructs.

Ubiquitination by FBXO11 possibly directly influences the DNA binding of SAMD1

To investigate the role of FBXO11 in the ubiquitination of SAMD1, we established HEK293 and PaTu8988t cells with FBXO11 knockout (**Figure 5a**). Consistent with the idea that FBXO11 ubiquitinates SAMD1, we found a reduced level of ubiquitination of the SAMD1 WH domain in the FBXO11 knockout HEK293 cells (**Figure 5b, c**).

Notably, in the FBXO11 knockout PaTu8988t cells, we did not observe an altered protein level of SAMD1 (**Figure 5d**), suggesting that ubiquitination of SAMD1 by FBXO11 has no relevant influence on the turn-over of SAMD1. However, fractionation experiments showed that the chromatin association of SAMD1 was substantially augmented in the FBXO11 KO cells (**Figure 5e**), supporting that the FBXO11 is involved in modulating the chromatin association of SAMD1.

We speculated that the inhibitory effect of FBXO11 on SAMD1 chromatin binding is due to direct interference with SAMD1's DNA binding ability. Based on the PhosphoSitePlus database [34], the amino acids K75 and K88, which lie in the DNA-binding WH domain, can be ubiquitinated. K88 is directly involved in DNA binding via its ability to interact with the minor groove of the DNA [8], while K75 is not associated with DNA (**Figure 5f**). We mutated these amino acids into alanines to address whether they are relevant for SAMD1 ubiquitination. Indeed, we found an approximately 30% percent reduction in ubiquitination levels with the K88A mutant, suggesting that this site is important for this process (**Figure 5g, h**).

Ubiquitination of K88 would lead to sterical hindrance and would prevent the formation of H-bridges of K88 with the DNA. Thus, it is likely that this ubiquitination would at least reduce the DNA binding capacity of the WH domain. However, previous experiments showed that K88 is less critical for DNA binding than R44 and K45, which bind to the major groove of the DNA [8]. Also, the homologous WH domain of KAT6A shows only subtle association with the minor groove [10], indicating that this interaction is not absolutely required for the DNA binding. Thus, we cannot exclude the possibility that ubiquitination of K88 still allows binding of SAMD1 to the DNA, albeit likely less efficiently.
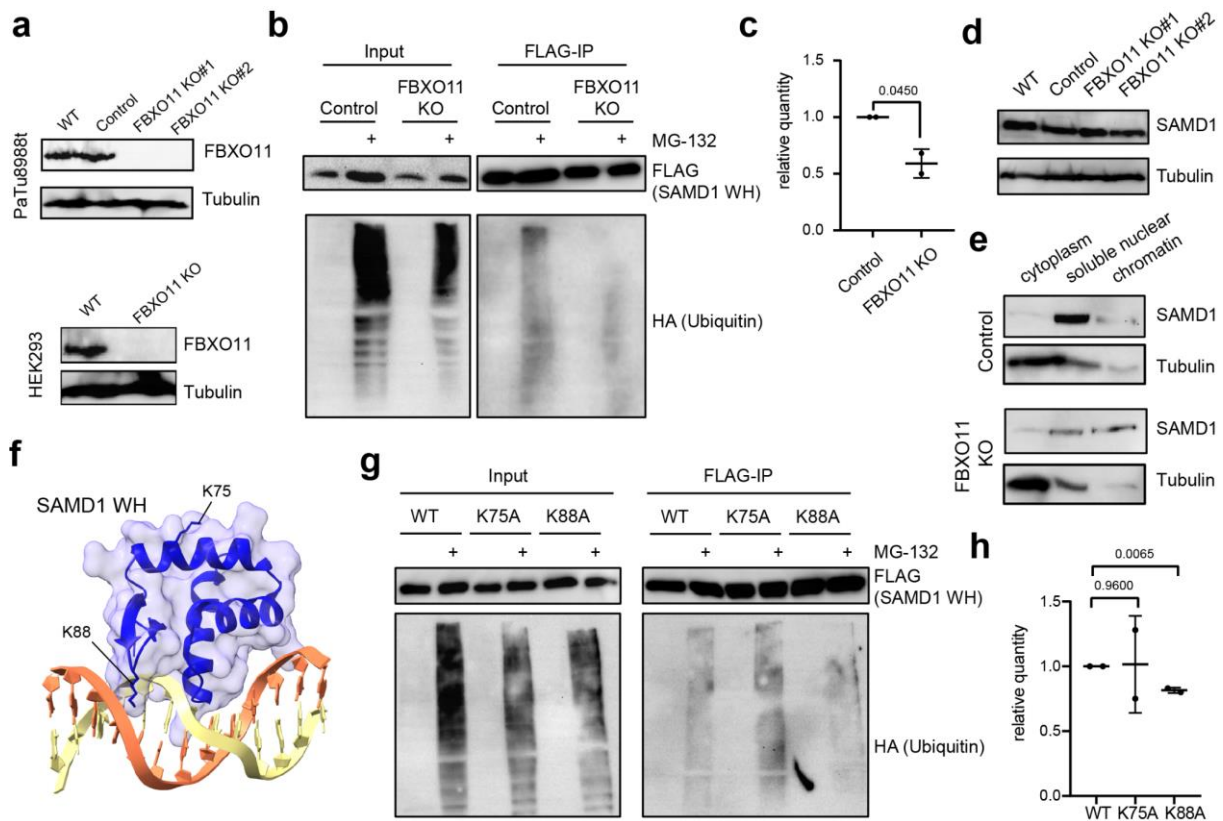
Figure 5: FBXO11 affects SAMD1 ubiquitination and chromatin association.

a) Western blot showing PaTu8988t wild-type cells, control cells, and two different FBXO11 knockout clones; western blot showing HEK293 cells with FBXO11 KO.
b) Ubiquitination assay of SAMD1 winged helix (WH) domain in HEK293 control and FBXO11 KO cells.
c) Quantification of b) using two biological replicates.
d) Western blot showing SAMD1 expression in PaTu8988t wild-type cells, control cells, and two different FBXO11 knockout clones.
e) Fractionation of PaTu8988t control and FBXO11 KO cells, followed by SAMD1 western blotting.
f) Structure of the SAMD1 WH domain (PDB: 6LUI)[8], with K75 and K88 indicated.
g) Ubiquitination assay of wild-type and mutated SAMD1 winged helix domains in HEK293 cells.
h) Quantification of g) using two biological replicates.

FBXO11 deletion influences SAMD1 chromatin binding genome-wide

To address the role of FBXO11 in SAMD1 chromatin binding in further detail, we performed ChIP-qPCR using our SAMD1 antibody. Consistent with the fractionation experiment (**Figure 5e**), we observed an enhanced chromatin binding of SAMD1 at the *CDH2* promoter in the FBXO11 KO PaTu8988t cells (**Figure 6a**). To investigate this effect at the genome-wide level, we performed ChIP-seq of SAMD1 in wild-type and FBXO11 KO cells. We found an increased SAMD1 binding at many locations in FBXO11 knockout cells (**Figure 6b, Supplementary Figure 7a**), supporting that FBXO11 is involved in inhibiting the chromatin binding of SAMD1 at a global level.

The impact of FBXO11 deletion was particularly evident at locations where SAMD1 chromatin binding was low under wild-type conditions (**Figure 6c, d,** groups 1 and 2, **Supplementary Figure 7b**). In contrast, in places where SAMD1 was already strongly present in the wild-type cells, the FBXO11 knockout had only minor effects (**Figure 6c, d,** group 3, **Supplementary Figure 7b**). Only at a very small fraction, with very high SAMD1 levels, does SAMD1 occupancy become weaker (**Figure 6c, d,** group 4, **Supplementary Figure 7b**). Closer inspection of the distinct groups showed that the locations more susceptible to FBXO11 deletion have smaller CpG islands (**Supplementary Figure 7c**) and are transcriptionally more active, signified by higher H3K4me3 and RNA Polymerase II levels (**Supplementary Figure 7d-e**). This observation raises the possibility that FBXO11 more strongly regulates SAMD1 chromatin binding at locations with fewer CpG binding motifs and higher transcriptional activity. On the other hand, no substantial differences regarding the gene expression levels of the associated genes could be recognized (**Supplementary Figure 7f**).

To investigate whether the enhanced binding of SAMD1 contributes to changes in gene expression, we analyzed several SAMD1 target genes via RT-qPCR. We found a significant reduction in *MTSS1* gene expression but no changes in other investigated genes (**Figure 6e**). This observation suggests that the increased chromatin binding of SAMD1 upon FBXO11 deletion has an impact on the gene expression of specific genes, consistent with the overall relatively low number of dysregulated genes upon SAMD1 deletion (**Figure 2a**).

Our biochemical and genome-wide data support the hypothesis that FBXO11 ubiquitinates SAMD1, which impairs its chromatin association and thereby may inhibit the gene repressive function of SAMD1. This mechanism could be important during cancer progression to overcome the tumor-suppressive role of SAMD1. This idea is further supported by the fact that *FBXO11* is commonly upregulated in PDAC (**Figure 6f**), and its high expression correlates with worse patient prognosis (**Figure 6g**). Furthermore, a high *FBXO11* expression level was linked to increased expression of EMT pathway genes (**Figure 6h**), which is opposite to what we observed before for *SAMD1* (**Figure 1b**). Together, these data support that in PDAC cells, the tumor-suppressive function of SAMD1 is counteracted by FBXO11.
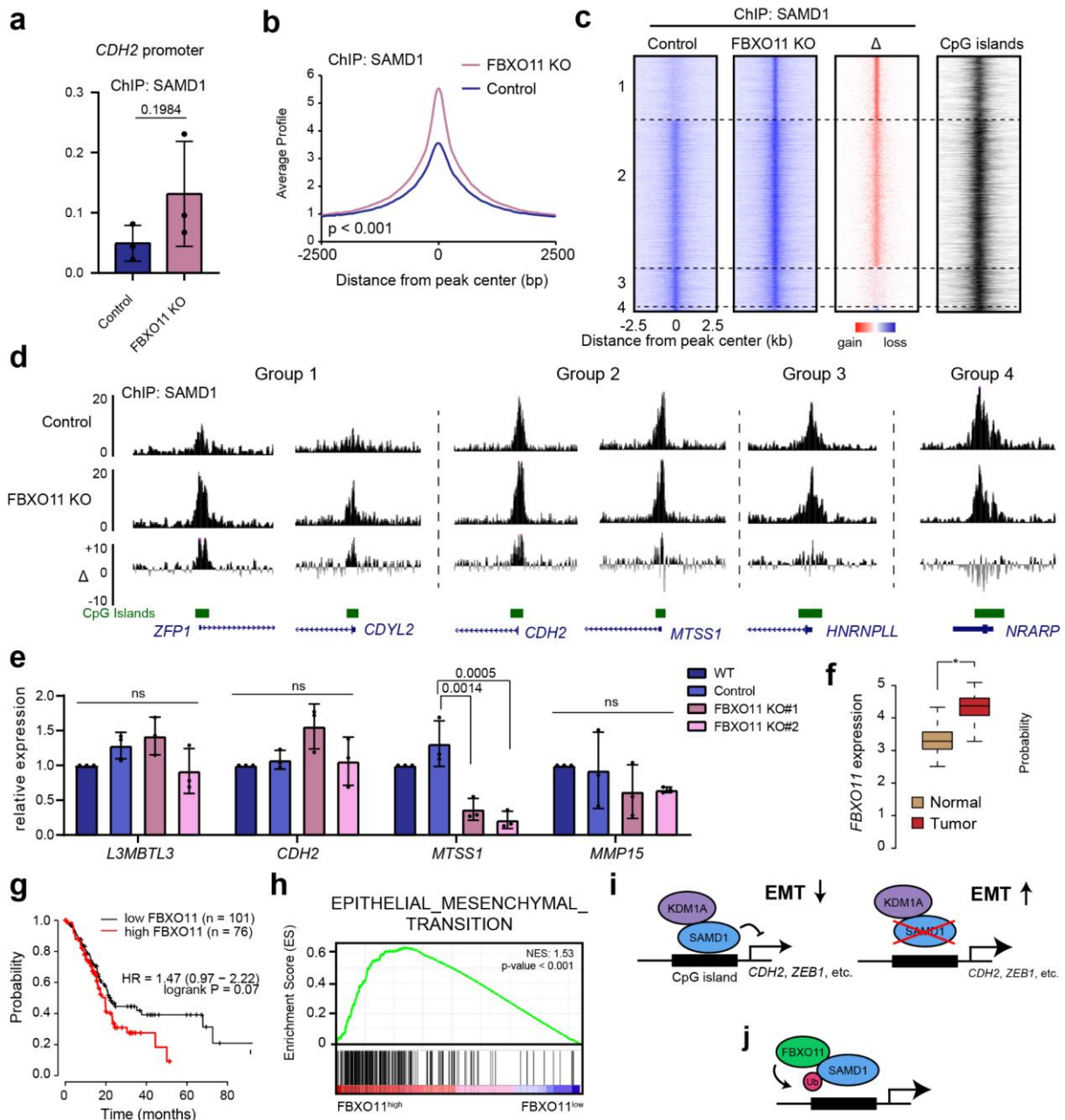
Figure 6: FBXO11 counteracts SAMD1.

a) ChIP-qPCR at the CDH2 promoter in PaTu8988t Control and FBXO11 KO cells, using a SAMD1 antibody. Data represent the mean ± SD of three biological replicates.

b) Profile of SAMD1 at SAMD1 peaks, in wild-type and FBXO11 KO cells. Significance was evaluated via a Kolmogorov-Smirnov test. See also **Supplementary Figure 7a**.

c) Heatmaps showing all SAMD1 peaks in PaTu8988t control and FBXO11 KO cells. Peaks were grouped according to the gain or loss after FBXO11 KO (Group 1, n = 3028; Group 2, n = 6895; Group 3, n = 1883; Group 4, n = 229). Δ indicates the difference between control and FBXO11 KO. Peaks overlap with CpG islands. See also **Supplementary Figure 7b**.

d) Snapshots of the USCS browser showing examples of the four groups observed in c).

e) RT-qPCR analysis of SAMD1 target genes with enhanced SAMD1 chromatin binding upon FBXO11 KO. Data represent the mean ± SD of three biological replicates. Significance was analyzed using one-way ANOVA.

f) Expression of FBXO11 in PDAC versus normal tissues. Data from TCGA [20] and visualized via GePIA [21].

g) Kaplan Meier survival curve showing the correlation of *FBXO11* expression with patient survival. Graph was visualized via KM-Plotter [24].

h) GSEA for epithelial-mesenchymal transition using TCGA data analyzed for high and low FBXO11 expression.

i) Model of the role of SAMD1 in PDAC. SAMD1 represses EMT-related genes, thereby suppressing migration. In SAMD1 KO cells, the EMT genes become upregulated leading to an enhanced EMT-related phenotype.

j) FBXO11 is involved in counteracting SAMD1 by ubiquitination, which inhibits SAMD1 chromatin association.

Discussion

Pancreatic ductal adenocarcinoma (PDAC) is a highly lethal form of cancer [4]. Local invasion and metastasis, driven by uncontrolled epithelial-mesenchymal transition (EMT), are the main reasons for the aggressive nature of PDAC [23]. Unfortunately, current strategies for inhibiting EMT in PDAC are not sufficiently effective. In this work, we identified the chromatin regulator SAMD1 as an important suppressor of EMT-related pathways in PDAC.

Analysis of patient data revealed that *SAMD1* is frequently dysregulated in cancer, and its expression often correlates with favorable or unfavorable prognoses. In the context of PDAC samples, *SAMD1* expression is commonly upregulated, (**Supplementary Figure 1a**), but Kaplan Meier survival curves demonstrated that high SAMD1 expression is associated with a better outcome (**Figure 1a**), suggesting a tumor-suppressive role for SAMD1. Biological assays utilizing SAMD1 KO cells were conducted to investigate this further, revealing increased migration rates and upregulation of EMT-related pathways upon SAMD1 loss (**Figure 1h, i, Figure 6i**), further supporting a potential tumor-suppressive function of SAMD1.

During this work, we noticed a decreased chromatin binding of SAMD1 in PDAC cell lines compared to other tumor cell lines (**Figure 4a**). This finding led us to speculate that the correlation with survival may not be solely determined by *SAMD1* expression levels but rather by the chromatin-binding level of SAMD1 itself. Via IP-MS experiments we revealed the FBXO11-containing E3-ubiquitin ligase complex as a novel interactor of SAMD1 (**Figure 4b**), which acts as an inhibitor of SAMD1 chromating binding (**Figure 6j**). FBXO11 plays a versatile role in cancer, acting both as an oncogene and as a tumor-suppressor. It targets oncogenic proteins, such as BCL-6 or the Snail family of transcription factors, for degradation, thereby exhibiting a tumor-suppressive function in various cancer types, such as diffuse large B-cell lymphomas and lung cancer [31,35]. In lung cancer cell lines, the FBXO11-containing complex was found to neddylate p53, thereby inhibiting its transcriptional activity [36]. In contrast, in pancreatic ductal adenocarcinoma, silencing FBXO11 suppresses tumor development [37], which may involve the ubiquitination of p53, supporting an oncogenic role in this cancer type. As a result, high *FBXO11* expression is associated with a worse prognosis in PDAC (**Figure 6d**) [37]. Our finding that FBXO11 inhibits SAMD1 chromatin association represents a novel regulatory mechanism of FBXO11 in cancer (**Figure 6j**). The FBXO11-SAMD1 axis could potentially be relevant for more cancer types beyond PDAC.

In our IP-MS experiments, we successfully identified another key interactor of SAMD1 – the KDM1A histone demethylase complex (**Figure 4b**). KDM1A, also referred to as LSD1, has previously been characterized as an interactor of SAMD1 [8]. In mouse embryonic stem cells, the deletion of SAMD1 results in a reduction of KDM1A-binding on specific promoters [8], implying that SAMD1 plays a role in the recruitment of KDM1A to chromatin. We confirmed that KDM1A is also an interactor of SAMD1 in PDAC (**Figure 3c**) and that deletion of SAMD1 leads to reduced recruitment of KDM1A to chromatin (**Figure 3a**) and altered H3K4 methylation levels (**Figure 3b**). Besides the involvement of SAMD1 in the recruitment of KDM1A, our work supports the idea that the presence of SAMD1 in the KDM1A complex influences the catalytic activity of KDM1A (**Figure 3g, h, Figure 6i**). One could speculate that the association of SAMD1 with KDM1A affects the conformation of the KDM1A complex, which in turn may allow a more efficient demethylation reaction. Another possibility is that SAMD1, when bound to the KDM1A complex, enhances the association of the KDM1A complex with its nucleosomal substrate, which increases the efficiency of demethylation. It is also possible that the association of KDM1A with the FBXO11 complex contributes to the altered demethylase activity of KDM1A. Thus, more research will be required to clarify the potentially sophisticated interplay of KDM1A, SAMD1, and FBXO11.

CDH2 is one of the top SAMD1 target genes and shares occupancy with KDM1A (**Figure 2, Figure 3a**). Our investigations have demonstrated that *CDH2*, the gene encoding N-cadherin, serves as the main driver of enhanced migration after SAMD1 KO (**Figure 2j, k**). During EMT progression, cadherins play a central role. The downregulation of epithelial cadherin (E-cadherin), which is critical for adherens junction formation, is accompanied by an elevation in neural cadherin (N-cadherin) expression. This shift contributes to heightened cell mobility and a more mesenchymal phenotype [38]. While N-cadherin is only detectable in nearly 50% of all PDAC patient samples, its presence in metastatic lesions significantly correlates with augmented neural invasion and a higher histological grade [26]. Remarkably, tumors exhibiting high N-cadherin levels also show elevated *TGFB* expression, which resonates with our observations indicating the upregulation of TGFB-signaling consequent to SAMD1 knockout (**Figure 2b**). Besides *CDH2*, many other EMT-related genes, including *ZEB1* [39], *BMP2* [40], and Netrin-1 (*NTN1*) [41], are targeted and repressed by SAMD1 (**Figure 6i**).

Based on our RNA-seq experiment performed here (**Figure 2a**) and previously [8,19], most SAMD1 target genes are only subtly influenced by SAMD1 deletion, leading to less than 10-fold upregulation. However, given that SAMD1 targets thousands of genes (**Figure 2d**), it is likely that SAMD1 has a substantial influence on the transcriptional landscape in the cells. This hypothesis is supported by the intense dysregulation of cellular pathways during differentiation processes [8] and by the severe phenotype of the SAMD1 knockout mice [16]. It is also notable that the impact of SAMD1 on gene transcription is rather cell-type specific. Except for the *L3MBTL3* gene, which appears to be commonly dysregulated upon SAMD1 deletion (**Figure 2a**) [8,19], possibly due to a feedback mechanism, most other genes are affected by SAMD1 in a cell type-specific manner [19]. Thus, the role of SAMD1 is likely highly context-dependent, and more work will be necessary to fully understand the specific functions of SAMD1 in the various physiological and pathophysiological contexts.

In PDAC, our work identified a tumor-suppressive role of SAMD1 by inhibiting EMT-related genes. By exploiting this functionality, such as enhancing SAMD1 chromatin binding by disrupting its interaction with FBXO11 (**Figure 6j**), it may be possible to impede EMT during

PDAC progression. Thus, our findings serve as a foundation for future investigations into the therapeutic possibilities of targeting SAMD1 in PDAC and other diseases.

Data availability

ChIP-seq and RNA-seq data were uploaded to the Gene Expression Omnibus (GEO) database, with the accession numbers GSE239415 (review token: mvkfecmqxhanzcx) and GSE239414 (reviewer token: ipkjaeqgdxqhfyd), respectively. Genome browser link to ChIP-seq data: http://genome.ucsc.edu/s/oasufai23/sjbgsiud-SAMD1PDAC-Review. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [42] partner repository with the dataset identifier PXD044104 (Reviewer account details: Username: reviewer_pxd044104@ebi.ac.uk, Password: lb4cOc3O).

Funding

Author Contributions

C.S. performed most biochemical and biological experiments, I.D.B performed most ubiquitination experiments and created FBXO11 KO cells, B.S. performed most ChIP experiments and prepared ChIP-seq samples, A.M.S. performed transwell migration assays, L.M.J. and M.G. performed mapping experiments between KDM1A and FBXO11, M.B. helped with analysis of time-lapse analysis experiments, I.F. performed mass-spectrometry analysis, I.R. created plasmids for most experiments, M.G, L.M.W. and S.F. contributed experiments and/or material, A.N. performed next-generation sequencing, R.L. performed bioinformatic analyses, T.S., M.H. and R.L supervised the work. C.S. and R.L. conceived the study. R.L. and C.S. wrote the manuscript with input from all authors.

Acknowledgments

Material and Methods

Cell culture

Patu8988t cells were cultured in DMEM, high glucose, GlutaMAX™ Supplement (Thermo Fisher Scientific; 61965026) supplemented with 5% fetal bovine serum (Thermo Fisher Scientific; 10270106). PANC-1 and CFPAC cells were cultured in DMEM, high glucose, GlutaMAX™ Supplement supplemented with 10% fetal bovine serum respectively. BxPC3 cells were kept in RPMI 1640 Medium, GlutaMAX™ Supplement (Thermo Fisher Scientific; 61870036) supplemented with 10% fetal bovine serum whereas HL-60 cells obtained 15%

fetal bovine serum. HEK 293 cells were grown in DMEM/F-12, GlutaMAX™ Supplement supplemented with 10% fetal bovine serum. HepG2 cells were cultured in MEM, GlutaMAX™ (Thermo Fisher Scientific; 41090036) supplemented with 10% fetal bovine serum and 1x nonessential amino acids (Thermo Fisher Scientific; 11140050). All cell lines were cultured with 1% penicillin-streptomycin (Thermo Fisher Scientific; 15140122).

Antibodies

All antibodies used are described in the methods subsections and in **Supplementary Table S1**.

Stable cell line generation

A SAMD1 knockout was conducted using the Lenti-CRISPR V2 plasmid containing either an unspecific control or guide RNAs targeting *SAMD1* (sg1: AGCGCATCTGCCGGATGGTG; sg2: GAGCATCTCGTACCGCAACG), *CDH2* (sg1: GCCTGAAGCCAACCTTAACTG; sg2: GAGACAATTCAGTAAGCACAG; sg3: GAACTTGCCAGAAAACTCCAG), *FBXO11* (sg1: GAGCCTCTTGTACCCCACCA; sg2: GTGTCCCACAAAGAACAGTA; sg3: GTTTTCTGTAGTTGAAGTTG).

Cells were transfected using Polyethylenimine, Linear, MW 25000, Transfection Grade (Polysciences; 23966) and Opti-MEM™ (Thermo Fisher Scientific; 31985062). Selection for single clones was performed using 2 µg/µl puromycin (Merck; 58-58-2) for PaTu8988t cells and 0.3 µg/µl for BxPC3 cells. The knockout was confirmed by western blot or immunofluorescence.

To rescue SAMD1, PaTu8988t cells were transfected with SAMD1 constructs containing a FLAG-HA-ER tag. Positive clones were selected using 10 µg/ml blasticidin. After selection, the concentration was reduced to 5 µg/ml blasticidin. Nuclear translocation of FLAG-ER-SAMD1 was induced by adding 200 nM 4-OHT (Merck; 68392-35-8) for 24 h.

Nuclear Extract Preparation

To obtain the nuclear extract, the cytoplasmic fraction was removed by incubating harvested cells for 10 min at 4 °C in low salt buffer (10 mM HEPES/KOH (pH=7.9), 10 mM KCl, 1.5 mM MgCl2, 1xPIC (cOmplete™, Protease Inhibitor Cocktail (Roche; 04693116001)), 0.5 mM PMSF). After centrifugation, the remaining pellet was dissolved in high salt buffer (20 mM HEPES/KOH pH=7.9, 420 mM NaCl, 1.5 mM MgCl2, 0.2 mM EDTA, 20% glycerol, 1x PIC, 0.5 mM PMSF) and incubated for 20 min at 4 °C while shaking. Subsequently, the lysates were centrifuged, and the supernatant containing the nuclear fraction was further analyzed by western blotting.

Subcellular Fractionation

A subcellular protein fractionation kit for cultured cells (Thermo Fisher Scientific; 78840) was used for fractionation experiments according to the manufacturer's instructions. A 10 cm dish format was applied, which corresponded to a packed cell volume of 20 µl per well.

Western Blot

Western blots were conducted using the Trans-Blot® Turbo™ Transfer System (BioRad; 1704150). The following antibodies were used: anti-tubulin (Merck; MAB3408), anti-SAMD1 antibody (Bethyl; A303-578A), anti-FBXO11 (Novus Biologicals; NB100-59826), anti-KDM1A (Abcam; AB17721), anti-HA (Merck; 11867423), and anti-FLAG (Merck; F3165). Full Western blots are presented in **Supplementary Figure S8**.

Immunofluorescence Staining

For immunofluorescence staining, cells were seeded on coverslips. On the next day, the cells were fixed with 4% methanol-free formaldehyde (Thermo Fisher Scientific; PI28906), and subsequently permeabilized with 0.5% Triton X-100 in PBS. Blocking was performed with 10% FBS + 0.5% Triton X-100 in PBS. Primary antibody incubation was performed for 1 h in a wet chamber. The following primary antibodies were used at a 1:500 dilution in blocking solution: a homemade SAMD1 antibody recognizing the SAM domain, an HA-antibody (Merck; 11867423), and an N-cadherin antibody (Thermo Fisher Scientific; 33-3900). Next, the cells were washed three times with 0.5% Triton X-100 in PBS. Secondary antibody incubation was conducted using Alexa Fluor 488 and 546 coupled antibodies (Thermo Fisher Scientific; A-11008, A-11081; A-11001) at a 1:1000 dilution. To stain the actin cytoskeleton, cells were stained with 1x Phalloidin-California Red Conjugate (Santa Cruz; sc-499440) for 20 min. Following three washing steps, the coverslips were mounted onto microscopy slides using ProLong™ Diamond mounting medium (Thermo Fisher Scientific; P36961). Photos were taken using a Leica DM 5500 microscope.

Proliferation Assay

To determine proliferation rates, cells were seeded in technical triplicates on 6-well plates at a density of $5x10^4$ cells per well. Cell viability was determined 1, 3, and 7 days after seeding for BxPC3 cells and 1, 3, and 5 days after seeding for PaTu8988t cells using the MTT assay. Therefore, 90 µl of 5 mg/ml thiazolyl blue ≥98% (Carl Roth; 4022) was added to each well. After 1 h, the medium was aspirated, and stained cells were dissolved in 400 µl of lysis buffer (80% isopropanol, 10% 1 M HCl, 10% Triton X-100) and diluted further with PBS if necessary. Absorption was measured at 595 nm using a plate reader. All values were normalized to day 1 to compensate for variations in seeding density. The mean value of three biological replicates was determined.

Wound Healing Assay

To determine the migration rate of SAMD1 knockout cells, PaTu8988t and BxPC3 cells were seeded in culture inserts (Ibidi; 80209). A total of 70 µl of cell suspension at a density of $6x10^5$ cells per ml was applied. On the next day, the insert was directly removed for BxPC3 cells whereas PaTu8988t cells were starved with medium containing 0.5% FBS for 6 h before removing the insert. Photos were taken using an Olympus CKX53 microscope. After 7 h for BxPC3 cells and after 24 h for PaTu8988t cells, photos were taken on the same spots and the cell-free area was measured for both timepoints using ImageJ Fiji (version: 2.1.0/1.53r).

Transswell Migration Assay

To determine whether SAMD1 KO has any effects on the migratory potential of PaTu8988t cells, transwell migration assays were performed. Therefore, transwell inserts with a pore size of 8.0 µm (BD Biosciences; 353097) were placed in the wells of a corresponding 24-well plate (Corning; 353504) containing 600 µl serum-free DMEM, high glucose, and GlutaMAX™ Supplement with or without 5% FBS as a chemoattractant. $2x10^4$ PaTu8988t cells in 300 µl serum-free DMEM medium were seeded per transwell insert. The cells were allowed to migrate through the filter for 18 h. Non-migrated cells were removed from the upper transwell insert by wiping them out and performing thorough washing steps in PBS. The migrated cells present on the bottom side of the transwell filter were fixed in methanol for at least 3 minutes and stained with crystal violet solution (0.2% in 20% methanol, 1:5 dilution in dH_2O) for 10 minutes at room temperature. Membranes were washed in aqua bidest and dried prior to fixing them on microscopy coverslips using Vectashield® with DAPI (Vector Laboratories; H-1200). Evaluation of migrated tumor cells was performed under a Leica DMI3000B microscope. Migrated cells were counted in seven visual fields per filter using ImageJ (version: 2.0.0-rc-43/1.52n). Migration was depicted relative to control.

Time Lapse Analysis

To perform time lapse analysis, PaTu8988t cells were seeded on collagen-coated 6-well plates at a density of $5x10^4$ cells per well. Coating was performed with 30 mg collagen (Merck; 50201) per ml acetic acid (0.02 M) for 2 h at 37 °C. Afterwards, the plates were washed three times with PBS before seeding the cells.

On the next day, the cells were placed in a Zeiss LSM 780 microscope and every 10 min photos were taken for 24 h using a 10x DIC objective. Migration of cells was analyzed using the "Time Lapse Analyzer" (University of Ulm, version tla_src_v01_33). As a setup file, "DIC tracking 1" was used. Migration was measured in µm per min.

Measurement of cell shape

To determine the cell shape, cells were seeded on 6-well plates at low density ($3x10^4$/well). Photos were taken three days after seeding. For treatment with ADH-1, cells were seeded on 24-well plates ($1x10^4$) and treated directly after seeding. Photos were taken one day later using an Olympus CKX53 microscope. For each condition, three photos were taken and 10 cells per photo were analyzed by measuring the circularity of single cells using ImageJ Fiji (version: 2.1.0/1.53r).

RNA Preparation

For RNA isolation, cells were cultivated on 6-well plates up to 80-100% confluency. RNA was prepared according to the manufacturer's manual using the RNeasy Mini Kit (Qiagen; 74004) including an on-column DNA digest.

cDNA Synthesis

The Tetro cDNA Synthesis Kit (Bioline; BIO-151 65043) was used to transcribe mRNA into cDNA according to the manufacturer's manual. Samples were incubated at 45 °C for 50 min

followed by 5 min at 85 °C to inactivate Tetro RT. Subsequently, cDNA was diluted 1:20 for use in RT-qPCR.

RT-qPCR

For analysis by real-time quantitative PCR, MyTaq™ Mix (Bioline; BIO-25041) was used. For gene expression analysis, values were normalized to GAPDH. Primers are displayed in **Supplementary Table S2.**

Ectopic Co-immunoprecipitation

All ectopic coimmunoprecipitation (Co-IP) experiments were performed in HEK293 cells. Cells were seeded in 10-cm dishes at 2×10^6 cells per dish. One day later, the expression constructs for 3xHA or FLAG–tagged proteins were transfected using Polyethylenimine, Linear, MW 25000, Transfection Grade (Polysciences; 23966) and Opti-MEM™ (Thermo Fisher Scientific; 31985062). When the interaction of two proteins should be studied in the presence of a KDM1A inhibitor, the medium was exchanged 5 h after transfection to medium containing either DMSO or 20 nM ORY-1001 (Cay19136; Biomol).

Two days after transfection, extract was prepared using Co-IP buffer (50 mM Tris/Cl (pH=7.5), 150 mM NaCl, 1% Triton X-100, 1 mM EDTA, 10% glycerol, 1xPIC (cOmplete™, Protease Inhibitor Cocktail (Roche; 04693116001), 0.5 mM PMSF). ANTI-FLAG M2 Affinity Gel (Merck, A2220) beads were equilibrated by washing two times with 1× TBS and one time with Co-IP buffer. To bind FLAG-tagged proteins, extracts were added to 50 µl of beads and incubated for approximately 3 hours head over tail at 4°C. After incubation, three washing steps with Co-IP buffer were performed. The FLAG beads were boiled for 3 min in 2× Laemmli buffer without β-mercaptoethanol. Subsequently, β-mercaptoethanol was added to the supernatant and the samples were analyzed via western blotting.

For Co-IP experiments, the following constructs were used:

SAMD1:

construct	amino acids
FL	1-538
ΔSAM	1-450
ΔWH	111-538
WH	1-110
SAM	451-538

KDM1A:

construct	amino acids
FL	1-852
IDR/SWIRM/AOD-N	1-417

Tower	418-513
AOD-C	514-852
IDR	1-172
SWIRM	173-272
AOD-N	273-417

Ubiquitination Assay

To perform a ubiquitination assay, HEK 293 cells stably overexpressing 3xHA-tagged ubiquitin were seeded on 15 cm dishes and transfected one day afterwards with the respective FLAG-tagged constructs using Polyethylenimine, Linear, MW 25000, Transfection Grade (Polysciences; 23966) and Opti-MEM™ (Thermo Fisher Scientific; 31985062). Before preparing extracts, cells were treated with 10 µM MG-132 (M7449; Merck) or DMSO as control for 5 h. Extract preparation and IP were performed according to the ectopic co-immunoprecipitation protocol (see above) and samples were analyzed via western blotting.

Histone Demethylase Assay

The protocol for a histone demethylase assay was modified after Laurent et al (2015). Wild-type HEK293 cells, HEK293 cells stably overexpressing FLAG-KDM1A and HEK293 cells with SAMD1 KO stably overexpressing FLAG-KDM1A were used. Per reaction, an extract was prepared with buffer A (10 mM HEPES (pH=7.6), 3 mM $MgCl_2$, 300 mM KCl, 5% glycerol, 0.5% NP-40, 1x PhosSTOP™ (Roche; 4906845001), 0.5 mM PMSF, 1 µg/µl pepstatin, 10 µg/µl aprotinin, 10 µg/µl leupeptin) using five 15 cm dishes.Subsequently, the extract was diluted by half with buffer B (10 mM HEPES (pH=7.6), 3 mM $MgCl_2$, 10 mM KCl, 5% Glycerol, 0.5% NP-40, 1x PhosSTOP™ (Roche; 4906845001), 0.5 mM PMSF, 1µg/µl Pepstatin I 10 µg/µl Aprotinin I 10 µg/µl Leupeptin). 50 µl of ANTI-FLAG M2 Affinity Gel (Merck, A2220) beads were equilibrated with buffer B. Extracts were added to the equilibrated beads and incubated for 3 h head over tail at 4 °C. Next, the beads were washed three times with buffer B and two volumes (100 µl) of demethylase buffer (50 mM Tris/Cl (pH=8.5), 50 mM KCl, 5 mM $MgCl_2$, 0.5% BSA, 5% glycerol, 0.5 mM PMSF, 1 µg/µl pepstatin, 10 µg/µl aprotinin, 10 µg/µl leupeptin, 500 µg/mL FLAG® peptide (Merck; F3290)) were added to the beads. To start the demethylase reaction, 3 µg of calf histones (Merck; H9250) was added, and the samples were incubated for 4 h at 37 °C while shaking. The reaction was stopped by boiling the samples for 5 min in 5x lämmli buffer. Subsequently, the samples were analyzed via western blotting.

Extract preparation and IP for Mass Spectrometry

For IP mass spectrometry, PaTu8988t cells with SAMD1 KO stably expressing either FH-ER as a control or FH-ER SAMD1 were used. For each construct, 20 15 cm dishes were seeded and the nuclear translocation of SAMD1 was induced 24 h before extract preparation by adding 200 nM 4-OHT (Merck; 68392-35-8). After collection, cells were centrifuged at 2,000 rpm and 4 °C for 10 min. The cell pellet was resuspended in 5x pellet volume hypotonic buffer (10 mM Tris (pH=7.3), 10 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM PMSF, 10 mM ß-mercaptoethanol, 1xPIC (cOmplete™, Protease Inhibitor Cocktail (Roche; 04693116001)) and shaken at 4 °C for 10-

15 min. Next, cells were centrifuged at 2,500 rpm and 4 °C for 10 min. The cell pellet was resuspended again in 5x pellet volume hypotonic buffer and denounced 40x in a cell douncer. To remove cell debris, lysates were centrifuged at 3,500 rpm and 4 °C for 15 min. The pellet was resuspended in 1x pellet volume low salt buffer (20 mM Tris/Cl (pH=7.3), 20 mM KCl, 1.5 mM MgCl$_2$, 0.2 mM EDTA, 25% glycerol, 0.2 mM PMSF, 10 mM ß-mercaptoethanol, 1xPIC) and dounced 10x. The sample was shaken in a thermomixer and 0.66x pellet volume of high salt buffer (20 mM Tris/Cl (pH=7.3), 1.2 M KCl, 1.5 mM MgCl$_2$, 0.2 mM EDTA, 25% glycerol, 0.2 mM PMSF, 10 mM β-mercaptoethanol) was added dropwise. The extract was shaken for 45 min and centrifuged afterwards for 30 min at 13,000 rpm and 4 °C.

The supernatant containing the proteins was transferred to a a Slide-A-Lyzer™ G2 Dialysis Cassette (3.5K) (Thermo Fisher Scientific; 87724) and dialyzed against 3 L of dialysis buffer (20 mM Tris/Cl (pH=7.3), 100 mM KCl, 0.2 mM EDTA, 20% glycerol, 0.2 mM PMSF, 1 mM DTT) overnight.

To perform the FLAG-IP, the material was retrieved from the dialysis chambers and centrifuged at 13,000 rpm and 4 °C for 30 min. Afterwards, a benzonase® nuclease (Merck;70664) digest (1 µl of benzonase per 500 µl extract) was performed for 1 h on ice. 40 µl of ANTI-FLAG M2 Affinity Gel (Merck, A2220) beads were equilibrated per IP by washing once with TAP buffer (50 mM Tris/Cl (pH=7.9), 100 mM KCl, 5 mM MgCl$_2$, 0.2 mM EDTA, 10% glycerol, 0.1% NP-40, 0.2 mM PMSF, 1 mM DTT), three times with 100 mM glycine (pH=2.5), once with 1 M Tris/Cl (pH=7.9) and finally once again with TAP buffer. Subsequently, the extracts were added to the prepared beads and incubated for 3 h, head over tail at 4 °C. Afterwards the beads were washed 3x with TAP buffer and 3x with 50 mM ammonium hydrogen carbonate. The washed beads were then sent in for mass-spectrometry analysis at the Biomedical Center Munich, protein analysis unit (Head: Axel Imhof), where the enriched proteins were analyzed using a Q Exactive HF Orbitrap Mass Spectrometer, as described previously [10].

Endogenous Co-IP

For endogenous Co-IP between SAMD1 and KDM1A, an extract was prepared according to the extract preparation protocol for mass spectrometry and the extract was dialyzed overnight (see above). For each IP, one 15 cm dish of PaTu8988t cells was used. Dynabeads™ Protein A (Thermo Fisher Scientific; 10008D) were equilibrated with TAP buffer (50 mM Tris/Cl (pH=7.9), 100 mM KCl, 5 mM MgCl$_2$, 0.2 mM EDTA, 10% glycerol, 0.1 % NP-40, 0.2 mM PMSF, 1 mM DTT) and subsequently the extract was precleared for 30 min with 10 µl of beads per IP before adding 2 µg of antibody per IP. Self-made IgG and SAMD1 antibodies and an anti-KDM1A (Abcam; AB17721) antibody were applied. After incubation for 3h, 20 µl of equilibrated Dynabeads™ Protein A per IP was added and incubated for another 2 h. The beads were washed 3x with TAP buffer before boiling in 2x Lämmli buffer.

Chromatin Preparation

To prepare chromatin, cells were seeded on 15 cm plates at 3x10^6 cells per plate and cultivated until reaching 70-90% confluency. First, 1% formaldehyde was added to the medium and the plates were slowly swayed for 10 min to fix the cells. The fixation was stopped by adding 125 mM glycine for 5 min. Subsequently, the cells were washed twice with PBS and scraped in 1 ml cold buffer B (10 mM HEPES/KOH (pH=6.5), 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100) per 15 cm plate. All plates containing the same cell line were pooled in a 15 ml tube.

The tubes were centrifuged for 5 min at 2000 rpm and 4 °C. The supernatant was removed, and the pellet was resuspended in 1 ml cold buffer C (10 mM HEPES/KOH (pH=6.5), 10 mM EDTA, 0.5 mM EGTA, 200 mM NaCl) per 15 cm plate followed by a 15 min incubation time on ice. Then the tubes were centrifuged with the same settings as mentioned before. After removing the supernatant, the pellet was resuspended in 200 µl cold buffer D (50 mM Tris/HCl (pH=8.0), 10 mM EDTA, 1% SDS, 1xPIC (cOmplete™, Protease Inhibitor Cocktail (Roche; 04693116001)) per 15 cm plate, vortexed, and incubated for 10-20 min on ice. For shearing the chromatin, the samples were sonicated two times for 7 min each using a precooled Bioruptor® (Diagenode). The samples were centrifuged for 10 min at 13.000 rpm and 4 °C. The supernatant contained the sheared chromatin.

Chromatin Immunoprecipitation

Chromatin-immunoprecipitation (ChIP) for ChIP-qPCR was performed according to the One-day ChIP kit protocol (Diagenode; C01010080). Beads were exchanged to Dynabeads™ Protein A (Thermo Fisher Scientific; 10008D), and Chelex to Chelex 100 Resin (BioRad;142-1253) and ChIP buffer was replaced by a homemade buffer (50 mM Tris/Cl (pH=7.5), 150 mM NaCl, 5 mM EDT, 1% Triton X-100, 0.5% NP-40). For each ChIP, 3 µg of either IgG control antibody (Diagenode; C15410206) or of a specific antibody were applied. For histone marks only 1 µg of antibody was used. The following antibodies were used: a self-made SAMD1 antibody recognizing the SAM domain, a self-made L3MBTL3 antibody recognizing the SAM domain, anti-KDM1A (Abcam; AB17721), anti-H3K3me2 (Diagenode; C15410035), and anti-H3K4me3 (Diagenode; C15410003).

To prepare samples for ChIP-sequencing, the One-day ChIP kit protocol was used as described above, but the DNA-purification was modified. For DNA elution, beads were incubated with 230 µl elution buffer (100 mM NaHCO$_3$, 1% SDS) for 30 min at room temperature while shaking. Afterward, the samples were centrifuged at 13.000 rpm for 1 min and 200 µl of supernatant was transferred to a fresh tube. The input DNA was dissolved in 50 µl of dH2O and 150 µl of elution buffer was added to obtain an equal volume in all samples. 8 µl of 5 M NaCl were added to each sample and the samples were incubated at 65 °C overnight to reverse the cross-linking.

On the next day, 8 µL of 1M Tris/Cl (pH=6.5), 4 µL 0.5 M EDTA, and 2 µL of Proteinase K (10 µg/µL) were added to each sample and all samples were incubated at 45 °C for 1 h whilst shaking. DNA was purified using the QIAquick PCR Purification Kit (Qiagen; 28104) whereby all samples prepared with the same antibody were pooled on the same column. To elute the DNA, columns were incubated for 1 min with 30 µl of sterile 2 mM Tris/Cl (pH=8.5), and centrifuged at 13.000 rpm for 1 min.

The concentration of the samples was determined using the Quant-iT™ dsDNA Assay Kit (Thermo Fisher Scientific; Q33120) and the NanoDrop™ 3300 (Thermo Fisher Scientific). At least 4 ng of DNA was used for library preparation.

Library Preparation and Next Generation Sequencing

Next generation sequencing was performed at the Genomics Core Facility Marburg (Center for Tumor Biology and Immunology, Hans-Meerwein-Str. 3, 35043 Marburg, Germany). For ChIP-seq, the Microplex library preparation kit v2 (Diagenode, C05010012) was used for

indexed sequencing library preparation with chromatin immunoprecipitated DNA. Libraries were purified on AMPure magnetic beads (Beckman; A6388). RNA was prepared as described in *RNA preparation* and integrity was assessed on an Experion StdSens RNA Chip (Bio-Rad; 7007103). RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Prep kit (Illumina, 2002059). RNA-seq and ChIP-seq libraries were quantified on a Bioanalyzer (Agilent Technologies). Next-generation sequencing was performed on an Illumina NextSeq 550.

Bioinformatic analyses

ChIP-seq data were mapped to the human genome hg38 using bowtie [44], allowing 1 mismatch. BigWig files were obtained using DeepTools/bamCoverage [45]. Significant peaks were obtained using Galaxy/MACS2 (2.2.7.1) [46]. Heatmaps and profiles were created using Galaxy/DeepTools [45]. The top target genes were identified based on the SAMD1 ChIP-seq signal at promoters.

RNA-seq data were aligned to the human transcriptome (GenCode 43) using Galaxy/RNA-Star (2.7.10b) [47]. Differentially expressed genes were obtained using DeSeq2 (2.11.40.7) [48]. Genes with a log2-fold change of more than 0.5 and a p-value lower than 0.01 were considered significantly dysregulated. Gene set enrichment analysis was performed using GSEA software with standard settings [22].

The following internet databases and tools were used: Galaxy Europa [49], DeepTools [45], GREAT (4.0.4) [43], Bioconductor/R [50], GSEA (4.3.2) [22], GePIA [21], GDC Data Portal [20] and Kaplan-Meier-Plotter [24].

The following public ChIP-seq data were used: ChIP-seq of H3K4me3 (GSM945261)[51] and RNA Polymerase II (GSM1010788)[52] in PANC-1 cells.

Statistical analysis

Statistical analysis was performed as described in the figure legends. Error bars indicate standard deviation (SD). The significance of the qPCR results was analyzed via ANOVA or Student's t-tests. The significance of the GSEA was evaluated by the GSEA software. The significance of changes in SAMD1 ChIP-seq levels was evaluated using a two-sided Kolmogorov-Smirnov test. All biological experiments were performed in at least three replicates. RNA-seq was performed with three replicates, using three independent SAMD1 KO clones.

References

1   Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* **71**, 209-249, doi:10.3322/caac.21660 (2021).
2   Cronin, K. A. *et al.* Annual report to the nation on the status of cancer, part 1: National cancer statistics. *Cancer* **128**, 4251-4284, doi:CNCR34479 [pii]
10.1002/cncr.34479 (2022).
3   Oberstein, P. E. & Olive, K. P. Pancreatic cancer: why is it so hard to treat? *Therap Adv Gastroenterol* **6**, 321-337, doi:10.1177_1756283X13478680 [pii]
10.1177/1756283X13478680 (2013).
4   Kleeff, J. *et al.* Pancreatic cancer. *Nat Rev Dis Primers* **2**, 16022, doi:nrdp201622 [pii]
10.1038/nrdp.2016.22 (2016).

5   Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. *Cell* **166**, 21-45, doi:S0092-8674(16)30796-6 [pii]
10.1016/j.cell.2016.06.028 (2016).

6   Samuel, N. & Hudson, T. J. The molecular and cellular heterogeneity of pancreatic ductal adenocarcinoma. *Nat Rev Gastroenterol Hepatol* **9**, 77-87, doi:nrgastro.2011.215 [pii]
10.1038/nrgastro.2011.215 (2011).

7   Uzunparmak, B. & Sahin, I. H. Pancreatic cancer microenvironment: a current dilemma. *Clin Transl Med* **8**, 2, doi:10.1186/s40169-019-0221-1 [pii]
221 [pii]
10.1186/s40169-019-0221-1 (2019).

8   Stielow, B. *et al.* The SAM domain-containing protein 1 (SAMD1) acts as a repressive chromatin regulator at unmethylated CpG islands. *Sci Adv* **7**, doi:7/20/eabf2229 [pii]
abf2229 [pii]
10.1126/sciadv.abf2229 (2021).

9   Stielow, B., Simon, C. & Liefke, R. Making fundamental scientific discoveries by combining information from literature, databases, and computational tools - An example. *Comput Struct Biotechnol J* **19**, 3027-3033, doi:S2001-0370(21)00164-1 [pii]
10.1016/j.csbj.2021.04.052 (2021).

10   Weber, L. M. *et al.* The histone acetyltransferase KAT6A is recruited to unmethylated CpG islands via a DNA binding winged helix domain. *Nucleic Acids Res* **51**, 574-594, doi:6946999 [pii]
gkac1188 [pii]
10.1093/nar/gkac1188 (2023).

11   Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010-1022, doi:25/10/1010 [pii]
10.1101/gad.2037511 (2011).

12   Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A. & Skalnik, D. G. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol* **20**, 2108-2121, doi:2052 [pii]
10.1128/MCB.20.6.2108-2121.2000 (2000).

13   Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082-1086, doi:nature08924 [pii]
10.1038/nature08924 (2010).

14   Li, H. *et al.* Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* **549**, 287-291, doi:nature23881 [pii]
10.1038/nature23881 (2017).

15   Fischer, S. & Liefke, R. Polycomb-like Proteins in Gene Regulation and Cancer. *Genes (Basel)* **14**, doi:genes14040938 [pii]
genes-14-00938 [pii]
10.3390/genes14040938 (2023).

16   Campbell, B. *et al.* Investigation of SAMD1 ablation in mice. *Sci Rep* **13**, 3000, doi:10.1038/s41598-023-29779-3 [pii]
29779 [pii]
10.1038/s41598-023-29779-3 (2023).

17   Dungan, C. M. *et al.* Muscle-Specific Cellular and Molecular Adaptations to Late-Life Voluntary Concurrent Exercise. *Function (Oxf)* **3**, zqac027, doi:zqac027 [pii]
10.1093/function/zqac027 (2022).

18   Annear, D. J., Vandeweyer, G., Sanchis-Juan, A., Raymond, F. L. & Kooy, R. F. Non-Mendelian inheritance patterns and extreme deviation rates of CGG repeats in autism. *Genome Res* **32**, 1967-1980, doi:gr.277011.122 [pii]
10.1101/gr.277011.122 (2022).

19   Simon, C. *et al.* The CpG Island-Binding Protein SAMD1 Contributes to an Unfavorable Gene Signature in HepG2 Hepatocellular Carcinoma Cells. *Biology (Basel)* **11**, doi:biology11040557 [pii]

biology-11-00557 [pii]
10.3390/biology11040557 (2022).

20   Weinstein, J. N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-1120, doi:ng.2764 [pii]
10.1038/ng.2764 (2013).

21   Tang, Z., Li, C., Kang, B., Gao, G. & Zhang, Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* **45**, W98-W102, doi:3605636 [pii]
10.1093/nar/gkx247 (2017).

22   Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:0506580102 [pii]
10.1073/pnas.0506580102 (2005).

23   Wang, S., Huang, S. & Sun, Y. L. Epithelial-Mesenchymal Transition in Pancreatic Cancer: A Review. *Biomed Res Int* **2017**, 2646148, doi:10.1155/2017/2646148 (2018).

24   Nagy, A., Munkacsy, G. & Gyorffy, B. Pancancer survival analysis of cancer hallmark genes. *Sci Rep* **11**, 6047, doi:10.1038/s41598-021-84787-5 [pii]
84787 [pii]
10.1038/s41598-021-84787-5 (2021).

25   Derycke, L. D. & Bracke, M. E. N-cadherin in the spotlight of cell-cell adhesion, differentiation, embryogenesis, invasion and signalling. *Int J Dev Biol* **48**, 463-476, doi:041793ld [pii]
10.1387/ijdb.041793ld (2004).

26   Nakajima, S. *et al.* N-cadherin expression and epithelial-mesenchymal transition in pancreatic carcinoma. *Clin Cancer Res* **10**, 4125-4133, doi:10/12/4125 [pii]
10.1158/1078-0432.CCR-0578-03 (2004).

27   Shintani, Y. *et al.* ADH-1 suppresses N-cadherin-dependent pancreatic cancer progression. *Int J Cancer* **122**, 71-77, doi:10.1002/ijc.23027 (2008).

28   Maes, T. *et al.* ORY-1001, a Potent and Selective Covalent KDM1A Inhibitor, for the Treatment of Acute Leukemia. *Cancer Cell* **33**, 495-511 e412, doi:S1535-6108(18)30023-0 [pii]
10.1016/j.ccell.2018.02.002 (2018).

29   Shi, Y. J. *et al.* Regulation of LSD1 histone demethylase activity by its associated factors. *Mol Cell* **19**, 857-864, doi:S1097-2765(05)01569-8 [pii]
10.1016/j.molcel.2005.08.027 (2005).

30   Zheng, N. *et al.* Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**, 703-709, doi:416703a [pii]
10.1038/416703a (2002).

31   Duan, S. *et al.* FBXO11 targets BCL6 for degradation and is inactivated in diffuse large B-cell lymphomas. *Nature* **481**, 90-93, doi:nature10688 [pii]
10.1038/nature10688 (2011).

32   Rossi, M. *et al.* Regulation of the CRL4(Cdt2) ubiquitin ligase and cell-cycle exit by the SCF(Fbxo11) ubiquitin ligase. *Mol Cell* **49**, 1159-1166, doi:S1097-2765(13)00131-7 [pii]
10.1016/j.molcel.2013.02.004 (2013).

33   Xu, P. *et al.* FBXO11-mediated proteolysis of BAHD1 relieves PRC2-dependent transcriptional repression in erythropoiesis. *Blood* **137**, 155-167, doi:S0006-4971(21)00021-5 [pii]
2020/BLD2020007809 [pii]
10.1182/blood.2020007809 (2020).

34   Hornbeck, P. V. *et al.* PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512-520, doi:gku1267 [pii]
10.1093/nar/gku1267 (2015).

35   Jin, Y. *et al.* FBXO11 promotes ubiquitination of the Snail family of transcription factors in cancer progression and epidermal development. *Cancer Lett* **362**, 70-82, doi:S0304-3835(15)00239-6 [pii]
10.1016/j.canlet.2015.03.037 (2015).

36   Abida, W. M., Nikolaev, A., Zhao, W., Zhang, W. & Gu, W. FBXO11 promotes the Neddylation of p53 and inhibits its transcriptional activity. *J Biol Chem* **282**, 1797-1804, doi:S0021-9258(20)72179-4 [pii]
10.1074/jbc.M609001200 (2007).
37   Xue, J., Chen, S., Ge, D. & Yuan, X. Ultrasound-targeted microbubble destruction-mediated silencing of FBXO11 suppresses development of pancreatic cancer. *Hum Cell* **35**, 1174-1191, doi:10.1007/s13577-022-00700-w [pii]
10.1007/s13577-022-00700-w (2022).
38   Loh, C. Y. *et al.* The E-Cadherin and N-Cadherin Switch in Epithelial-to-Mesenchymal Transition: Signaling, Therapeutic Implications, and Challenges. *Cells* **8**, doi:cells8101118 [pii]
cells-08-01118 [pii]
10.3390/cells8101118 (2019).
39   Zhang, P., Sun, Y. & Ma, L. ZEB1: at the crossroads of epithelial-mesenchymal transition, metastasis and therapy resistance. *Cell Cycle* **14**, 481-487, doi:1006048 [pii]
10.1080/15384101.2015.1006048 (2015).
40   Gordon, K. J., Kirkbride, K. C., How, T. & Blobe, G. C. Bone morphogenetic proteins induce pancreatic cancer cell invasiveness through a Smad1-dependent mechanism that involves matrix metalloproteinase-2. *Carcinogenesis* **30**, 238-248, doi:bgn274 [pii]
10.1093/carcin/bgn274 (2009).
41   Lengrand, J. *et al.* Pharmacological targeting of netrin-1 inhibits EMT in cancer. *Nature*, doi:10.1038/s41586-023-06372-2 [pii]
10.1038/s41586-023-06372-2 (2023).
42   Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* **50**, D543-D552, doi:6415112 [pii]
gkab1038 [pii]
10.1093/nar/gkab1038 (2022).
43   McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:nbt.1630 [pii]
10.1038/nbt.1630 (2010).
44   Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:gb-2009-10-3-r25 [pii]
10.1186/gb-2009-10-3-r25 (2009).
45   Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191, doi:gku365 [pii]
10.1093/nar/gku365 (2014).
46   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:gb-2008-9-9-r137 [pii]
10.1186/gb-2008-9-9-r137 (2008).
47   Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:bts635 [pii]
10.1093/bioinformatics/bts635 (2013).
48   Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:s13059-014-0550-8 [pii]
550 [pii]
10.1186/s13059-014-0550-8 (2014).
49   Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res* **50**, W345-W351, doi:6572001 [pii]
gkac247 [pii]
10.1093/nar/gkac247 (2022).
50   Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**, R80, doi:gb-2004-5-10-r80 [pii]
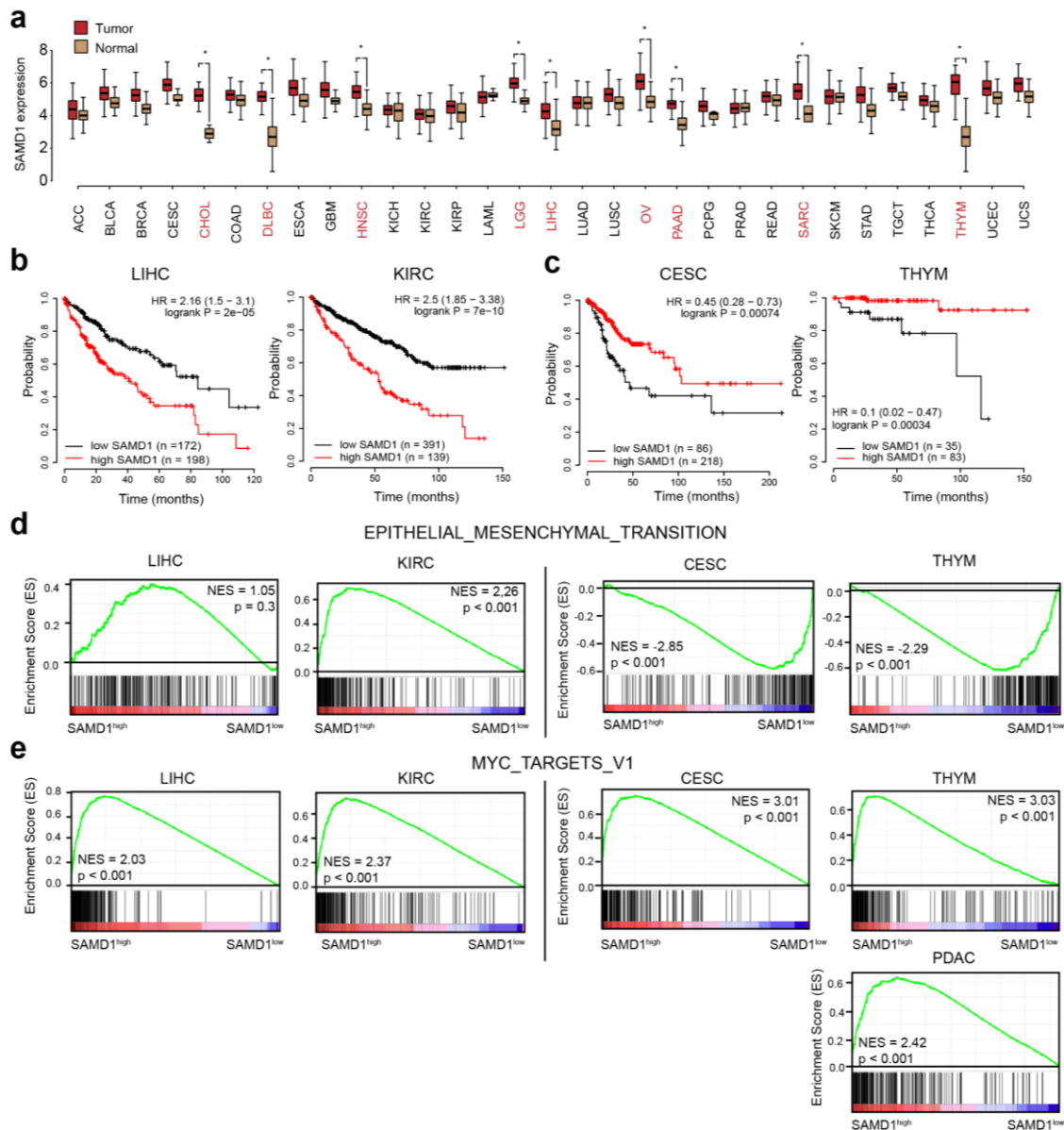
10.1186/gb-2004-5-10-r80 (2004).

51   Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:nature11232 [pii]

10.1038/nature11232 (2012).

52   Gertz, J. *et al.* Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol Cell* **52**, 25-36, doi:S1097-2765(13)00637-0 [pii]

10.1016/j.molcel.2013.08.037 (2013).

Supplementary Figure 1: SAMD1 has distinct roles in various cancer types.

a) Expression of *SAMD1* in cancer versus normal tissues. Data from TCGA [20] and visualized via GePIA [21]. Cancer types highlighted in red indicate significantly upregulated *SAMD1* expression.

b) Kaplan-Meier survival curves (overall survival) in liver hepatocellular carcinoma (LIHC) and kidney renal clear cell carcinoma (KIRC), using auto-selected cut-offs. High SAMD1 expression correlates with a worse prognosis.

c) Kaplan-Meier survival curves (overall survival) in cervical cancer (CESC) and thymoma (THYM). High *SAMD1* expression correlates with a better prognosis. Data in b) and c) are derived from TCGA and visualized via the Kaplan-Meier-Plotter tool [22] using auto-selected cut-off.

d) GSEA of the epithelial-mesenchymal transition (EMT) pathway in the cancer types presented in b) and c).

e) GSEA of the MYC target genes in the cancer types presented in b) and c) and of PDAC.
In d) and e), tissue samples with high *SAMD1* expression are compared to samples with low *SAMD1* expression.
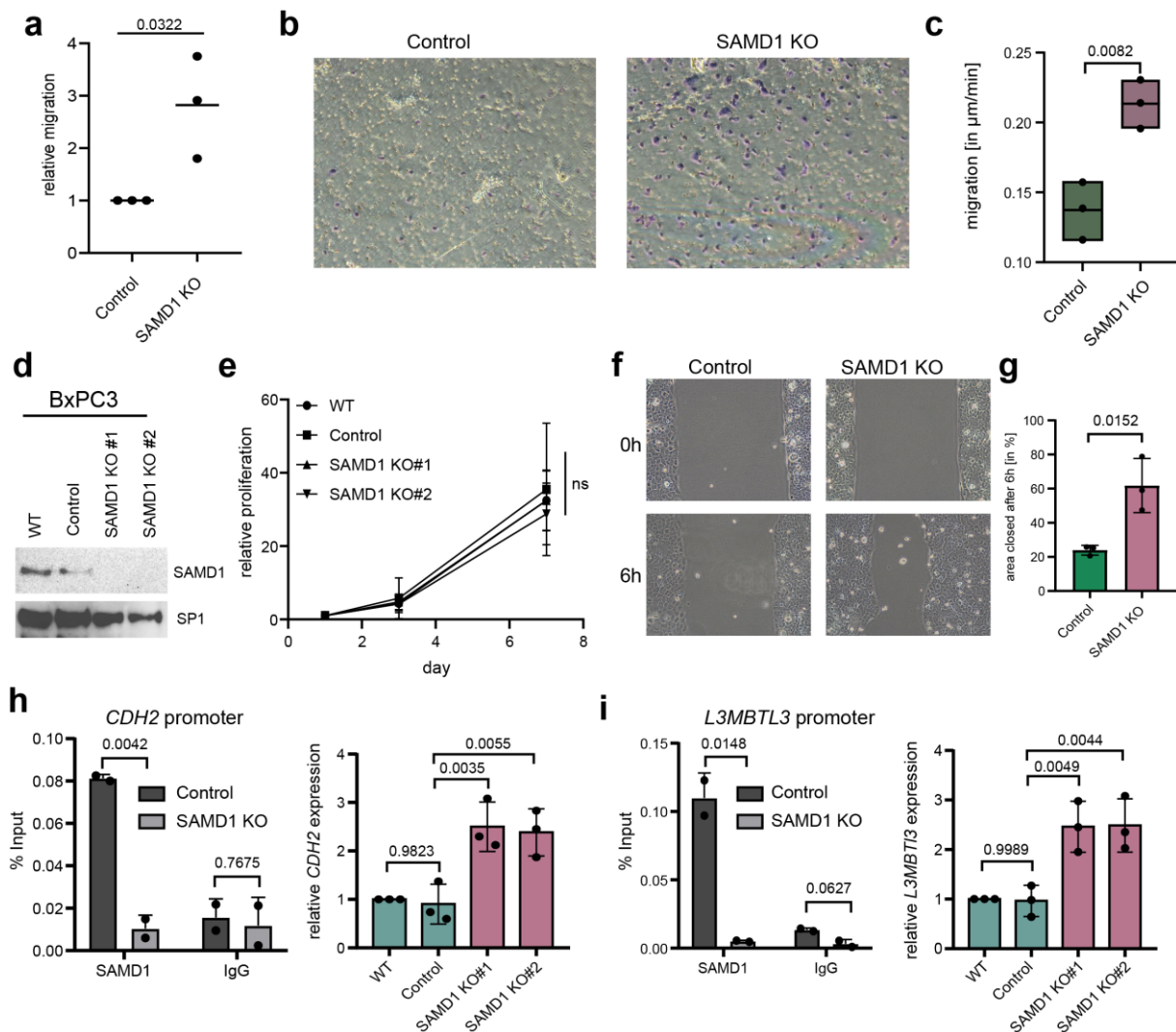
Supplementary Figure 2: transwell and time-lapse assays in PaTu8988t cells and SAMD1 KO in BxPC3 cells.

a) Transwell migration assay of PaTu8988t control and SAMD1 KO cells. Data represent the mean ± SD of three biological replicates. Significance was analyzed using Student's t-test.

b) Representative crystal violet staining of one transwell migration assay.

c) Migration of PaTu8988t control and SAMD1 KO cells in µm/min based on time-lapse analysis. See also **Supplementary Video 1 and 2**. Data represent the mean ± SD of three biological replicates. Significance was analyzed using Student's t-test.

d) Western blot showing BxPC3 wild-type cells, control cells, and two different SAMD1 knockout clones.

e) Proliferation assay of BxPC3 wild-type cells, control cells, and two different SAMD1 knockout clones. Data represent the mean ± SD of three biological replicates. Significance was analyzed using one-way ANOVA.

f) Representative picture of one wound healing assay of BxPC3 control cells and one SAMD1 knockout clone.

g) Quantification of the wound healing assay from c). Data represent the mean ± SD of three biological replicates, and significance was analyzed using Student's t-test.

h) ChIP-qPCR of *CDH2* promoter in BxPC3 Control and SAMD1 KO cells using IgG or SAMD1 antibodies. Significance was analyzed using Student's t-test, RT-qPCR showing *CDH2* expression in BxPC3 wild-type cells, control cells, and two different

SAMD1 knockout clones. Data represent the mean ± SD of three biological replicates. Significance was analyzed using one-way ANOVA.

i) ChIP-qPCR at the *L3MBTL3* promoter in BxPC3 control cells and SAMD1 KO cells, using SAMD1 or IgG antibodies. Significance was analyzed using Student's t-test, RT-qPCR showing *L3MBTL3* expression in BxPC3 wild-type cells, control cells, and two different SAMD1 knockout clones. Data represent the mean ± SD of three biological replicates. Significance was analyzed using one-way ANOVA.
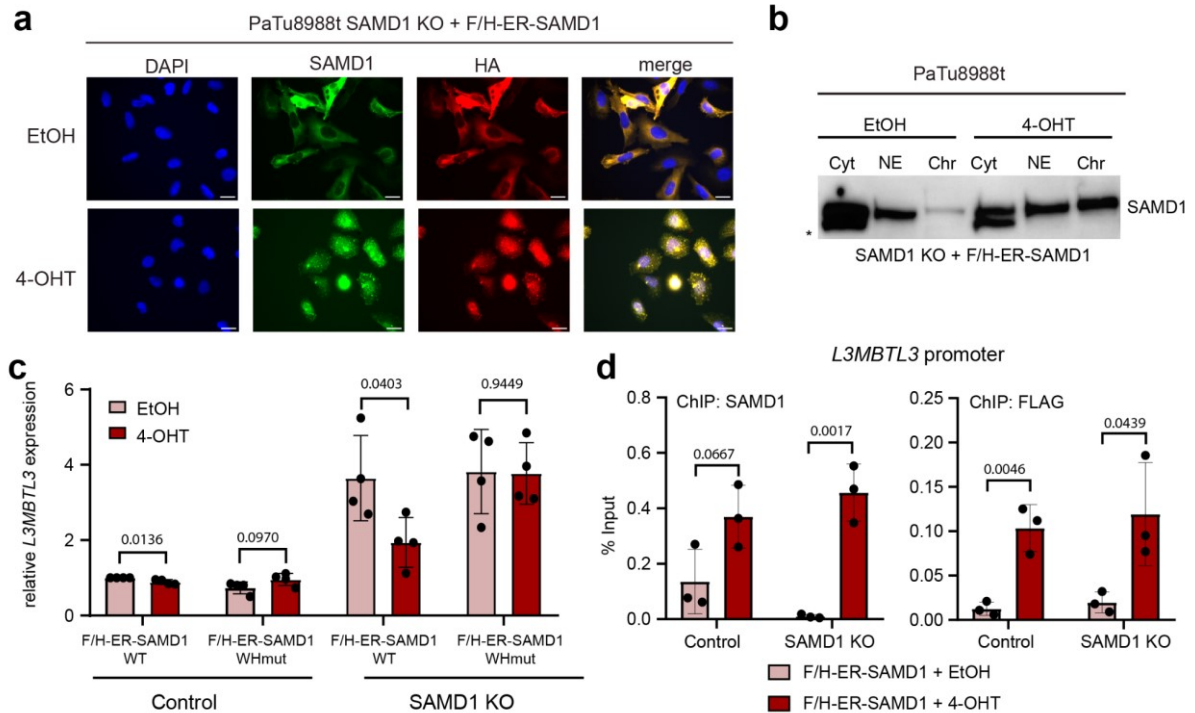
Supplementary Figure 3: SAMD1 rescue experiments in PaTu8988t cells.

a) Immunofluorescence of PaTu8988t SAMD1 knockout cells with or without induction of SAMD1 rescue, Bar=20 µM.

b) Western blot after fractionation of PaTu8988t SAMD1 knockout cells with or without induction of SAMD1 rescue.

c) RT-qPCR showing *L3MBTL3* expression with or without induction of SAMD1 rescue in PaTu8988t control and SAMD1 KO cells. WHmut=RK-45/46-AA mutation of SAMD1. Data represent the mean ± SD of four biological replicates. Significance was analyzed using Student's t-test.

d) SAMD1 ChIP-qPCR at the *L3MBTL3* promoter with or without induction of SAMD1 rescue in PaTu8988t Control and SAMD1 KO cells. Data represent the mean ± SD of three biological replicates. Significance was analyzed using Student's t-test.
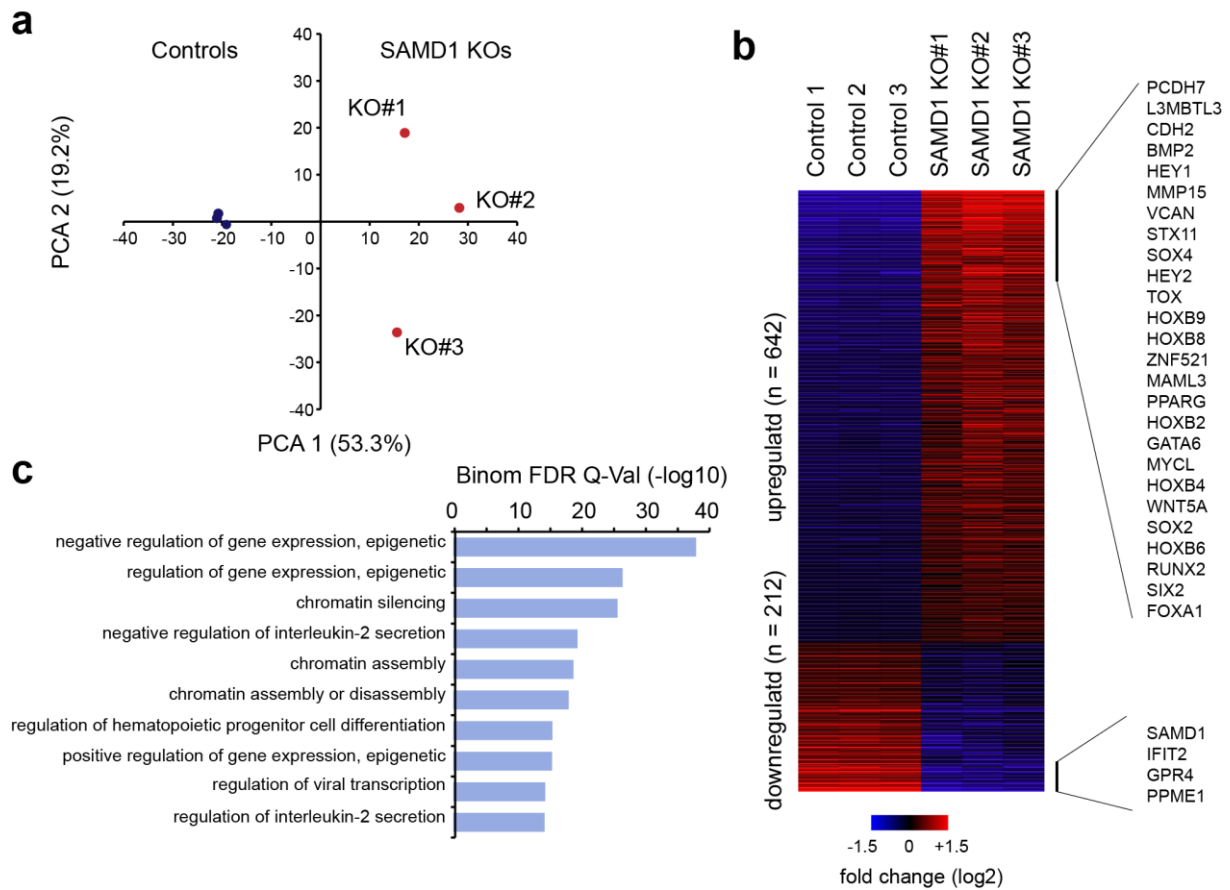
Supplementary Figure 4: Transcriptional regulation of SAMD1 in PaTu8998t cells.

a) Principal component analysis (PCA) of RNA-Seq data upon SAMD1 KO. Three clonally independent SAMD1 KO clones were used.

b) Heatmap of the significantly dysregulated genes. Examples of the most dysregulated genes are shown on the right.

c) Gene ontology analysis of SAMD1 genomic targets using GREAT [43].
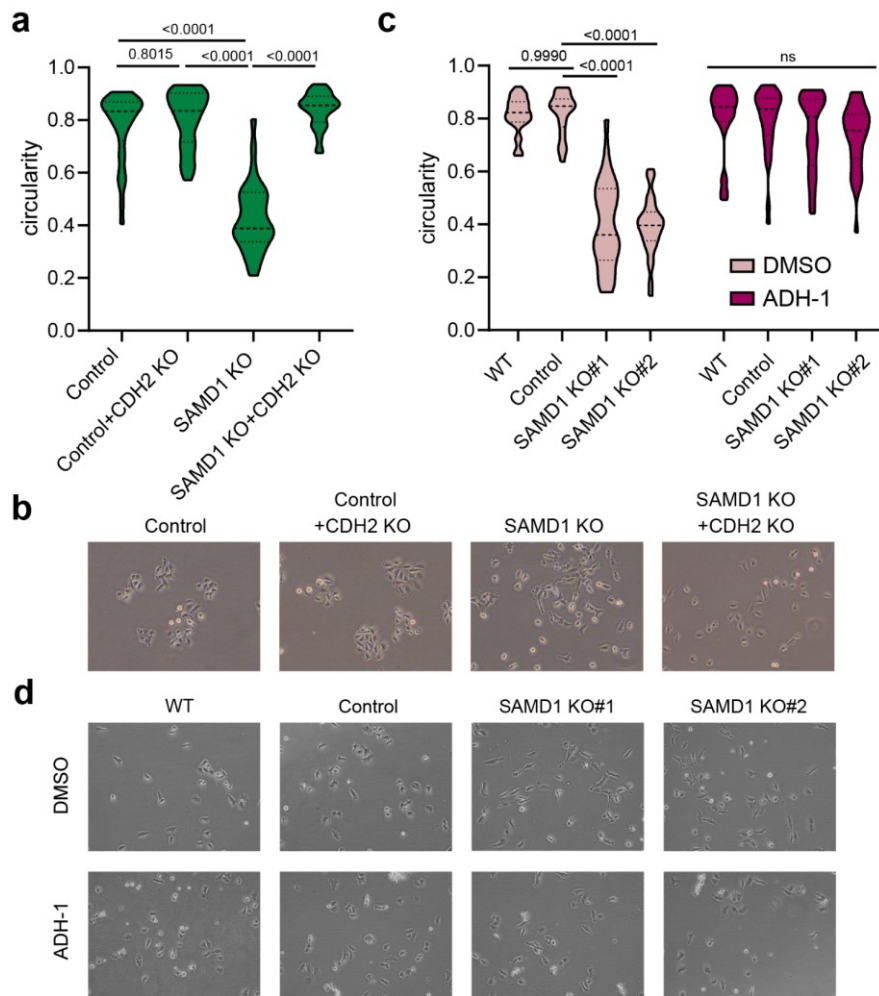
Supplementary Figure 5: CDH2 KO or inhibition rescues the migration phenotype upon SAMD1 deletion.

a) Cell shape of control, CDH2 KO, SAMD1 KO, and CDH2/SAMD1 double KO PaTu8988t cells. Circularity was determined using ImageJ Fiji. Significance was analyzed using one-way ANOVA.

b) Example bright field microscopy for a).

c) Cell shape of PaTu8988t wild-type cells, control cells and two different SAMD1 knockout clones with or without application of the N-cadherin inhibitor ADH-1. Circularity was determined using ImageJ Fiji. Significance was analyzed using one-way ANOVA.

d) Example bright field microscopy for c).
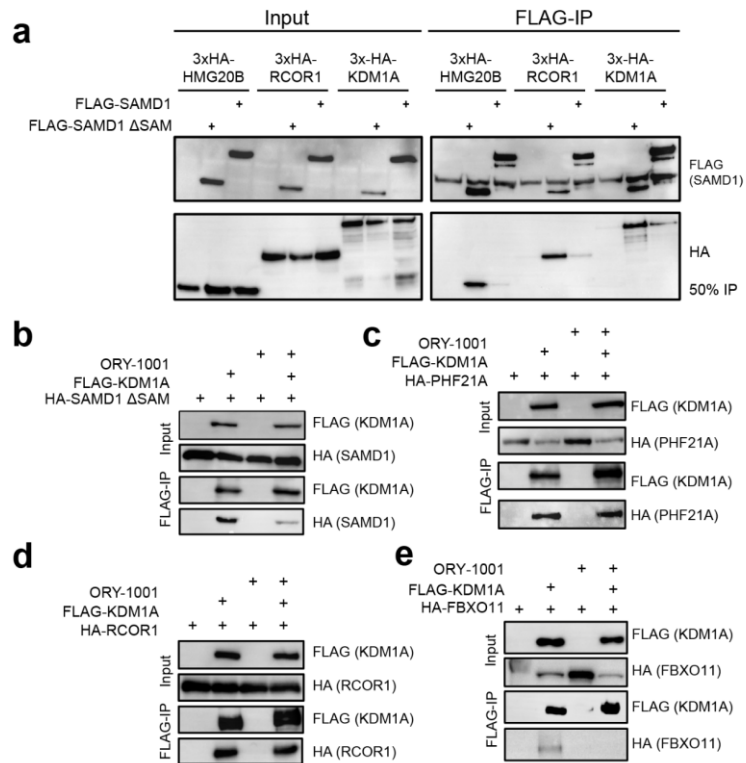
Supplementary Figure 6: SAMD1/KDM1A interaction is influenced by SAM domain and ORY-1001.

a) Co-immunoprecipitation in HEK293 cells showing the interaction between SAMD1 full-length or SAMD1 ΔSAM and the KDM1A-complex.

b) Co-immunoprecipitation in HEK293 cells showing the interaction between SAMD1ΔSAM and KDM1A upon treatment with the KDM1A inhibitor ORY-1001.

c) Co-immunoprecipitation in HEK293 cells showing the interaction between PHF21A and KDM1A upon treatment with the KDM1A inhibitor ORY-1001.

d) Co-immunoprecipitation in HEK293 cells showing the interaction between RCOR1 and KDM1A upon treatment with the KDM1A inhibitor ORY-1001.

e) Co-immunoprecipitation in HEK293 cells showing the interaction between FBXO11 and KDM1A upon treatment with the KDM1A inhibitor ORY-1001.
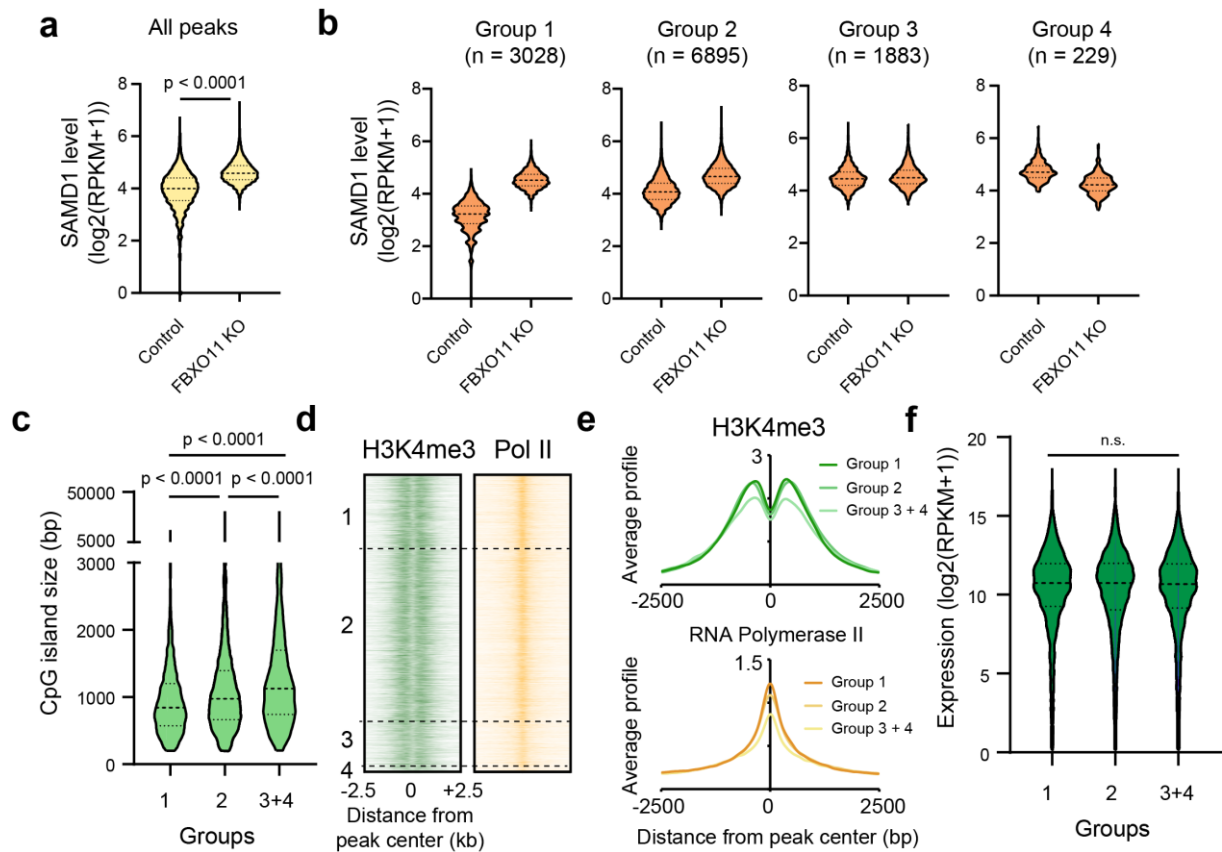
Supplementary Figure 7: Detailed analysis of the consequence of FBXO11 deletion on SAMD1 chromatin binding.

a) Violin plots showing the SAMD1 level in PaTu8988t control and FBXO11 KO cells. Statistical significance was evaluated using a Kolmogorov-Smirnov test.

b) SAMD1 levels in four different groups identified in **Figure 6g** in PaTu8988t control and FBXO11 KO cells.

c) CpG island size of four different groups identified in **Figure 6c**. Statistical significance was evaluated using a Kolmogorov-Smirnov test.

d) Heatmap of H3K4me3 and Pol II in the four different groups identified in **Figure 6c**.

e) Profiles of H3K4me3 and RNA Polymer II at the four different groups identified in **Figure 6c**.

f) Expression of genes at the four different groups identified in **Figure 6c**.

Curriculum Vitae

[removed for final print]

Ehrenwörtliche Erklärung

[removed for final print]

Danksagung

[removed for final print]