The tumour antigen PRAME is a subunit of a Cul2 ubiquitin ligase and associates with active NFY promoters

Adalberto Costessi^{1,5}, Nawel Mahrouf^{2,5}, Esther Tijchon¹, Rieka Stunnenberg¹, Marieke A Stoel¹, Pascal W Jansen¹, Dotan Sela², Skylar Martin-Brown², Michael P Washburn^{2,3}, Laurence Florens², Joan W Conaway^{2,4}, Ronald C Conaway^{2,4,*} and Hendrik G Stunnenberg^{1,*}

¹Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University, Nijmegen, The Netherlands, ²Stowers Institute for Medical Research, Kansas City, MO, USA, ³Department of Pathology and Laboratory Medicine, Kansas University Medical Center, Kansas City, KS, USA, and ⁴Department of Biochemistry and Molecular Biology, Kansas University Medical Center, Kansas City, KS, USA

The human tumour antigen PRAME (preferentially expressed antigen of melanoma) is frequently overexpressed in tumours. High PRAME levels correlate with poor clinical outcome of several cancers, but the mechanisms by which PRAME could be involved in tumourigenesis remain largely elusive. We applied protein-complex purification strategies and identified PRAME as a substrate recognition subunit of a Cullin2-based E3 ubiquitin ligase. PRAME can be recruited to DNA *in vitro*, and genome-wide chromatin immunoprecipitation experiments revealed that PRAME is specifically enriched at transcriptionally active promoters that are also bound by NFY and at enhancers. Our results are consistent with a role for the PRAME ubiquitin ligase complex in NFY-mediated transcriptional regulation.

The EMBO Journal (2011) 30, 3786–3798. doi:10.1038/emboj.2011.262; Published online 5 August 2011

Subject Categories: chromatin & transcription

Keywords: Cullin2; NFY; PRAME; transcription; ubiquitination

Introduction

PRAME is the founding member of a large family of PRAME genes (Birtle *et al*, 2005). PRAME genes have been implicated in oogenesis (Dadé *et al*, 2003), spermatogenesis (Wang *et al*, 2001) and stemness (Bortvin *et al*, 2003; Cinelli *et al*, 2008), but their function is poorly understood. Human PRAME

was initially identified as the protein triggering an autologous anti-tumour immune response in a melanoma patient (Ikeda *et al*, 1997). Since then, many reports have shown that PRAME is not expressed in most normal adult tissues, while it is overexpressed in a wide variety of solid and haematological malignancies (Kilpinen *et al*, 2008). Although not prognostic for all cancers (van Baren *et al*, 1998; Steinbach *et al*, 2002; Tajeddine *et al*, 2005), high PRAME expression has been found to correlate with the stage of melanoma lesions (Haqq *et al*, 2005), and with shorter overall survival of neuroblastoma (Oberthuer *et al*, 2004) and serous ovarian adenocarcinoma patients (Partheen *et al*, 2008). High PRAME mRNA levels were also shown to be an independent prognostic factor of poor clinical outcome in breast cancer (Doolan *et al*, 2007; Epping *et al*, 2008). In addition, PRAME is significantly activated in BCR-ABL-positive chronic myeloid leukaemia (Ph+ CML) patients at the onset of advanced phase and blast crisis (Radich *et al*, 2006), and hypomethylation of the PRAME locus seems to have a role in this process (Roman-Gomez *et al*, 2007; Schenk *et al*, 2007). A growing body of evidence indicates that PRAME contributes markedly to the oncogenic phenotype of several cancers. PRAME has been recently implicated in retinoic acid signalling and polycomb-mediated transcriptional repression in melanoma cells (Epping *et al*, 2005); however, these mechanisms do not seem to have a role in breast cancer (Epping *et al*, 2008) and acute myeloid leukaemia (Steinbach *et al*, 2007). While evidence that expression of PRAME is restricted mainly to tumour cells has designated it as a promising therapeutic target, efforts to develop strategies to interfere with the oncogenic activity of PRAME are hindered by a lack of information about its function(s) (Paydas, 2008).

In this study, we combined biochemical and genomics approaches to identify PRAME as a subunit of Cullin2-based E3 ubiquitin ligases. In addition, we show that PRAME is recruited in cells to epigenetically and transcriptionally active promoter regions bound by nuclear transcription factor Y (NFY), a transcription factor that is essential for early embryonic development (Bhattacharya *et al*, 2003) and has been implicated in the maintenance of the high proliferative capacity of ES cells (Grskovic *et al*, 2007) as well as in the inhibition of differentiation (Gurtner *et al*, 2003).

Results

PRAME associates with Cullin2-based E3 ubiquitin ligases

To investigate the molecular function of human PRAME, we used a proteomic approach to identify PRAME-interacting proteins. To this end, we stably expressed epitope-tagged versions of PRAME suitable for protein purifications in two cellular backgrounds: human K562 leukaemia cells, which

*Corresponding author. RC Conaway, Stowers Institute for Medical Research, Kansas City, MO 64112, USA. Tel.: +1 816 926 4092; Fax: +1 816 926 2092; E-mail: rcc@stowers.org or HG Stunnenberg, Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University, Geert Grooteplein 28, PO Box 9101, Nijmegen, HB 6500, The Netherlands. Tel.: +31 24 361 0524; Fax: +31 24 361 0520; E-mail: h.stunnenberg@ncmls.ru.nl
⁵These authors contributed equally to this work

Received: 7 February 2011; accepted: 4 July 2011; published online: 5 August 2011

Figure 1 PRAME assembles with Cullin2 ubiquitin ligases. **(A)** Ectopic expression of tagged versions of PRAME suitable for purifications. Whole-cell extracts of K562 cells or HA-immunoprecipitates from 293 cells were analysed by western blot with anti-PRAME antibodies. K562 cells expressing StreptII-MYC-3 × HA-PRAME (TAG-PRAME) at near-endogenous levels were selected, and 293 Flp-in cells stably expressing HA-PRAME were generated using the Invitrogen Flp-in system. **(B)** Superose 6 gel filtration analysis of nuclear extracts from K562-TAG-PRAME and K562 wild-type cells. Western blot was performed with anti-HA or affinity-purified anti-PRAME antibodies. Both TAG-PRAME and endogenous PRAME partially elute in high-molecular weight fractions. The numbers represent eluates fractions. A less prominent and faster migrating band for TAG-PRAME was detected, that most likely corresponds to a C-terminal truncation. **(C)** TAG-PRAME co-immunoprecipitates Cullin2-EloBC complex. Nuclear extracts from K562 wild-type and K562-TAG-PRAME cells were subjected to anti-TAG immunoprecipitation and western blotting as indicated; 0.8% of input and 33% of IP were loaded. A more slowly migrating form of Cullin2 was enriched after IP; this band is consistent with being post-translationally neddylated-Cullin2. Asterisks indicate antibody heavy chains. **(D)** Endogenous PRAME interacts with Cullin2-EloBC complex. Nuclear extracts from K562 wild-type cells were subjected to immunoprecipitations with anti-PRAME or preimmune sera and western blotting as indicated; 1% of input, 1% of supernatant and 20% of IP were loaded. PRAME was detected using a light-chain specific secondary antibody. PRAME was readily depleted from the extract after incubation with the specific antibodies (compare lanes 1, 2 and 4). Asterisks indicate protein A that dissociated from the beads after elution; the open arrow indicates PRAME; s/n, supernatant; IP, immunoprecipitates.

endogenously express detectable levels of PRAME, and human embryonic kidney HEK293 cells, which lack detectable endogenous PRAME (Figure 1A). When analysed by size exclusion chromatography, both endogenous and epitope-tagged PRAME from K562 cells showed a broad elution profile, with a substantial fraction of the protein eluting earlier than expected of free PRAME, which has a predicted molecular mass of 58 kDa (Figure 1B). This observation suggests that PRAME may be part of one or more protein complexes.

We purified PRAME-associated proteins from K562-TAG-PRAME or HEK293-HA-PRAME cells by anti-epitope immunoaffinity chromatography and analysed them by mass spectrometry. As shown in Tables I and II, eluates from

K562-TAG-PRAME and HEK293-HA-PRAME included a collection of proteins not found in control samples. Among the most highly represented were Cullin family member 2 (Cul2), Elongins B (TCEB2) and C (TCEB1), and RING finger protein Rbx1, all of which are subunits of a large family of multisubunit, Cullin-based, ubiquitin ligases. In these ubiquitin ligases, the Elongin BC dimer acts as an adaptor that links an Elongin BC-box containing substrate recognition subunit, or BC-box protein, to heterodimeric modules composed of either Cullin2 and Rbx1, or Cullin5 and Rbx2 (Kamura *et al*, 1998, 2004; Lonergan *et al*, 1998). The RING finger domains of the related Rbx1 and Rbx2 proteins in turn recruit and activate an E2 ubiquitin-conjugating enzyme for

ubiquitination of substrates (Kamura *et al*, 1999; Petroski and Deshaies, 2005).

The interaction of TAG-PRAME with ubiquitin ligase subunits was confirmed by western blotting of TAG-eluates. As shown in Figure 1C, these fractions contained PRAME, Elongins B and C, and two forms of Cul2 that most likely correspond to unmodified and a more slowly migrating Neddylated-Cul2, a hallmark of active ubiquitin ligases. Supporting the physiological relevance of these interactions, endogenous PRAME, Cul2, Elongin B and Elongin C were all detected in anti-PRAME immunoprecipitates from nuclear extracts of parental K562 cells (Figure 1D). Notably, while PRAME was clearly depleted from the extract after incubation with the specific antibodies (Figure 1D, compare lanes 1, 2 and 4), Cul2, Elongin B and Elongin C were not, consistent with their function as components of many Cullin-based ubiquitin ligases present in cells (Kamura *et al*, 1998; Mahrour *et al*, 2008).

PRAME is a BC-box protein

Although we identified all of the other subunits of Cul2-based ubiquitin ligases as PRAME-interacting proteins, we did not detect any of the previously identified BC-box proteins in our fractions. BC-box proteins are characterized by a sequence motif of about 10 amino acids, called the BC-box, which binds to Elongin C. In addition, a downstream Cullin-box specifies binding to either Cul2 or Cul5 (Kamura *et al*, 2004; Mahrour *et al*, 2008). Multiple sequence alignments of PRAME with several known BC-box proteins identified putative BC- and Cul2-boxes near the PRAME N-terminus (Figure 2A), raising the possibility that PRAME might function as a substrate recognition subunit of a Cul2-based ubiquitin ligase.

To assess the functionality of the predicted motifs, we introduced point mutations in key residues of the putative

BC-box (L26P/A30F) or the putative Cul2-box (L48A/P49A). We observed that the BC-box mutation greatly reduced the binding of PRAME to Elongin C and Cul2 in baculovirus-infected insect cells (Figure 2B). We next generated K562 and HEK293 cells stably expressing epitope-tagged PRAME carrying the BC-box or Cul2-box mutations and performed immunoprecipitations from nuclear extracts. In these cells, both mutations disrupted the interaction of PRAME with endogenous Cullin2 complex components (Figure 2C; Table II), demonstrating the functionality of the motifs.

In order to test the biochemical activity of the Cullin2-PRAME complex, we performed *in vitro* ubiquitination assays using purified recombinant proteins. As shown in Figure 2D, we observed significant synthesis of polyubiquitin chains only when all the subunits were combined in the presence of ATP, confirming that the Cullin2-PRAME complex is a catalytically active ubiquitin ligase.

Taken together, these results indicate that PRAME's BC- and Cul2-boxes are essential for the assembly of a PRAME-containing Elongin BC ubiquitin ligase, and they identify PRAME as a BC-box substrate recognition subunit of a Cul2-based ubiquitin ligase (Figure 2E).

PRAME associates with chromatin

Cell fractionation and cesium chloride gradient purification of crosslinked chromatin from K562 cells showed that PRAME associates with chromatin (data not shown). In addition, we observed that PRAME can be recruited to DNA *in vitro*. Consistent with the absence of recognizable DNA-binding domains in PRAME and other PRAME complex subunits, purified PRAME complexes did not bind directly to DNA immobilized on agarose beads. When incubated with bead-bound DNA in the presence of nuclear extract, however, PRAME complexes were retained on the DNA beads, suggesting that activity(s) in the extract are needed to bridge PRAME's interaction with DNA (Supplementary Figure S1A).

To investigate the chromatin-binding profile of PRAME, we performed ChIP assays followed by deep sequencing (ChIP-seq) using anti-PRAME antibodies or preimmune serum as control for the specificity of the signals obtained. Overlapping sequence tags were joined and displayed as tag-density profiles in the UCSC genome browser. Strikingly, inspection of the profiles revealed that the promoter regions of many protein-coding genes had specific enrichment for PRAME, as exemplified by the screenshots in Figure 3A. Validation experiments by ChIP-qPCR confirmed the high quality of the sequencing data (Supplementary Figure S1B).

Table I PRAME-associated ubiquitin ligase subunits identified by mass spectrometry of purified complexes from K562 cells

Protein name	Protein ID	TAG-PRAME (emPAI)	TAG-PRAME (pept.)	Control (pept.)
Elongin C	IPI00300341.5	4.6	6	1
Elongin B	IPI00026670.3	4.2	5	0
Rbx1	IPI00003386.3	0.8	1	0
PRAME	IPI00019282.1	0.7	6	0
CUL2	IPI00014311.4	0.7	12	0

Pept., number of peptides.

Table II PRAME-associated ubiquitin ligase subunits identified by mass spectrometry of purified complexes from HEK 293 cells expressing epitope-tagged wild-type or mutant PRAME proteins

Protein name	Gen bank acc. no.	PRAME WT				PRAME BC-box mutant				PRAME Cul2-box mutant				Control
		Pept.	SpC	Cov. (%)	NSAF × 1000	Pept.	SpC	Cov. (%)	NSAF × 1000	Pept.	SpC	Cov. (%)	NSAF × 1000	
PRAME	NP_996839	11	255	24.9	22	11	275	23	29.5	15	267	27.1	17.3	0
Elongin B	NP_009039	9	235	48.3	87.6	2	3	21.2	1.4	4	26	31.4	7.3	1
Elongin C	NP_005639	6	160	50.9	62.8	0	0	0	0	4	13	38.4	3.8	0
CUL2	NP_003582	14	455	18.4	26.9	0	0	0	0	0	0	0	0	0
NEDD8	NP_006147	2	9	17.3	4.9	0	0	0	0	0	0	0	0	0

Cov., sequence coverage; NSAF, normalized spectral abundance factor; Pept., number of peptides; SpC., spectral counts.

Figure 2 PRAME is a BC-box substrate receptor for Cullin2-EloBC ligases. **(A)** Multiple sequence alignment of known Elongin BC-associating proteins with PRAME. Sequences were aligned using the blosum62 mt2 alignment matrix with the AlignX program of the Vector NTI software package. Colour coding is based on the AlignX default similarity table. Identical amino acids are highlighted in yellow, very similar ones in blue, and similar amino acids are shown in green. The canonical BC-box and Cul2-box sequences are indicated below the alignment. **(B)** PRAME has a functional BC-box. SF21 cells were co-infected with the baculoviruses as indicated. Total cell lysates and anti-FLAG immunoprecipitates were analysed by western blotting. IP, immunoprecipitation; mBC, BC-box mutant. **(C)** BC-box defective PRAME lacks interaction with endogenous Cul2-EloBC ligases. Immunoblot analysis of 12CA5-HA immunoprecipitates from nuclear extracts of K562 cells stably expressing TAG-PRAME wild-type or TAG-PRAME BC-box mutant (PRAME mBC). Asterisks indicate antibody heavy chains. **(D)** The Cul2-PRAME ligase has *in vitro* ubiquitination activity. Epitope-tagged subassemblies of the PRAME Cul2-based ubiquitin ligase complex were purified from insect cells and used in an *in vitro* ubiquitin assay (as described in Materials and methods). The results were analysed by western blot using rabbit anti-FLAG and mouse anti-ubiquitin antibodies. **(E)** Model summarizing Cul2-EloBC-PRAME interactions.

Furthermore, ChIP-seq with stably expressed, epitope-tagged PRAME resulted in very similar if not identical patterns of PRAME enrichment (data not shown). To substantiate our finding, we used the MACS algorithm (Zhang *et al*, 2008) at a false discovery rate of 10^{-6} and identified a total of 14 132 PRAME-binding sites. Location analysis showed that 43% of the PRAME sites are closer than 1 kb upstream or downstream of a transcription start site (TSS) (Figure 3B). Importantly, promoters were nine-fold enriched as compared with the background distribution of genomic DNA.

Recently, Heintzman *et al* (2009) identified enhancer elements in K562 cells based on ChIP-on-chip histone modification signatures and binding of the enhancer-associated acetyltransferase p300. Most enhancers were cell-type specific, and they were enriched in genomic domains containing genes specifically expressed in K562 cells. When we compared the binding profile of PRAME to the predicted enhancers, we found a significant overlap between promoter-distal PRAME-binding sites and enhancer regions: 44% of the distal PRAME sites overlapped predicted enhancers, whereas only

4.1% of random genomic sites did (11-fold enrichment). Moreover, similarly to the annotated enhancers, the distal PRAME sites were not randomly distributed in the genome, but clustered in the vicinity of PRAME-bound promoters (Supplementary Figure S1C), suggesting possible promoter-enhancer interactions.

PRAME associates with NFY promoters

We next analysed all PRAME-binding sites falling within 1 kb of promoters. When we plotted their distance to the closest TSS, we found that 83% of promoter-proximal sites bound by PRAME are concentrated in a narrow window between -400 bp and +100 bp from TSS, peaking around -87 bp relative to the TSS (Figure 3C).

Promoters are characterized by the presence of transcription factor binding sites that critically contribute to the regulation of gene expression. Since neither PRAME, nor other PRAME complex subunits contain recognizable DNA-binding domains, we reasoned that other transcription factor(s) might mediate the recruitment of PRAME to these

Figure 3 Genome-wide profiling of PRAME association with chromatin. **(A)** Overview of PRAME-binding sites. UCSC genome browser views of ChIP-seq profiles generated with preimmune serum (negative control) and PRAME antibodies in K562 cells. Specific PRAME peaks are detected at the promoters of the *DNM2*, *EIF2AK3*, *TOP2A* and *MYST2* genes. **(B)** Genomic distribution of PRAME-binding sites (left) compared with the distribution of genomic DNA (right). The majority of sites (43%) are located within ± 1 kb from a TSS. **(C)** Promoter binding of PRAME occurs at the TSS. Frequency plot of the distances of PRAME-binding sites from TSS shows that 83% of PRAME promoter sites are located between -400 bp and $+100$ bp, with a peak at -87 bp. **(D)** Motif analysis identifies a CCAAT motif under PRAME promoter peaks. Left, logo of the Jaspas motif matching the most enriched sequence identified by GimmeMotifs. Right, the positional preference plot represents the frequency of the locations of the CCAAT motifs with respect to the summits of PRAME promoter summits, and indicate that the CCAAT motif is enriched at the centre of PRAME-binding regions.

regions. We therefore performed *de novo* motif discovery with GimmeMotifs (van Heeringen and Veenstra, 2011), using the DNA sequences underlying the promoter regions bound by PRAME. This analysis identified a CCAAT sequence as the top-scoring motif, which stood out with highly significant *P*-values ($P < 10^{-20}$) and the highest enrichment and MNCP scores as compared with a pool of random promoters (Figure 3D; Supplementary Figure S2). The CCAAT sequence can be bound by the NFY complex, a heterotrimeric transcription factor ubiquitously expressed in cycling cells (Maity and de Crombrughe, 1998). The NFYA subunit harbours the DNA-binding domain that recognizes the CCAAT sequence, while NFYB and NFYC are characterized

by histone-like folds (Maity and de Crombrughe, 1998). Similar *de novo* motif analysis performed on the distal PRAME-binding sites identified two significant motifs: a GATA/Evi1 motif, and a Fos motif.

ChIP-seq experiments with antibodies against the NFYA and NFYB subunits revealed that PRAME sites at promoters are indeed bound by NFY *in vivo*. Screenshots of the sequencing data show clear enrichment of NFY at the representative PRAME sites *DNM2* and *EIF2AK3*, as well as PRAME enrichment at the classical NFY promoters *TOP2A* and *CDC25C* (Figure 4A). Consecutive ChIP experiments (ChIP-re-ChIP) using crosslinked NFYA antibodies detected PRAME and NFY concomitantly on the same regions, indicating that these

Figure 4 PRAME associates with NFY-bound promoters genome wide. (A) Overview of NFY binding to PRAME loci. UCSC genome browser views show clear binding of PRAME, NFYA and NFYB at the promoters of the *DNM2*, *EIF2AK3*, *TOP2A* and *MYST2* genes. (B) Sequential ChIP (ChIP-re-ChIP) experiments show co-occupancy of PRAME and NFY. The first ChIP was performed with covalently crosslinked NFYA beads, followed by either anti-PRAME serum (NFYA-PRAME) or preimmune serum as negative control (NFYA-preimmune). (C) Promoter binding of NFY and PRAME occurs at the TSS. Frequency plot as in Figure 3C shows very similar distributions of distances from TSS for NFY and PRAME. (D) PRAME associates with a subset of NFY-promoters genome wide. Left, Venn diagram of the overlap between NFY-bound (green) and PRAME-bound (blue) promoters shows that PRAME promoters overlap almost completely with NFY promoters. On the contrary, a minority of NFY promoters are not bound by PRAME. An example is the *TSPAN15* gene (right).

proteins are simultaneously localized at the same promoter regions (Figure 4B).

Using the MACS algorithm, we identified a total of 28 844 regions that were bound by NFYA and/or NFYB. The distance distributions of both NFYA- and NFYB-binding sites from TSSs showed a distinct peak around $-78/-84$ bp very similar to PRAME (Figure 4C). We next counted the number of sequence tags within NFY-binding sites and performed a regression curve analysis (Supplementary Figure S3). The binding of NFYA and NFYB were very closely correlated at the majority of the sites, as expected for stoichiometric binding and in agreement with *in vitro* experiments, showing that all subunits are required for efficient DNA binding (Sinha *et al*, 1995).

Globally, 71% of all PRAME-binding sites overlapped NFY regions. Importantly, virtually all PRAME-bound promoters were also bound by NFY, corroborating and extending the

relevance of the CCAAT motif (Figure 4D). In contrast, a sizeable number of NFY promoters did not contain PRAME peaks (Figure 4D). We performed functional annotation of the associated genes and found that NFY-only genes as well as NFY genes with PRAME enrichment are associated with metabolic and cancer-related pathways. Furthermore, NFY-only genes are specifically enriched for immunological processes and intercellular junctions, while NFY genes with PRAME binding are significantly associated with numerous transcription-related pathways and cell cycle regulation (Supplementary Table SI).

PRAME is specifically recruited to transcriptionally active NFY promoters

Post-translational modifications of the histone proteins, like lysine methylations and acetylations, are implicated in the regulation of gene transcription. Therefore, we used histone

marks to assess the transcriptional status of PRAME loci. For this purpose we processed ChIP-seq data sets generated by the ENCODE project for K562 cells (see Materials and methods), and we quantitated the signals at all NFY promoters. We found that the active histone marks H3K9ac and H3K4me3 showed a strong correlation with the binding of PRAME (Figure 5A). In contrast, NFY-only promoters had near-background levels of these marks.

To investigate the transcriptional status of the associated genes, we quantitated the occupancy in gene bodies of RNA Polymerase II and of the histone marks H3K36me3, which has been linked to transcriptional elongation (Krogan *et al*, 2003), and H3K27me3, which has been linked to repression (Barski *et al*, 2007). As shown in Figure 5B, NFY genes without PRAME binding displayed low levels of PolII and H3K36me3, and high levels of the repressive H3K27me3 mark. In contrast, PRAME-bound genes had significantly higher occupancy of both PolII and H3K36me3 (Figure 5B and D; Supplementary Figure S4), suggesting that PRAME is associated specifically with transcriptionally active genes. To substantiate these findings, we analysed gene expression levels by RNA-seq in our K562 cells. In agreement with the ChIP-seq data, we detected few, if any, transcripts from NFY-only genes, while PRAME-bound NFY genes were expressed at high levels (Figure 5C).

We next tested if components of the Cullin2 complex are present together with PRAME on chromatin. In the absence of ChIP-grade antibodies, we established stable K562 cell lines expressing Elongin C tagged with a double TY1 epitope (TTE-EloC), for which a ChIP-grade antibody was validated in our laboratory. ChIP-qPCR experiments showed specific binding of EloC at all the PRAME promoters tested (Figure 5E). As a control, we performed similar experiments with TTE-PRAME.

Taken together, our data support a model in which a Cul2/EloBC/PRAME complex is specifically bound to the subset of epigenetically and transcriptionally active NFY promoters and nearby enhancer regions.

Discussion

Our proteomic data reveal that the human oncoprotein PRAME is a BC-box substrate recognition subunit of a Cul2-based E3 ubiquitin ligase, suggesting a functional link between ubiquitination and the oncogenic properties of PRAME. Notably, the Von Hippel-Lindau tumour suppressor protein is one of the best-characterized BC-box proteins, and its tumour suppressor activity is due in part to its role as the substrate recognition subunit of a Cul2-based ubiquitin ligase (Kaelin, 2007). Intriguingly, all the annotated PRAME-like proteins display a complete conservation of the BC-box and Cul2-box at their N-terminus (Supplementary Figure S5), suggesting that all members of the large PRAME protein family might function as substrate recognition subunits of Cul2-based ubiquitin ligases, possibly displaying different degrees of specificity and/or redundancy in substrate selection.

To date, ubiquitination has been demonstrated to have important roles in gene regulation, affecting several different steps in the transcriptional process (Conaway *et al*, 2002; Collins and Tansey, 2006). Many activators are subject to ubiquitin-mediated degradation at a rate that parallels their

transactivation abilities (Salghetti *et al*, 2001). Histones, transcription factors and RNA polymerase II (RNAPII) are targets of ubiquitination, and ubiquitination events are required for the initiation-competent RNAPII complex to exchange factors and become competent for cotranscriptional mRNA processing (Muratani *et al*, 2005). A recent study indicated that the stability of NFYA can be modulated by ubiquitination and degradation via the proteasome, and that degradation-resistant NFYA mutants increased cell proliferation (Manni *et al*, 2008). Interestingly, some of the lysine residues in NFYA that are targets of ubiquitination could also be acetylated by p300, suggesting a regulatory competition between these post-translational modifications (Manni *et al*, 2008).

Our genome-wide analyses reveal that PRAME is strongly enriched at NFY promoters, and binding of PRAME predicts a permissive epigenetic landscape and transcriptional activity at these regions. Moreover, we found that PRAME is also significantly enriched at adjacent enhancer regions, suggesting that it could participate in promoter-enhancer interactions. Recently, Ceribelli *et al* (2008) reported a ChIP-on-chip analysis of NFYB in HeLa cells with 2.2% coverage of the human genome. Similar to our K562 data, the majority of the NFY-bound promoters identified in HeLa cells were transcriptionally active and carried H3K9/14ac and/or H3K4me3 marks, while about 27% of NFY-bound promoters were devoid of these active marks and mostly silent. Further experiments implicated NFY not only in the expression of active genes, but also in the repression of silent ones, revealing an unexpected dual nature of the NFY complex. The basis of this dual functionality, however, is still unclear, and genome-wide data suggest that the activity and inactivity of NFY promoters is determined by additional factors and mechanisms that are yet to be identified.

PRAME is the first factor reported to discriminate between active and inactive NFY loci. Our genome-wide analyses, summarized in Figure 5F, reveal that PRAME is specifically recruited to NFY loci associated with the promoters of genes that are highly transcribed and that have high levels of H3K9ac and H3K4me3 at their promoters and high H3K36me3 but low H3K27me3 levels in their gene bodies. These data suggest a role for PRAME in transcriptional regulation, and possibly in the switch between activity and inactivity of NFY loci. It is tempting to speculate that PRAME might mediate ubiquitination, and possibly degradation, of one or more of the factors that are involved in the regulation of NFY promoters. Despite evidence that PRAME and NFYA can co-occupy the same genetic loci (Figure 4B), we have not been able to detect stable and direct interactions between PRAME and NFY subunits by co-immunoprecipitation following their expression in transfected mammalian cells or in baculovirus-infected insect cells (data not shown). We have also observed that PRAME can be recruited *in vitro* to a DNA fragment containing the NFY-dependent *HSPA5* promoter by a mechanism that depends on activity(s) present in a nuclear extract. It seems likely that PRAME-NFY interactions, if they exist, occur only on DNA and depend on additional, yet to be identified factors, since (i) purified NFY is not sufficient to recruit PRAME to *HSPA5* promoter DNA and (ii) supplementing nuclear extract with exogenously added NFY does not further enhance PRAME recruitment (Supplementary Figure S1A).

Figure 5 PRAME discriminates between active and inactive NFY loci. **(A)** PRAME binding positively correlates with H3K9ac and H3K4me3 modifications at NFY promoters. Left, heat map representing median tag densities (tags/kb) of H3K9ac and H3K4me3 at promoter-associated NFY-binding sites without ('No PRAME') or with ('PRAME') PRAME binding. Higher levels of both histone marks are found if PRAME is present. Right, UCSC genome browser views of two representative genes: *TSPAN15* and *PHB*. **(B)** PRAME binding of NFY promoters correlates with high H3K36me3 and RNA-PolIII, and with low H3K27me3 in the gene body. Box plots show the tag density (tags/kb) over the bodies of the genes associated with NFY promoters without or with PRAME binding. **(C)** PRAME binding correlates with high levels of gene expression. Box plot of normalized expression values computed from RNA-seq experiments show that PRAME-bound NFY genes are expressed at higher levels than those without PRAME. **(D)** Genome browser view of the genome-wide data combined. Two neighbouring tail-to-tail NFY genes are shown: *SPRYD4* is bound by PRAME and is active, while *GLS2* is not bound by PRAME and is in a repressed state. **(E)** Elongin C associates with PRAME promoters. ChIP-qPCR assays were performed using BB2 antibody on the cell lines indicated. Wild-type and empty vector (K562-TTE) cell lines were used as negative controls; K562-TTE-PRAME as positive control. **(F)** Model summarizing the genome-wide data.

Interactions have been reported between NFY and the histone acetyltransferases GCN5 and PCAF (Currie, 1998), and single-gene studies showed that NFY can activate at least some of its target genes in cooperation with the acetyltransferase p300, which is able to acetylate histone tails as well as NFY subunits (Li *et al*, 1998; Caretti *et al*, 2003; Salsi *et al*, 2003; Gurtner *et al*, 2008). While these acetyltransferases constitute potential targets of PRAME, other unknown factors could also take part to these mechanisms. Notably, p300 is not only recruited to promoters, but it is also a hallmark of enhancer elements. Heintzman *et al* (2007) used p300 ChIP-on-chip binding profiles to identify an enhancer-specific chromatin signature, which was successfully used to predict enhancer elements in several cell lines, including K562. Significantly, we found that about half of the promoter-distal PRAME sites are localized at predicted enhancers, suggesting that PRAME might be implicated in promoter–enhancer crosstalk.

In an effort to gain insight into PRAME's potential role in transcription regulation, we knocked down *PRAME* with retroviral shRNA vectors and measured the mRNA levels of several target genes. Despite a 70% decrease of *PRAME* transcripts, we did not observe reproducible changes in the transcript levels of the target genes tested (data not shown). This is perhaps not surprising given that PRAME is a member of a large family of closely related proteins that could have overlapping or redundant functions in cultured cells. Further, the fact that the PRAME gene family underwent an extraordinary expansion specifically in rodents and primates (Birtle *et al*, 2005; Consortium *et al*, 2007) suggests that they are more likely to fine-tune transcriptional regulatory processes rather than to be essential for transcription. Regrettably, the lack of a one-to-one orthology between the human and mouse PRAME genes (Birtle *et al*, 2005) has thus far precluded the development of suitable knockout mouse models for studying PRAME function.

Notably, it is becoming clear that many proteins associate with the promoter regions of genes, but their depletion does not seem to cause significant or widespread transcriptional defects. For example, the yeast Set1 histone methyl transferase is widely associated with active promoters, but deletion of the Set1 gene causes expression changes of 1.5-fold or more in only about 20 genes (Miller *et al*, 2001). Similarly, the Jmjd3 histone demethylase was recently shown to associate with the promoters of LPS-induced genes, but the expression of most of them was unaffected by knockout of Jmjd3 (De Santa *et al*, 2009). These observations are consistent with the model that some factors are not strictly required for the bulk of transcription, but instead have ancillary or regulatory functions that become apparent only in very specific conditions or subpopulations of cells.

While numerous reports established a role for PRAME in human cancers, several studies indicated that PRAME-like genes have roles in the early stages of spermatogenesis (Wang *et al*, 2001) and oogenesis (Dadé *et al*, 2003), as well as in embryonic development and embryonic stem cells (Bortvin *et al*, 2003). Recently, overexpression of *Pramel7* in E14 embryonic stem cells was reported to maintain a pluripotent state in the absence of the anti-differentiation factor LIF (Cinelli *et al*, 2008). These findings are consistent with the notion that stem cells and neoplastic tissues share many properties, and several oncogenic

pathways were recently found to also regulate self-renewal mechanisms in stem cells (Reya *et al*, 2001). Interestingly, recent studies suggest important roles for NFY in stem cell biology. Knockout mice demonstrated that NFYA is essential for early mouse development (Bhattacharya *et al*, 2003), and NFYA was found to promote the self-renewal of haematopoietic stem cells (Zhu *et al*, 2005). Moreover, a systematic analysis of *cis*-regulatory elements identified the CCAAT box and NFY as crucial regulators of both mouse and human stem cells-specific genes, and NFY was involved in the maintenance of the proliferative capacity of ES cells (Grskovic *et al*, 2007).

Taken together, our data unambiguously show that the PRAME oncoprotein is part of a Cullin2-based ubiquitin ligase complex and that PRAME localizes to active NFY promoters, paving the road for understanding the functions of PRAME in normal cells and in human malignancies. Future experiments will address the role of the PRAME ubiquitin ligase at TSSs, its role in transcription of NFY-regulated genes, and its contribution to the oncogenic properties of PRAME.

Materials and methods

Cell culture and stable cell lines

HEK293 cells were cultured in DMEM and K562 in RPMI medium (Gibco, Invitrogen) at 37°C in 5% CO₂. Both media were supplemented with 5% Glutamax, 10% fetal bovine serum, 100 U/ml of penicillin, 100 µg/ml of streptomycin (Gibco, Invitrogen). Stable HEK293–HA–PRAME cells were generated using pcDNA5/FRT/HA–PRAME and the Flp-in system (Invitrogen). K562 stable cell lines were generated with retroviral constructs according to the following 5 days protocol. On the first day, HEK293T cells at 50–60% confluence in a 6-cm dish were transfected using Lipofectamine2000 (Invitrogen) and 1.8 µg of MD-MLV, 300 ng VSV-G, 3.6 µg retroviral construct, 100 ng pGFP. After 24 h (day 2), the virus-containing supernatant was harvested, centrifuged for 10 min at 3000 r.p.m. and used for a first round of transduction of 10⁶ K562 cells in the presence of 8 µg/ml of Polybrene (Sigma) for 24 h. A second round of transduction was performed on day 3, and on day 4, the target cells were transferred to fresh RPMI and selection was started on day 5 with either 6–10 µg/ml of puromycin (Sigma) or 1–2 mg/ml Zeocin (Invitrogen).

Antibodies and western blot

PRAME antisera and affinity-purified antibodies were generated by immunizing rabbits with the previously described peptide FPE-PEAAQPMTKKRKVDG (Epping *et al*, 2005) (V27 serum). Mouse monoclonal HA-12CA5 and Myc-9E11 were produced in-house from hybridoma cultures. Commercial antibodies used were mouse monoclonal FLAG-M2 (Sigma), c-Myc monoclonal (Roche Applied Science), rabbit HA (Bethyl Laboratories, or Abcam ab9110), mouse monoclonal HA.11 (Covance MMS-101P), rabbit Cul2 (Zymed Laboratories 51-1890), mouse ElonginC (BD Transduction Laboratories 610760), goat ElonginB (Santa Cruz P-16, sc-23407), α -tubulin (Santa Cruz B-7, sc-5286), HDAC2 (Biomol SA-402).

Proteins were separated by conventional SDS–PAGE or NuPAGE 4–12% Bis–Tris gels (Invitrogen) run with MES buffer. Western blots were visualized by ECL (GE Healthcare) or Odyssey (LiCor).

Plasmids and cloning

Full-length human PRAME was PCR amplified and cloned into *NotI* and *XhoI* sites of pcDNA5/FRT/TO/StrepII-2 × TEV–Myc-3 × HA (Le Guezennec *et al*, 2006) to generate pcDNA5–TAG–PRAME. A *BamHI*–*XhoI* fragment encoding TAG–PRAME was further subcloned into LZRS(Zeo) to generate the retroviral expression vector LZRS(Zeo)–TAG–PRAME. Full-length human PRAME (ATCC) was subcloned with N-terminal HA epitope tags into pcDNA5/FRT (Invitrogen). Mutagenesis to generate PRAME mutants was performed using the QuikChange II XL site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions. To create the pTTE retroviral vector for ChIP, the ProtA-2 × TEV–Myc

cassette of the pZXN plasmid (Le Guezennec *et al*, 2006) was removed by digestion with *Bam*HI and *Sal*I. The resulting backbone was ligated to overlapping synthetic oligos with *Bam*HI and *Xho*I overhangs, encoding two consecutive TY1 epitopes, a single-ERalpha epitope and an extended MCS. PRAME was cloned using the *Eco*RI site to generate pTTE-PRAME plasmid; EloC was cloned using *Bgl*II and *Xho*I to generate pTTE-EloC plasmid. All constructs were checked by sequencing.

Gel filtration analysis

Size exclusion chromatography was performed with a Superose 6 column on a SMART system (Amersham Pharmacia Biotech) with buffer containing 300 mM NaCl, 10% glycerol, 20 mM TRIS, 0.2 mM EDTA, 1 mM DTT and 1 mM PMSF. In all, 50 µl nuclear extract was separated in 100 µl fractions and analysed by SDS-PAGE and western blot.

Purification of protein complexes from K562 cells

K562 cells were grown in spinner flasks and harvested at a confluence of about 10⁶ cells/ml. For each protein extraction, about 10¹⁰ cells were centrifuged for 15 min at 1600 r.p.m. and washed with PBS. Cell pellets were resuspended in 2.5 volumes of hypotonic buffer T1 (10 mM Hepes pH 7.9, 10 mM KCl, 0.1 mM MgCl₂, 0.1 mM EDTA pH 8, 1 mM DTT, 10% glycerol) supplemented with phosphatase and protease inhibitors (1 mM NaVO₄, 1 mM NaF, 1 mM PMSF, Protease inhibitor cocktail (Roche)). After 15 min of incubation on ice, the cell membranes were disrupted with 20 strokes of Pestle A, or until 80–90% of the cells were stained by trypan blue. After centrifugation for 20 min at 3000 r.p.m. at 4°C, the supernatant was harvested and regarded as the cytoplasmic fraction. The pellet containing the nuclei was resuspended in 1 volume of T2 buffer (20 mM Hepes pH 7.9, 420 mM NaCl, 1.5 mM MgCl₂, 0.1 mM EDTA pH 8, 1 mM DTT, 10% glycerol) supplemented with inhibitors, and incubated in rotation for 1 h at 4°C to extract the nuclear proteins. The suspension was cleared by ultracentrifugation for 30 min at 100 000 g at 4°C (Sorvall RC M150 GX). The supernatant was the nuclear fraction. The quality of the cytoplasmic/nuclear extractions was determined by western blot of α-tubulin (cytoplasmic marker) and HDAC2 (nuclear marker).

About 5–10 ml (60–120 mg) of nuclear extracts were incubated overnight at 4°C with anti-Myc (9E11) antibodies covalently cross-linked to protA-sepharose CL-4B (GE Healthcare) (see below). The following day, beads were washed three times with 10 ml buffer IPP300/0.5 (300 mM NaCl, 10 mM Tris-HCl pH 8, 0.5 mM EDTA and 0.5% NP-40), and the protein complexes were eluted with 20 µg/ml of Myc peptide in a thermoshaker for 30 min at 28°C.

Crosslinking of antibody to beads

ProtA-sepharose CL-4B (GE Healthcare) or protG-sepharose (Sigma) beads were incubated with antibody in PBS at room temperature for 1–2 h, then washed once with 10 volumes of sodium borate solution (0.2 M, pH 9). Chemical crosslinking of the antibodies to the beads was performed resuspending the beads in a sodium borate solution containing 5.2 mg/ml DMP (dimethyl-pimelimidate; Sigma) and incubating for 30 min at room temperature with gentle rotation. The beads were washed once in 10 volumes ethanolamine (0.2 M, pH 8), incubated with ethanolamine for 2 h at room temperature, washed twice with glycine (0.1 M, pH 3) to remove uncrosslinked antibodies, and eventually with PBS to neutralize the pH.

Mass spectrometry and emPAI calculation for K562 complexes

The eluates of large-scale immunoprecipitations from K562 cells were concentrated by speedvac and run shortly (~1 cm) on SDS-PAGE gels to remove detergents and the excess of peptide. The gel lane was then fixed, cut in small pieces, reduced and alkylated. Proteins were digested overnight with trypsin (Promega) and eluted from the gel with trifluoroacetic acid. Peptides were sequenced using a nano-high-pressure liquid chromatography Agilent 1100 nanoflow system connected online to a 7-T linear quadrupole ion trap-Fourier transform mass spectrometer (Thermo Electron, Bremen, Germany) essentially as described (Olsen *et al*, 2004). Proteins were identified using the Mascot search algorithm (Matrix Science) with IPI human database v3.37, which contains 69164 unique entries. First ranked peptides (Mascot peptide scores ≥20) were parsed from Mascot database search html files with MSQuant (<http://www.msquant.sourceforge.net>) to generate unique first

ranked peptide lists that were further filtered for absolute calibrated mass error <5 and peptide δ score >5. To remove redundancy at the protein level and to uniquely assign peptides to one IPI Db entry, the peptides were remapped to the IPI human Db using the program Protein Coverage Summarizer (<http://ncrr.pnl.gov/software/>) as described (Lasonder *et al*, 2008). To determine the protein abundance, we computed the exponentially modified Protein Abundance Index (emPAI) values for our data (Ishihama *et al*, 2005; Lasonder *et al*, 2008); emPAI values for all proteins were calculated as 10^{PAI-1} (PAI = observed peptides/observable peptides). The number of observable peptides was calculated by *in silico* tryptic digestion of the IPI human Db v3.37 using the program Protein Digestion Simulator (<http://ncrr.pnl.gov/software/>).

Purification of protein complexes from 293 cells and MudPIT analysis

To prepare protein samples for mass spectrometry, 293-HA-PRAME and control wild-type cells were grown to 70–80% confluence in 10–20 roller bottles then washed, lysed and subjected to anti-HA agarose immunoaffinity chromatography as described (Mahrouir *et al*, 2008). TCA-precipitated proteins were urea-denatured, reduced, alkylated and digested with endoproteinase Lys-C (Roche) followed by modified trypsin (Roche) as described (Washburn *et al*, 2001; Florens and Washburn, 2006). Peptide mixtures were loaded onto 100 µm fused silica microcapillary columns packed with 5 µm C₁₈ reverse phase (Aqua, Phenomenex), strong cation exchange particles (Partisphere SCX, Whatman), and reverse phase. Loaded microcapillary columns were placed in-line with a Quaternary Agilent 1100 series HPLC pump and a LTQ ion trap mass spectrometer equipped with a nano-LC electrospray ionization source (ThermoFinnigan). Fully automated 10-step MudPIT runs were carried out on the electrosprayed peptides, as described (Florens and Washburn, 2006). Tandem mass (MS/MS) spectra were interpreted using SEQUEST against a database of 61 430 sequences, consisting of 37 742 human proteins (downloaded from NCBI on 2008-03-04), 177 usual contaminants (such as human keratins, IgGs and proteolytic enzymes), and to estimate false discovery rates, 30 716 randomized amino-acid sequences derived from each non-redundant protein entry. Peptide/spectrum matches were sorted and selected using DTASelect with the following criteria set: spectra/peptide matches were only retained if they had a DeltCn of at least 0.08, and minimum XCorr of 1.8 for singly, 2.0 for doubly and 3.0 for triply charged spectra. In addition, peptides had to be fully tryptic and at least 7 amino acids long. Combining all runs, proteins had to be detected by at least two such peptides, or one peptide with two independent spectra. Under these criteria, the final false discovery rate at the peptide level is 0.151 ± 0.009. Peptide hits from multiple runs were compared using CONTRAST. To estimate relative protein levels, we took advantage of evidence that normalized spectral abundance factors (NSAFs) vary with the relative abundance of proteins in a sample. NSAFs were calculated for each detected protein, using the formula:

$$(\text{NSAF})_k = \frac{(\text{SpC}/\text{Length})_k}{\sum_{i=1}^N (\text{SpC}/\text{Length})_i}$$

where SpC = spectral count, Length = protein length in amino acids and *i* = all proteins detected in the MudPIT runs, as described (Florens *et al*, 2006; Paoletti *et al*, 2006).

Expression of recombinant proteins in Sf21 insect cells

cDNAs encoding wild-type or mutant PRAME, Cul2, Elongins B and C were subcloned into pBacPAK8 and recombinant baculoviruses were generated. Sf21 cells were cultured at 27°C in SF-900 medium supplemented with 10% fetal bovine serum. Sf21 cells were co-infected with the recombinant baculoviruses indicated in the legend to Figure 2. At 50 h after infection, cells were collected, lysed and processed as described (Mahrouir *et al*, 2008).

In vitro ubiquitination assays

The PRAME-Elongin BC complex and the His-T7-Xpress-Cullin2/MyC-RBX1 complex were obtained by co-infecting Sf21 cells with baculoviruses encoding for each individual protein and purifying through Flag tag or nickel-agarose chromatography, respectively. Different combinations of these complexes were then incubated for

1 h at 37°C with ~50 nM Uba1, ~0.2 µM UbcH5a and 0.25 mg/ml ubiquitin, in 20 µl reactions containing 50 mM Tris-HCl pH 7.6, 50 mM NaCl, 5 mM MgCl₂, 0.5 mM EDTA (pH 7.9), 5% glycerol, 2 mM ATP, 0.5 mM dithiothreitol, 0.3 units/ml pyrophosphatase, 60 mM creatine phosphate and 0.3 mg/ml creatine phosphokinase. N-terminally (His)6-tagged mouse E1 was expressed in Sf21 insect cells and purified by nickel-agarose chromatography as described. Human UbcH5a with an N-terminal 6-histidine tag and a C-terminal Flag tag were expressed in *Escherichia coli* strain BL21(DE3) and purified by nickel agarose.

ChIP and qPCR

Chromatin harvests, ChIPs and qPCR analyses were performed as previously described (Denissov *et al*, 2007), with the following minor adjustments. Cells were crosslinked for 15 min in 1% formaldehyde, they were resuspended in ChIP incubation buffer at a concentration of 16 millions/ml and sonicated using a Bioruptor sonicator (Diagenode) for 15 min with settings: high power, 30 s ON, 30 s OFF. For every ChIP reaction, 200 µl of sonicated chromatin were used. The immunoprecipitated chromatin was eluted with 200 µl of elution buffer at room temperature for 20 min and DNA was purified with QIAquick PCR purification kit (Qiagen). For ChIP-seq, DNA eluted from 5 to 7 ChIP reactions was pooled and the DNA concentration was measured with a Qubit fluorometer and Quant-iT dsDNA HS Assay (Invitrogen).

The following antibodies were used for ChIP: preimmune serum (preV27, 'no antibody' control) and PRAME antiserum (V27); NFYA (Santa Cruz H-209, sc-10779); NFYB (pAb-007-100) and BB2 (against TY1 epitope) were from Diagenode.

For sequential ChIP experiments (ChIP-re-ChIP), in order to avoid antibody carry-over, the first ChIP was performed with antibody covalently crosslinked to ProteinA/G-Sepharose beads (Santa Cruz) with DMS as described above. Immunoprecipitated chromatin was then eluted in a smaller volume and further processed as a normal ChIP with PRAME immune serum or the preimmune serum as negative control. Real-time qPCR was performed using the SYBR Green mix (Bio-Rad) with the MyIQ thermocycler (Bio-Rad). Specific primers for qPCR (Biolegio) were designed as previously described (Denissov *et al*, 2007) and are listed in Supplementary Table SII. Primers amplifying exon 12 of *albumin* were used as negative control.

RNA-seq library preparation and data analysis

Total RNA was extracted from K562 cells with the RNeasy kit and on-column DNase treatment (Qiagen). In all, 100 µg of total RNA were subjected to two rounds of poly(A) selection using the Oligotex mRNA Mini Kit (Qiagen) and the concentration of poly(A) + RNA was measured with a Qubit fluorometer (Invitrogen). In all, 200 ng of purified poly(A) + RNA were diluted in 160 µl RNase-free water and fragmented by addition of 40 µl 5 × fragmentation buffer (200 mM Tris acetate pH 8.2, 500 mM potassium acetate and 150 mM magnesium acetate) and incubation at 94°C for 120 s. After purification with RNeasy Minelute Kit (Qiagen), the quality of fragmented RNA was analysed on an Experion (Bio-Rad). Double-stranded cDNA was synthesized from the fragmented RNA with SuperscriptIII (Invitrogen) using random hexamers, and then used for Illumina sample prepping and sequencing (see below).

A total of 7 246 489 RNA-seq reads were uniquely mapped to hg18 and used for bioinformatical analysis. Normalized gene expression values (NE) for RefSeq genes were computed with Genomatix Region Miner tools.

Illumina high-throughput sequencing

A starting amount of 10–20 ng of ChIP DNA or double-stranded cDNA was used for sample prepping according to the standard protocols. Briefly, end repair was performed with T4 DNA ligase, Klenow DNA polymerase and T4 PNK; a 3' protruding adenosine was added using Klenow DNA polymerase, and T-overhang adaptors were ligated. DNA molecules with a size of 270–300 bp were selected from gel, purified and amplified with 14 cycles of PCR. Quality controls were performed by qPCR of DNA before and after sample prepping, and by analysis of the PCR products with an Experion (Bio-Rad). Cluster generation and sequencing-by-synthesis (35 bp) were performed using the Illumina Genome Analyser II according to standard protocols (Illumina). The image files were

processed to extract DNA sequence data, and sequences were mapped to the human hg18 reference genome using the eland program allowing at most one base mismatch. Sequence reads were filtered based on quality controls and sequences that could not be mapped uniquely on the reference genome were discarded. For visualization purposes, ChIP-seq reads were directionally extended to the length of the fragments used for sequencing and the number of overlapping sequence reads was determined for each basepair in the genome, averaged over a 10-bp window, and visualized in the University of California Santa Cruz genome browser (<http://genome.ucsc.edu>). ChIP-seq and RNA-seq data are published in the GEO database (GSE26439), and in the published history 'PRAME' in Galaxy (<http://main.g2.bx.psu.edu/>).

ChIP-seq analysis and peak detection

Bioinformatical analysis was performed using a combination of in-house developed scripts, a Genomatix Genome Analyser server (<http://www.genomatix.de>), and Galaxy tools (<http://main.g2.bx.psu.edu/>).

ChIP-seq experiments generated the following number of uniquely mapped reads: 5.5 millions for PRAME, 5.6 millions for preimmune (control), 8.9 millions for NFYA and 8.8 millions for NFYB. In order to compensate for the small difference in sequencing depth between the PRAME and preimmune data sets, the number of tags of preimmune ChIP-seq was equalized to the number of tags of PRAME ChIP-seq by uniform removal of tags as described (Nielsen *et al*, 2008). PRAME-binding sites (peaks) were subsequently identified with the MACS peak calling algorithm (Zhang *et al*, 2008) using the two-sample method with a *P*-value (*e*) of 10^{−6}, which compared the PRAME profile to the preimmune (control) profile. NFYA- and NFYB-binding sites were determined with MACS using the standard method with extended lambda (10 kb, 50 kb 100 kb) and *P*-value (*e*) of 10^{−6}. MACS identified a total of 14 132 PRAME peaks, 25 151 NFYA peaks and 16 350 NFYB peaks. NFYA and NFYB peaks were combined in a total of 28 844 NFY-binding sites that were tested for tag density. The number of sequence reads were counted in 500 bp regions centred on the summits of the 28 844 NFY-binding sites (with the summit being the site of highest enrichment as determined by MACS), and the data were visualized in dot plots using R.

Location analysis and motif discovery

To analyse the genomic distribution of the binding sites identified, we first determined the fraction of peak summits localized within a 2-kb window centred around the TSS of primary transcripts (TSS ± 1 kb) using the Genomatix Genome Inspector. The genomic distribution of the binding sites located outside promoter regions ('distal sites') relative to genes was analysed with PinkThing (<http://pinkthing.cmbi.ru.nl/>). The background distribution of genomic DNA was determined applying the same approach to pools of random genomic sites (*n* > 33 000).

The distances of promoter-associated binding sites from TSSs were computed with the Genomatix Genome Inspector tool, and plots were generated with R.

The localization of PRAME at enhancer elements was determined by overlaps between PRAME peaks summits and enhancer regions recently predicted for K562 cells (Heintzman *et al*, 2009). The genomic coordinates of the published enhancer regions were converted from hg17 to hg18 with the Galaxy Lift-over function, resulting in 24 545 enhancers.

De novo motif discovery was performed using the GimmeMotifs pipeline (van Heeringen and Veenstra, 2011), which incorporates a number of widely used motif analysis tools, including MDmodule, MEME, Weeder, MotifSampler and BioProspector. Based on the distribution of PRAME-binding sites, we performed motif searches on the DNA sequences underlying PRAME summits located within −400 bp and +100 bp of a primary TSS (*n* = 4981). A CCAAT motif scored the highest values of enrichment and MNCP scores, as compared with pools of background sequences with the same genomic distribution.

ENCODE data sets

We retrieved and analysed publicly available ChIP-seq data sets of H3K9ac, H3K4me3, H3K36me3, H3K27me3 and RNA-PolII generated from K562 cells by the 'ENCODE Chromatin Group at Broad

Institute and Massachusetts General Hospital' (<http://hgdownload.cse.ucsc.edu/goldenPath/hg18/encodeDCC/wgEncodeBroadChIPSeq/>).

Starting from 27 903 RefSeq genes with a RNA-seq expression value, we identified 7214 genes longer than 1 kb with a NFY peak summit in the proximal promoter region. Next, we quantified the signals of PRAME in a 500-bp window under the NFY-binding site, H3K9ac and H3K4me3 in a 3-kb window, H3K36me3 and H3K27 in the gene body and RNA-PolIII in the gene body excluding the first 500 bp.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

We are grateful to B Scheijen for plasmids and protocols, A Kaan for technical assistance with DNA cloning, S Hall for assistance with protein expression and purification, A Gottschalk for recombinant

NFY, E Janssen-Megens and KJ Francois for the Illumina sequencing. We are indebted to S van Heeringen, A Brinkman, H Marks and W Welboren for scripts and support with the bioinformatical analyses. We also thank our colleagues for helpful discussions and suggestions. Work in the authors' laboratories was funded by Dutch Cancer Society KWF (to HS), the Stowers Institute for Medical Research and grant R37 GM41628 (to RCC) from the National Institute for General Medical Sciences (NIGMS).

Authors contributions: AC and NM designed and performed the experiments and contributed equally to this study. ET, RS, MAS and DS performed the experiments. PWJ, SM-B, MPW and LF performed mass spectrometry analyses. JWC, RCC and HGS conceived strategies and supervised the project. AC drafted the manuscript, which was edited by NM, JWC, RCC and HGS.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129**: 823–837
- Bhattacharya A, Deng JM, Zhang Z, Behringer R, de Crombrughe B, Maity SN (2003) The B subunit of the CCAAT box binding transcription factor complex (CBF/NFY) is essential for early mouse development and cell proliferation. *Cancer Res* **63**: 8167–8172
- Birtle Z, Goodstadt L, Ponting C (2005) Duplication and positive selection among hominin-specific PRAME genes. *BMC Genomics* **6**: 120
- Bortvin A, Eggan K, Skaletsky H, Akutsu H, Berry DL, Yanagimachi R, Page DC, Jaenisch R (2003) Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. *Development* **130**: 1673–1680
- Caretti G, Salsi V, Vecchi C, Imbriano C, Mantovani R (2003) Dynamic recruitment of NF-Y and histone acetyltransferases on cell-cycle promoters. *J Biol Chem* **278**: 30435–30440
- Ceribelli M, Dolfini D, Merico D, Gatta R, Viganò AM, Pavesi G, Mantovani R (2008) The histone-like NF-Y is a bifunctional transcription factor. *Mol Cell Biol* **28**: 2047–2058
- Cinelli P, Casanova EA, Uhlig S, Lochmatter P, Matsuda T, Yokota T, Rüllicke T, Ledermann B, Bürki K (2008) Expression profiling in transgenic FVB/N embryonic stem cells overexpressing STAT3. *BMC Dev Biol* **8**: 57
- Collins GA, Tansey WP (2006) The proteasome: a utility tool for transcription? *Curr Opin Genet Dev* **16**: 197–202
- Conaway RC, Brower CS, Conaway JW (2002) Emerging roles of ubiquitin in transcription regulation. *Science* **296**: 1254–1258
- Consortium RMGSaA, Consortium RMGSaA, Gibbs R, Rogers J, Katze M, Bumgarner R, Weinstock G, Mardis E, Remington K, Strausberg R, Venter J, Wilson R, Batzer M, Bustamante C, Eichler E, Hahn M, Hardison R, Makova K, Miller W, Milosavljevic A *et al* (2007) Evolutionary and biomedical insights from the Rhesus Macaque genome. *Science* **316**: 222
- Currie RA (1998) NF-Y is associated with the histone acetyltransferases GCN5 and P/CAF. *J Biol Chem* **273**: 1430–1434
- Dadé S, Callebaut I, Mermillod P, Monget P (2003) Identification of a new expanding family of genes characterized by atypical LRR domains. Localization of a cluster preferentially expressed in oocyte. *FEBS Lett* **555**: 533–538
- De Santa F, Narang V, Yap ZH, Tusi BK, Burgold T, Austenaa L, Bucci G, Caganova M, Notarbartolo S, Casola S, Testa G, Sung W-K, Wei C-L, Natoli G (2009) Jmjd3 contributes to the control of gene expression in LPS-activated macrophages. *EMBO J* **28**: 3341–3352
- Denissov S, van Driel M, Voit R, Hekkelman M, Hulsen T, Hernandez N, Grummt I, Wehrens R, Stunnenberg H (2007) Identification of novel functional TBP-binding sites and general factor repertoires. *EMBO J* **26**: 944–954
- Doolan P, Clynes M, Kennedy S, Mehta JP, Crown J, O'Driscoll L (2007) Prevalence and prognostic and predictive relevance of PRAME in breast cancer. *Breast Cancer Res Treat* **109** (2): 359–365
- Epping MT, Hart AAM, Glas AM, Krijgsman O, Bernards R (2008) PRAME expression and clinical outcome of breast cancer. *Br J Cancer* **99**: 398–403
- Epping MT, Wang L, Edel MJ, Carlée L, Hernandez M, Bernards R (2005) The human tumor antigen PRAME is a dominant repressor of retinoic acid receptor signaling. *Cell* **122**: 835–847
- Florens L, Carozza M, Swanson S, Fournier M, Coleman M, Workman J, Washburn M (2006) Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40**: 303–311
- Florens L, Washburn M (2006) Proteomic analysis by multidimensional protein identification technology. *Methods Mol Biol* **328**: 159–175
- Grskovic M, Chaivorapol C, Gaspar-Maia A, Li H, Ramalho-Santos M (2007) Systematic identification of cis-regulatory sequences active in mouse and human embryonic stem cells. *PLoS Genet* **3**: e145
- Gurtner A, Fuschi P, Magi F, Colussi C, Gaetano C, Dobbstein M, Sacchi A, Piaggio G (2008) NF-Y dependent epigenetic modifications discriminate between proliferating and postmitotic tissue. *PLoS ONE* **3**: e2047
- Gurtner A, Manni I, Fuschi P, Mantovani R, Guadagni F, Sacchi A, Piaggio G (2003) Requirement for down-regulation of the CCAAT-binding activity of the NF-Y transcription factor during skeletal muscle differentiation. *Mol Biol Cell* **14**: 2706–2715
- Haqq C, Nosrati M, Sudilovsky D, Crothers J, Khodabakhsh D, Pulliam BL, Ferman S, Miller JR, Allen RE, Singer MI, Leong SPL, Ljung B-M, Sagebiel RW, Kashani-Sabet M (2005) The gene expression signatures of melanoma progression. *Proc Natl Acad Sci USA* **102**: 6092–6097
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M *et al* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–112
- Heintzman ND, Stuart RK, Hon GC, Fu Y, Ching CW, Hawkins RD, Barrera LO, Calcar Sv, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318
- Ikedo H, Lethé B, Lehmann F, van Baren N, Baurain JF, de Smet C, Chambost H, Vitale M, Moretta A, Boon T, Coulié PG (1997) Characterization of an antigen that is recognized on a melanoma showing partial HLA loss by CTL expressing an NK inhibitory receptor. *Immunity* **6**: 199–208
- Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4**: 1265–1272
- Kaelin WG (2007) Von Hippel-Lindau disease. *Ann Rev Pathol* **2**: 145–173

- Kamura T, Koepp DM, Conrad MN, Skowyra D, Moreland RJ, Iliopoulos O, Lane WS, Kaelin WG, Elledge SJ, Conaway RC, Harper JW, Conaway JW (1999) Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science* **284**: 657–661
- Kamura T, Maenaka K, Kotoshiba S, Matsumoto M, Kohda D, Conaway RC, Conaway JW, Nakayama KI (2004) VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases. *Genes Dev* **18**: 3055–3065
- Kamura T, Sato S, Haque D, Liu L, Kaelin WG, Conaway RC, Conaway JW (1998) The Elongin BC complex interacts with the conserved SOCS-box motif present in members of the SOCS, ras, WD-40 repeat, and ankyrin repeat families. *Genes Dev* **12**: 3872–3881
- Kilpinen S, Autio R, Ojala K, Iljin K, Bucher E, Sara H, Pisto T, Saarela M, Skotheim RI, Björkman M, Mpindi J-P, Haapa-Paananen S, Vainio P, Edgren H, Wolf M, Astola J, Nees M, Hautaniemi S, Kallioniemi O (2008) Systematic bioinformatic analysis of expression levels of 17,330 human genes across 9,783 samples from 175 types of healthy and pathological tissues. *Genome Biol* **9**: R139
- Krogan NJ, Kim M, Tong A, Golshani A, Cagney G, Canadien V, Richards DP, Beattie BK, Emili A, Boone C, Shilatifard A, Buratowski S, Greenblatt J (2003) Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol* **23**: 4207–4218
- Lasonder E, Janse CJ, van Gemert G-J, Mair GR, Vermunt AMW, Douradinha BG, van Noort V, Huynen MA, Luty AJF, Kroeze H, Khan SM, Sauerwein RW, Waters AP, Mann M, Stunnenberg HG (2008) Proteomic profiling of Plasmodium sporozoite maturation identifies new proteins essential for parasite development and infectivity. *PLoS Pathog* **4**: e1000195
- Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WAM, Cohen A, Lasonder E, Stunnenberg HG (2006) MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol Cell Biol* **26**: 843–851
- Li Q, Herrler M, Landsberger N, Kaludov N, Ogryzko VV, Nakatani Y, Wolffe AP (1998) Xenopus NF-Y pre-sets chromatin to potentiate p300 and acetylation-responsive transcription from the Xenopus hsp70 promoter *in vivo*. *EMBO J* **17**: 6300–6315
- Loneragan KM, Iliopoulos O, Ohh M, Kamura T, Conaway RC, Conaway JW, Kaelin WG (1998) Regulation of hypoxia-inducible mRNAs by the von Hippel-Lindau tumor suppressor protein requires binding to complexes containing elongins B/C and Cul2. *Mol Cell Biol* **18**: 732–741
- Mahrour N, Redwine WB, Florens L, Swanson SK, Martin-Brown S, Bradford WD, Staehling-Hampton K, Washburn MP, Conaway RC, Conaway JW (2008) Characterization of Cullin-box sequences that direct recruitment of Cul2-Rbx1 and Cul5-Rbx2 modules to Elongin BC-based ubiquitin ligases. *J Biol Chem* **283**: 8005–8013
- Maity SN, de Crombrughe B (1998) Role of the CCAAT-binding protein CBF/NF-Y in transcription. *Trends Biochem Sci* **23**: 174–178
- Manni I, Caretti G, Artuso S, Gurtner A, Emiliozzi V, Sacchi A, Mantovani R, Piaggio G (2008) Posttranslational regulation of NF-YA modulates NF-Y transcriptional activity. *Mol Biol Cell* **19**: 5203–5213
- Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci USA* **98**: 12902–12907
- Muratani M, Kung C, Shokat KM, Tansey WP (2005) The F box protein Dsg1/Mdm30 is a transcriptional coactivator that stimulates Gal4 turnover and cotranscriptional mRNA processing. *Cell* **120**: 887–899
- Nielsen R, Pedersen TA, Hagenbeek D, Moulos P, Siersbaek R, Megens E, Denissov S, Børgesen M, Francoijs K-J, Mandrup S, Stunnenberg HG (2008) Genome-wide profiling of PPARgamma: RXR and RNA polymerase II occupancy reveals temporal activation of distinct metabolic pathways and changes in RXR dimer composition during adipogenesis. *Genes Dev* **22**: 2953–2967
- Oberthuer A, Hero B, Spitz R, Berthold F, Fischer M (2004) The tumor-associated antigen PRAME is universally expressed in high-stage neuroblastoma and associated with poor outcome. *Clin Cancer Res* **10**: 4307–4313
- Olsen JV, Ong S-E, Mann M (2004) Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **3**: 608–614
- Paoletti A, Parmely T, Tomomori-Sato C, Sato S, Zhu D, Conaway R, Conaway J, Florens L, Washburn M (2006) Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc Natl Acad Sci USA* **103**: 18928–18933
- Parthen K, Levan K, Osterberg L, Claesson I, Fallénius G, Sundfeldt K, Horvath G (2008) Four potential biomarkers as prognostic factors in stage III serous ovarian adenocarcinomas. *Int J Cancer* **123**: 2130–2137
- Paydas S (2008) Is everything known in all faces of iceberg in PRAME? *Leukemia Res* **32**: 1356–1357
- Petroski MD, Deshaies RJ (2005) Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* **6**: 9–20
- Radich JP, Dai H, Mao M, Oehler V, Schelter J, Druker B, Sawyers C, Shah N, Stock W, Willman CL, Friend S, Linsley PS (2006) Gene expression changes associated with progression and response in chronic myeloid leukemia. *Proc Natl Acad Sci USA* **103**: 2794–2799
- Reya T, Morrison SJ, Clarke MF, Weissman IL (2001) Stem cells, cancer, and cancer stem cells. *Nature* **414**: 105–111
- Roman-Gomez J, Jimenez-Velasco A, Agirre X, Castillejo JA, Navarro G, Jose-Eneriz ES, Garate L, Cordeu L, Cervantes F, Prosper F, Heiniger A, Torres A (2007) Epigenetic regulation of PRAME gene in chronic myeloid leukemia. *Leukemia Res* **31**: 1521–1528
- Salghetti SE, Caudy AA, Chenoweth JG, Tansey WP (2001) Regulation of transcriptional activation domain function by ubiquitin. *Science* **293**: 1651–1653
- Salsi V, Caretti G, Wasner M, Reinhard W, Haugwitz U, Engeland K, Mantovani R (2003) Interactions between p300 and multiple NF-Y trimers govern cyclin B2 promoter function. *J Biol Chem* **278**: 6642–6650
- Schenk T, Stengel S, Goellner S, Steinbach D, Saluz HP (2007) Hypomethylation of PRAME is responsible for its aberrant overexpression in human malignancies. *Genes Chromosomes Cancer* **46**: 796–804
- Sinha S, Maity SN, Lu J, de Crombrughe B (1995) Recombinant rat CBF-C, the third subunit of CBF/NFY, allows formation of a protein-DNA complex with CBF-A and CBF-B and with yeast HAP2 and HAP3. *Proc Natl Acad Sci USA* **92**: 1624–1628
- Steinbach D, Hermann J, Viehmann S, Zintl F, Gruhn B (2002) Clinical implications of PRAME gene expression in childhood acute myeloid leukemia. *Cancer Genet Cytogenet* **133**: 118–123
- Steinbach D, Pfaffendorf N, Wittig S, Gruhn B (2007) PRAME expression is not associated with down-regulation of retinoic acid signaling in primary acute myeloid leukemia. *Cancer Genet Cytogenet* **177**: 51–54
- Tajeddine N, Gala J-L, Louis M, Van Schoor M, Tombal B, Gailly P (2005) Tumor-associated antigen preferentially expressed antigen of melanoma (PRAME) induces caspase-independent cell death *in vitro* and reduces tumorigenicity *in vivo*. *Cancer Res* **65**: 7348–7355
- van Baren N, Chambost H, Ferrant A, Michaux L, Ikeda H, Millard I, Olive D, Boon T, Coulie PG (1998) PRAME, a gene encoding an antigen recognized on a human melanoma by cytolytic T cells, is expressed in acute leukaemia cells. *Br J Haematol* **102**: 1376–1379
- van Heeringen SJ, Veenstra GJC (2011) GimmeMotifs: a *de novo* motif prediction pipeline for ChIP-seq experiments. *Bioinformatics (Oxford, England)* **27**: 270–271
- Wang PJ, McCarrey JR, Yang F, Page DC (2001) An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* **27**: 422–426
- Washburn M, Wolters D, Yates J (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**: 242–247
- Zhang Y, Liu T, Meyer CA, Eickhout J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137
- Zhu J, Zhang Y, Joe GJ, Pompetti R, Emerson SG (2005) NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal. *Proc Natl Acad Sci USA* **102**: 11728–11733