



HAL
open science

Protein network study of human AF4 reveals its central role in the RNA Pol II-mediated transcription and in phosphorylation-dependent regulatory mechanisms

Gabriella Esposito, Armando Cevenini, Alessandro Cuomo, Francesca de Falco, Dario Sabbatino, Fabrizio Pane, Margherita Ruoppolo, Francesco Salvatore

► To cite this version:

Gabriella Esposito, Armando Cevenini, Alessandro Cuomo, Francesca de Falco, Dario Sabbatino, et al.. Protein network study of human AF4 reveals its central role in the RNA Pol II-mediated transcription and in phosphorylation-dependent regulatory mechanisms. *Biochemical Journal*, 2011, 438 (1), pp.121-131. 10.1042/BJ20101633 . hal-00612964

HAL Id: hal-00612964

<https://hal.science/hal-00612964v1>

Submitted on 2 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Protein network study of human AF4 reveals its central role in the RNA Pol II-mediated transcription and in phosphorylation-dependent regulatory mechanisms

Gabriella Esposito[†], Armando Cevenini*[†], Alessandro Cuomo*, Francesca De Falco[†], Dario Sabbatino[†], Fabrizio Pane*[†], Margherita Ruoppolo*[†], Francesco Salvatore*[‡]*

*CEINGE - Biotecnologie Avanzate, s.c. a r.l., Via G. Salvatore 486, 80145 Naples, Italy; [†]Dipartimento di Biochimica e Biotecnologie Mediche, Università di Napoli “Federico II”, Via S. Pansini 5, 80131 Naples, Italy, [‡]IRCCS-Fondazione SDN, Via E. Gianturco 113, 80143 Naples, Italy.

Corresponding author: Francesco Salvatore, CEINGE - Biotecnologie Avanzate, s.c. a r.l. c/o Università di Napoli “Federico II”, Via S. Pansini, 5 80131 Naples, Italy. E-mail: salvator@unina.it; Tel: 0039-081-7464966; Fax: 0039-081-7463650.

Short title: The human AF4 protein network

Abbreviations: AEP complex, AF4, ENL, P-TEFb complex; ALF, AF4/LAF4/FMR2; BIND, Biomolecular Interaction Network Database; CTD, C-terminal domain; DMEM, Dulbecco’s modified Eagle’s medium; EAF1, ELL-associated factor 1; ELL, Eleven-Nineteen Lysine-rich Leukemia; HEK, human embryonic kidney; HPRD; Human Protein Reference Database; IP, immunoprecipitation; LC-ESI-MS/MS, liquid chromatography-electrospray ionization-tandem mass spectrometry analyses; MLL, mixed-lineage leukemia; PIC, pre-initiation complex; Pol II, RNA polymerase II; P-TEFb, positive transcription elongation factor b; SEC, super elongation complex.

SYNOPSIS

AF4 belongs to a family of proteins implicated in childhood lymphoblastic leukaemia, FRAXE mental retardation and ataxia. AF4 is a transcriptional activator that is involved in transcriptional elongation. Although AF4 has been implicated in MLL-related leukaemogenesis, AF4-dependent physiological mechanisms are not clearly defined. Proteins that interact with AF4 may play important roles also in mediating oncogenesis, and are potential targets for novel therapies. Using a functional proteomic approach involving tandem mass spectrometry and bioinformatics, we identified 51 AF4-interacting proteins of various Gene Ontology Categories. About 60% participate in transcription regulatory mechanisms, including the Mediator complex in eukaryotic cells. Here we report the first extensive proteomic study aimed at elucidating AF4 protein cross-talk. Moreover, we found that the AF4 residues Thr-220 and Ser-212 are phosphorylated, which suggests that AF4 function depends on phosphorylation mechanisms. We also mapped the AF4 interaction site with CDK9, which is a direct interactor crucial for the protein's function and regulation. Our findings significantly expand the number of putative members of the multiprotein complex formed by AF4, which is instrumental in promoting the transcription/elongation of specific genes in human cells.

Key Words: AEP complex, AF4 interactome, lymphoblastic acute leukemia, MLL chimeras, protein-protein interactions, Mediator complex, Pol II-mediated transcription elongation.

INTRODUCTION

AF4 is the prevalent (37%) mixed-lineage leukemia (*MLL*) fusion gene associated with spontaneous acute lymphoblastic leukemia [1,2]. The *AF4* gene transcript is ubiquitously expressed in all types of hematopoietic cells and in other human tissues including brain [2,3]. The AF4 protein is a member of the AF4/LAF4/FMR2 family (ALF) of nuclear proteins, which include AF4, AF5q31, LAF4 and FMR2 [5-8]. *AF4*, *AF5q31*, and *LAF4* form fusion genes with *MLL* in leukemia [9]. There are three regions that are conserved in ALF family members: the N-terminal homology domain, the ALF domain, which contains a proline-serine-rich region, and the C-terminal homology domain (Figure 1) [9]. Furthermore, all the ALF family members, except FMR2, have a transactivation domain [9]. The ALF domain seems to promote ALF protein degradation through the proteasome pathway by mediating their interaction with SIAH ubiquitin ligases [10,11]. A murine AF4 knockout model demonstrated that AF4 is important for normal lymphocyte development and cell growth [12]. Furthermore, *Af4* was identified as the disease gene in the robotic mouse, a dominant N-ethyl-N-nitrosourea mutant that, in addition to defects in early T-cell maturation, develops ataxia because of Purkinje cell degeneration in the cerebellum [13]. This finding suggested that AF4 plays a hitherto unknown role in a function of Purkinje cells that is essential for the control of balance and motor coordination [14]. Moreover, AF4, also through the interaction with the positive transcription elongation factor b (P-TEFb) [15], ENL and/or AF9 [5], was found to have transcriptional regulatory properties that entail elongation and chromatin remodelling involving RNA polymerase II (Pol II) [15]. Noteworthy, ENL family proteins, ENL and AF4, associate with DOT1L, the histone methyltransferase that modifies H3K79 and marks actively transcribed genes [15]. A very recent study demonstrated that AF4 belongs to a high-order multiprotein complex, which is constituted by at least P-TEFb, ENL and AF5q31 (the "AEP complex"). This multiprotein complex is recruited by wild-type MLL only on some of its target promoters (i.e. *HOXA9* and *MEIS1*) [16] by context-dependent mechanisms that are still unknown. In contrast, chimeric oncoproteins originating from the fusion of MLL with AEP components (MLL-AF4, MLL-AF5q31, MLL-ENL) constitutively forms

hybrid complexes to cause sustained expression of MLL target genes that leads to leukemic transformation of hematopoietic cells [16].

In the present study, we used functional proteomics procedures to look for proteins that interact with human AF4 in the attempt to gain further insights into its function and the regulatory mechanism in which it is involved in physiological as well as in pathological transcriptional pathways.

EXPERIMENTAL

Plasmids. The full-length *AF4* cDNA (GenBank # NM_005935), and partial *AF4* cDNAs, designated AF4-1 (bp 4-1950; aa 2-650), AF4-1.1 (bp 4-833; aa 2-277), AF4-1.2 (bp 696-1497; aa 232-499) and AF4-1.3 (bp 1384-1950; aa 462-650), were cloned into the N-terminal p3X-Flag 7.1 Vector (Sigma-Aldrich, Saint. Louis, MO, USA), to obtain recombinant proteins tagged with a Flag epitope at the N-terminus (Figure 1). Primer sequences and cycling conditions are available upon request.

Antibodies. Mouse monoclonal anti-Flag M2 and anti- α -tubulin (Sigma-Aldrich, Inc., Santa Cruz, CA 95060 USA, No. F3165, T9026); rabbit polyclonal anti-AF4 for IP and anti-AF4 for WB (Bethyl Laboratories, Inc., Montgomery, TX, USA, No. 302-345A, 302-344A); rabbit polyclonal anti-CDK9, anti-ELL, anti-YWHAQ, anti-YWHAE, anti-MED1 (Santa Cruz Biotechnologies, No. sc-8338, sc-28702, sc-732, sc-1020 sc-8998); goat polyclonal anti-MED7, anti-CRSP3, anti-SIAH-1, anti-MED27, anti-MED24, anti-MED17, anti-MED6, anti-MED26 (Santa Cruz Biotechnologies, No. sc-9436, sc-12454, sc-5505, sc-9421, sc-5338, sc-12453, sc-9434, sc-9425); anti-mouse and anti-rabbit (GE Healthcare, No. NXA931, NA934); anti-goat (Santa Cruz Biotechnologies, No. sc-2020)

Cell Culture and Transfection. Human embryonic kidney (HEK)-293 cells (ATCC No. CRL-1573) were grown in Dulbecco's modified Eagle's medium (DMEM) (Lonza Group Ltd, Basel, Switzerland), supplemented with 10% FBS (Lonza) and 10 ml/l penicillin-streptomycin (Sigma-Aldrich, No P-0906). Cells were seeded 24 h before transfection at about $6 \times 10^4/\text{cm}^2$ confluency and transfected for 48 h using the calcium phosphate method. 697 cells (human pre-B lineage leukemia) (DSMZ no. ACC

42) were cultured in RPMI 1640 medium (Lonza) supplemented with 20% FBS (Lonza), L-Glutamine 2 mM (Sigma-Aldrich, No. G7513) and 10 ml/l penicillin-streptomycin (Sigma-Aldrich, No. P-0906).

Cell Lysis, Protein extraction, and Immunoprecipitation. HEK-293 cells were transfected with either a recombinant or empty vector (mock control) and lysed in immunoprecipitation (IP) buffer A (10% glycerol, 50 mM Tris-HCl pH 8, 150 mM NaCl, 0.1% NP-40, 0.5 M EDTA pH 8, 10 ml/ml inhibitor cocktail for mammalian tissues (PICM)) (Sigma-Aldrich, No. P8340). 697 cells were lysed in IP buffer B (50 mM Tris-HCl pH 8, 150 mM NaCl, 10 mM KCl, 1.5 mM MgCl₂, NaF 10 mM, 1.5 mM MgCl₂, 1 mM NaV, 0.2 mM EDTA pH8, 0.5% NP-40, 10 µl/ml PICM).

For Flag-IP assays, the lysate was incubated for 1 h at 4°C with anti-Flag M2-agarose affinity gel (Sigma-Aldrich, No. A2220) using 40 µl per 10 mg of total proteins. Immunocomplexes were eluted from the anti-Flag affinity gel using 1 µl elution buffer (IP buffer A with 200 µg/ml 3 x Flag peptide [Sigma-Aldrich, No. F4799]) per µl of gel. For IP assays using specific antibodies directed against endogenous interactors, lysates either from transfected HEK-293 or 697 cells were incubated overnight at 4°C with 1-3 µg of specific antibody, i.e. anti-MED7, anti-CRSP3, anti-CDK9, anti-ELL, anti-YWHAQ, anti-YWHAE, anti-SIAH-1 and, only for lysate from 697 cells, anti-AF4 antibody for IP, per 10 mg of total proteins. Subsequently, the mixture from HEK-297 was incubated with 30 µl of Protein A/G PLUS-Agarose (Santa Cruz) per 1 µg of antibody and the immunocomplexes were washed several times with IP buffer A. Otherwise, the mixture from 697 cells was incubated with 5 µl of nProtein A Sepharose 4 Fast Flow (GE Healthcare, Chalfont St. Giles, United Kingdom, No. 17-5280-01) per 1 µg of antibody and the sepharose-bound immunocomplexes were washed several times with IP buffer B. As negative control, we used cell extracts incubated with serum IgGs (goat or rabbit, Sigma Aldrich, No. I8140, I9140) then processed as the samples.

Western Blot. Either 40 µg of whole cell extract (WCE) from HEK-293, or 40 µg of WCE from 697 cells, or 15 µl of sample from each IP experiment were loaded onto SDS-PAGE and then transferred

onto a nitrocellulose membrane for western blot. Blots were incubated with the following primary antibodies: anti-Flag M2 (dil. 1:5000) or anti- α -tubulin (1:1000); anti-AF4 for WB (1:8000); anti-CDK9, anti-YWHAQ, anti-YWHAE, anti-SIAH-1 (1:1000); anti-ELL (1:500); anti-MED1, anti-MED7, anti-CRSP3, anti-MED27, anti-MED24, anti-MED17, anti-MED6, anti-MED26 (1:200). Membranes were then incubated with HRP-conjugated appropriated secondary antibodies, i.e. anti-mouse or anti-rabbit (1:5000) or anti-goat (1:8000). The bands were visualized with the ECL or ECL Plus detection system (GE Healthcare, No. RPN2106, RPN2132).

Preparative Immunoaffinity Purification. AF4-1 interacting proteins were purified using anti-Flag M2 affinity gel (Sigma-Aldrich) (as described above) on 50 mg of whole cellular extract from HEK-293 cells transfected with the Flag-AF4-1 construct. The same purification protocol was carried out also with an equal amount of extract from HEK-293 cells that were transfected with the empty p3X-Flag vector (mock control). The proteins eluted from the agarose beads were chloroform/methanol-precipitated and the pellets were dried [17].

SDS-PAGE, in Gel Digestion and Mass Spectrometric Analysis. The protein pellets obtained by affinity purification from transfected and from mock control cells were resuspended in SDS-PAGE buffer and fractionated on a 10% SDS-PAGE. Protein electrophoretic patterns were visualized using Gel Code Blue Stain Reagent (Pierce, Rockford, IL, USA). The two gel lanes were cut to generate 38 slices (~ 2 mm) per lane. Subsequent in gel digestion and Nano liquid chromatography-electrospray ionization-tandem mass spectrometry analyses (LC-ESI-MS/MS) of each protein slices were performed as described elsewhere [18,19].

Raw data from Nano LC-ESI-MS/MS analyses were converted into a peak list by using LC-MSD Trap 6.0 Build 485.0 (Agilent technologies) to create a Mascot format text. Proteins were identified by means of a Mascot software version 2.1 in house, (Matrix Science, Inc. Boston, MA 02110 USA) [20]. The protein search was governed by the following parameters: non-redundant protein sequence data base (NCBIInr, July 2009 - 20070908 database with 7,415,798 sequences and 2,558,340,887 residues

downloaded; Sprot, July 2009 - 57.6 database with 495,880 sequences and 174,780,353 residues downloaded; the protein sequence entries actually searched numbered 187,996 for NCBIInr and 20,331 for Sprot because of the taxonomy restriction to *Homo sapiens*); specificity of the proteolytic enzyme used for hydrolysis (trypsin); taxonomic category of the sample (*Homo sapiens*); no protein molecular weight was considered; up to 1 missed cleavage; carbamidomethyl cysteine was set as a fixed modification; protein N-acetylation, N-terminal pyroglutamate, oxidized methionines, and phosphorylation of serine, threonine or tyrosine were set as variable modifications; a precursor peptide maximum mass tolerance of 100 ppm and a maximum fragment mass tolerance of 150 ppm. According to the probability-based Mowse score [20], the ion score is $-10 \times \log(P)$, where P is the probability that the observed match is a random event. Individual scores >38 indicate identity or extensive homology ($p < 0.05$). Individual MS/MS spectra for peptides with a Mascot score equal to 38 were inspected manually. All peptides were inspected for redundancy and represent sequences that are unique within the database used. Phosphopeptide spectra with a minimum Mascot score of 38 ($p < 0.05$) were considered for further manual data interpretation. Phosphorylation site assignments by Mascot were validated and verified manually using the MS-Product tool (University of California San Francisco Mass Spectrometry Facility, <http://prospector.ucsf.edu/>) with loss of H_3PO_4 and multiple losses enabled as an additional option, focusing on the occurrence of b- and y-ions, internal fragments, and neutral losses of H_2O , NH_3 , or H_3PO_4 from these ions and the parent ion, respectively.

In silico Analysis. To define an *in silico* “interactome” map of AF4, we screened for AF4 interactions the Human Protein Reference Database (HPRD), which is freely available at www.hprd.org [21]. We created a list of the known protein-protein interactions of AF4, and a list of primary (direct) and secondary (indirect) interactions was reported for each identified protein. In April 2011, the HPRD database contained 30,047 human protein entries, 39,194 protein-protein interactions and 453,521 PubMed links. We used HPRD, Biomolecular Interaction Network Database (BIND),

(<http://bond.unleashedinformatics.com/Action>), DAVID bioinformatics resources (www.david.abcc.ncifcrf.gov) and STRING (<http://string.embl.de>) for protein clustering.

RESULTS AND DISCUSSION

We have used a functional proteomics approach to identify, in the HEK-293 cell line, proteins that interact with AF4, a member of the ALF family that supplies most of the partners of the oncogenic fusion proteins involved in childhood acute lymphoblastic leukemia [1]. Our study represents the first global functional proteomic analysis to determine the identity of proteins that form a multimeric complex with AF4.

In contrast to HeLa, SAOS, K562 and SK-NB-E cell lines, the human HEK-293 cell line efficiently expressed our “baits,” the recombinant proteins Flag-AF4-1 and Flag-AF4 (Figure 1). Moreover, as reported in the HPRD and Unigene databases, the AF4 transcript is expressed in kidney, and HEK-293 cells have previously served as a heterologous expression system to study mouse recombinant Af4 [15]. We transfected HEK-293 with the full-length AF4 and with the AF4-1 cDNA constructs (Figure 2A). Co-immunoprecipitation with the flagged full-length AF4 protein yielded protein amounts that were sufficient for western blot but not for an exhaustive functional proteomic analysis. Consequently, we used the Flag-AF4-1 construct for large-scale immunoaffinity purification, and then validated the results in western blot experiments with the full-length recombinant protein. It is noteworthy that Flag-AF4-1 encodes the N-terminal half of the protein containing the transactivation domain (Figure 1), but not the binding domains for AF9 and ENL, the other two most common MLL fusion partners [22].

Proteins from total lysates of HEK-293 cells transfected with Flag-AF4-1 and with the empty vector (mock control) were immunoprecipitated and immunocomplexes were fractionated on 10% SDS-PAGE (Figure 2B). From each gel lane, we obtained 38 peptide mixtures, each of which was analyzed in duplicate by mass spectrometry. Peptide mixtures from the control lane were always injected before peptides from the sample lane to avoid carry-over.

Our analysis, although unable to distinguish between direct and indirect interactors, revealed 51 proteins, most of which were not previously known to be molecular partners of AF4. They are listed in Table 1, grouped according to the function. Supplementary Table I shows, for each protein entry, the peptide sequence identified, the precursor mass (m/z), the charge state (z), the mass errors (ppm) on the precursor peptide, the MASCOT score for each peptide, and the protein sequence coverage. Each protein is reported only with the gene symbol from NCBI source to eliminate redundancy.

To select proteins that specifically interact with AF4-1, we subtracted species that were common to AF4-1 and non-transfected control lanes (Figure 2B). These common proteins are shown in Supplementary Table II. All protein species identified by a single peptide were checked further. First, the peptide sequence stretch was manually reconstructed, and then the peptide sequence and the peptide precursor ion mass were inserted in the MASCOT software using the sequence query mode. All searches were performed against the NCBI nr and Sprot databases. The peptide sequence was searched for using the BLAST software (<http://ncbi.nlm.nih.gov/blast>). Supplementary Table I shows the full MS scan and the properly annotated MS/MS scan (with masses and fragment assignments) of proteins identified by a single peptide.

We classified the human proteins identified according to Gene Ontology cellular localization (Figure 3A) and to the biological processes in which they are involved (Figure 3B). It is noteworthy that the nature of the identified proteins, which are mainly localized in the nucleus (68%), points to the existence of a large protein network involving AF4. About 60% of the identified proteins that take part in this cross-talk is clustered in the RNA polymerase II (Pol II) transcriptional complex.

To check the authenticity of the AF4 interacting proteins identified by MS/MS, we verify the presence of some of these proteins by anti-Flag co-immunoprecipitation and Western blot, depending on the availability and efficiency of commercial antibodies. We confirmed that the CRSP3/MED23, ELL, CDK9, CRSP33/MED7, YWHAQ and YWHAE specifically interact with AF4-1 and with the full-length AF4 (Figure 4A), thus demonstrating that the latter two, despite their different length, have similar

folding in common regions. We also identified SIAH-1, a known molecular direct interactor of AF4 [10,11] (Figure 4A). We validated the AF4 interactors further by reverse immunoprecipitation experiments using specific anti-CDK9, anti-YWHAQ, anti-CRSP3, anti-MED7, anti-ELL and anti-YWHAQ antibodies on extracts from HEK-293 cells transfected with Flag-AF4. Western blot with the anti-Flag antibody revealed the Flag-tagged protein in each immunocomplex (Figure 4B). These interactions were also confirmed by Western blot analysis of immunocomplexes obtained from protein extracts of an hematopoietic cell line, the human pre-B lineage leukemia 697 cells, by using specific antibodies directed against either endogenous AF4 or endogenous interactors (Figure 5). Overall, our immunoprecipitation experiments support the conclusion that endogenous CDK9, YWHAQ, CRSP3/MED23, CRSP33/MED7, ELL and YWHAQ interact with Flag-AF4 in HEK-293, but they also interact with endogenous AF4 in 697 leukemia cells.

We also tested antibodies directed against all the components of the mediator complex that we identified in our proteomic study, namely, CRSP8/MED27, THRAP4/MED24, PPARBP/MED1, CRSP6/MED17, MED6 and CRSP7/MED26. However, these antibodies failed to detect the respective target proteins in cellular extracts from HEK-293, HeLa, SAOS, SK-NB-E, K562 or 697 cells analyzed by western blot (data not shown).

It is noteworthy that in our system we did not identify all proteins previously reported to interact with AF4. This is probably due to the use of the N-terminal half of AF4 as "bait". In fact, we did not find AF9, ENL, DOT1L or AF5q31, the AF4 partners known to interact with its C-terminal half [23]. However, although our bait (Flag-AF4-1) lacks the crucial C-terminal binding domain, we were able to identify various known and new molecular partners, thereby providing further insight into AF4 function in normal and leukemic conditions. Obviously, we cannot exclude that some of the reported interactions may be cell-lineage specific, however, almost all the proteins that form the AF4 interactome as well as AF4 itself are variably expressed in various cell types, according to the UniGene Database (<http://www.ncbi.nlm.nih.gov/unigene>).

AF4 is a phosphorylated protein [15], but its phosphorylation sites are still unknown. LCMS/MS analysis of the bait protein revealed two phosphorylation sites (Figure 2B). The MS2 spectra from the identified phosphopeptides are shown in Figure 6. Peptide ELSPLISLPSPVPPLSPIHSNQQTLPR resulted to be phosphorylated on Thr-220 and Ser-212 (Figure 6A and B, respectively). We also identified the corresponding unmodified peptide (Figure 6C). The phosphorylation sites identified were manually confirmed (Supplementary Table III). Although this peptide was previously reported to be phosphorylated [24], we indicate for the first time the precise localization of the phosphorylation sites.

To extend our data, we carried out an *in-silico* proteomic analysis of proteins that are known to directly or indirectly interact with AF4. Examination of the HPRD database indicated two direct interactions for AF4: SMAD9 and ENL [23, 25]. In the present study, we identified SMAD9, aka SMAD8A or SMAD8B (Table 1 and Supplementary Table S1). We carried out a detailed clustering analysis using the HPRD database to look for a protein network involved in the same functional scenario. To this aim, we filled out a comprehensive list of known primary and secondary interactions for all proteins found in the present work (see Supplementary Table S4).

Several proteins reported in our study interact with each other, thus suggesting that they participate in the same multiprotein complex or complexes. The vast majority of proteins interacting with human AF4 are involved in the Pol II-mediated transcription. Thus, there is compelling evidence that AF4 plays a key role in the Pol II transcription machinery. Figure 7 shows the map of the AF4 interacting proteins, with Pol II as the core of the interactions.

Indeed, our proteomics analysis identified the serine/threonine kinase CDK9 and cyclin T1 (CCNT1) that interact each other and form the positive transcription elongation factor, P-TEFb, which activate the Pol II elongation machinery [26]. It is known that AF4 associates with P-TEFb, positively regulates its kinase activity and stimulates Pol II transcriptional elongation [15]. Through its ability to phosphorylate the Pol II C-terminal domain (CTD), P-TEFb controls productive elongation of most eukaryotic genes, but also coordinates downstream events including pre-mRNA splicing and 3'-end processing [27]. P-

TEFb recruits various positive regulators other than AF4, namely, the transcriptional activators HIV Tat, CIITA, c-Myc, NFkB, MyoD, Brd4, AF5q31 and Eleven-Nineteen Lysine-rich Leukemia (ELL), to specific gene promoters and stimulates transcriptional elongation [16, 26, 28, 29]. A very recent paper demonstrates that AF4, AF5q31 and ENL associate in an endogenous higher-order complex containing P-TEFb, in hematopoietic lineage cells [16]. Our data indicate that such a complex should be formed also in the HEK-293 cells. Therefore, we mapped the CDK9-binding region along AF4, the CDK9-binding region. To this aim, we produced three new flagged constructs, namely AF4-1.1 (base pairs 4-833), AF4-1.2 (base pairs 696-1497) and AF4-1.3 (base pairs 1384-1950) (Figure 1), which we used to transiently transfect HEK-293 cells. We analyzed the corresponding anti-Flag immunoprecipitates by western blot with the anti-CDK9 antibody (Figure 8) and found that, in accordance with Yokoyama and colleagues [16], only AF4-1.1 (aa 2-277) was able to interact with CDK9. Intriguingly, this peptide contains Ser-212 and Thr-220, which we found to be phosphorylated in the our structural characterization of AF4.

P-TEFb phosphorylates AF4 and down-regulates its transactivation activity [15]. Therefore, phosphorylation is a key control mechanism of AF4 activity and the AF4 N-terminal region might reasonably be the P-TEFb phosphorylation site. This type of control should thus be ineffective on MLL-AF4 chimeras that lack the AF4 N-terminal. It is notable that, phosphorylation events regulate the function of other components that we identified in the AF4 protein network (e.g., 14-3-3 proteins, FGFR2, Pol II, nucleolin and MCMs).

In eukaryotes, Pol II is the central component of the basal Pol II transcription machinery. It moves on the template as the transcript elongates [30]. Elongation is influenced by the P-TEFb-mediated phosphorylation of the CTD of the largest Pol II subunit [31]. In accordance with our study and with reports from other groups showing that AF4 directly interacts with P-TEFb [15,16], we identified, in the AF4 multiprotein complex, the largest subunit (POLR2A), the second largest subunit (POLR2B) and a small subunit of 33 kDa (POLR2C) Pol II. In addition, we found: 1) nucleolin, which forms complexes

with POLR2A, CDK9 and CCNT1, and plays a role in transcriptional elongation [32]; 2) TCEA1, another Pol II transcription elongation factor, which has a role in suppression of transient pausing and strongly synergizes with p300 histone acetylase at a step subsequent to preinitiation complex formation [33]; 3) MCMs, a family of proteins related to ATP-dependent helicases, which has been co-purified with Pol II after anti-MCM3 immunoaffinity chromatography [34]. Our results support the concept that MCM proteins are components of the Pol II transcriptional apparatus [34].

We also found that AF4-1 interacts with the Pol II elongation factor ELL that, as well as AF4, ENL and AF5q31, is frequently fused to MLL in childhood lymphoblastic leukemia [34]. ELL is a multifunction factor that exerts transcriptional elongation activity and inhibitory effects on the initiation of Pol II-mediated transcription [35, 36], besides acting as a transcription factor [37]. In particular, the C-terminal domain of ELL interacts with P-TEFb [36], a molecular interactor of the N-terminal domain of AF4, as we demonstrate herein. Our proteomic analysis also identified the ELL-associated factor 1 (EAF1) [38]. EAF1 interaction with the C-terminal domain of ELL is necessary and sufficient for the leukemogenic effect of the MLL-ELL fusion protein [38]. Notably, a heterologous MLL-EAF1 fusion protein recapitulates the phenotype of MLL-ELL *in vitro* and *in vivo* [38]. EAF1 is a strong positive regulator of ELL elongation activity and contains a transactivation domain that has high sequence identity with the transactivation domain of AF4 [38]. Therefore, it is not surprising that the latter may also associate with ELL and positively regulate its transactivating function [38,39]. The identification of ELL and EAF1 among AF4 interactors is in line with a very recent report demonstrating that ELL and EAF1 are part of a super elongation complex (SEC), which includes P-TEFb, AF5q31, ENL and AF4, that is crucial in the control of transcription elongation [40]. Such a complex should be recruited constitutively by MLL-chimeras that should thus be able to bypass the normal transcription initiation and elongation checkpoint steps and activate aberrant MLL-target gene expression [16, 40]. Indeed, knock-down of the central SEC component AF5q31, in MLL-AF4 leukaemia cells, causes a reduction in the expression of *HOXA9*, a key mediator of leukaemogenesis [40].

Although we cannot be sure about the interactions that were not validated with IP-western blot methods, the data discussed so far demonstrate the appropriateness of our proteomic approach. In fact, all the interactions we identified are in line with the most recent findings about the composition of the multiprotein complex involving AF4 [16, 40]. Indeed, the finding of Pol II subunits in this complex supports the crucial role of AF4 in the transcriptional machinery. However, the novelty of our study resides in the identification of 15 of the 28 proteins that form the mammalian Mediator complex (Table 1).

The Mediator complex is a multiprotein transcription factor that is evolutionarily conserved and ubiquitously expressed in eukaryotes from yeast to man [41, 42]. This large multisubunit complex is the primary regulator of the assembly of the pre-initiation complex (PIC), which includes the general transcription factors TFIID, TFIIB, TFIIF, TFIIE and Pol II [41–43]. In general, transcription factors and activators physically bind to a specific subunit of Mediator and thereby recruit other components of the complex to the target promoter. In particular, activator-binding provokes specific conformational shifts in Mediator that induce a conformational state in the neighbouring subunits of the complex. Indeed, Mediator can adopt various conformations after the binding of different transcriptional activators to different Mediator subunits [44]. Therefore, structural changes in Mediator afford additional opportunities to fine-tune the diverse regulatory inputs received from the DNA-binding factors and from other signals to the transcription machinery [45]. Genetic, biochemical and structural data revealed that Mediator comprises several modules (head, middle, tail and CDK8) [46]. The head, middle and tail modules form a relatively stable ‘core’ that is composed of 18-20 tightly associated subunits [40, 45]. Otherwise, the subunits PPARBP/MED1, MED25 and MED8, and the module constituted by CRSP3/MED23, THRAP4/MED24 and TRAP95/MED16 are variably and weakly associated with the central core [40]. The constant core and the variably associated proteins are considered components of the active forms of the Mediator complex. In contrast, SRB10/CDK8, MED13, MED12 and SRB11/CycC form a functional and physically separable module that has been

implicated in transcriptional repression [40, 45]. The potential microheterogeneity could result in a wide spectrum of mammalian Mediator complexes [47].

Our proteomic analysis of the AF4 complex identified 10 Mediator central core subunits (TRAP36/MED4, TRAP32/MED6, CRSP33/MED7, CRSP2/MED14, CRSP6/MED17, TRFP/MED20, MED21, CRSP7/MED26, PCQAP/MED15 and CRSP8/MED27) and four variable subunits (PPARBP/MED1, MED8, CRSP3/MED23, THRAP4/MED24). Apart from MED12, all the components identified belong to the active form of Mediator. It is noteworthy, however, that MED12 directly interacts with mammalian β -catenin and activates target genes, thus indicating that also MED12 has an alternative activating role [48]. Interestingly, β -catenin is activated during development of MLL leukemic stem cells [49].

As our identifications indicate that AF4 interacts with Pol II and the Mediator complex, we may speculate that AF4 binding induces structural change in Mediator thereby activating stalled Pol II to transition to a positively-elongating state and enabling effective transcription to take place. Therefore, the AF4-Mediator complex interaction might be crucial for activation of specific gene expression. In this context, it is noteworthy that Yokoyama and colleagues [16] demonstrate that the AEP complex (AF4, AF5q31, ENL and p-TEFb) co-localizes with wild-type MLL on specific target promoters, thereby indicating that this complex plays a role in physiologic as well as in oncogenic MLL-dependent transcriptional pathways. However, they also showed that recruitment of AEP to MLL-target loci was not constitutive, because some MLL-occupied loci (e.g., *HOXA7*) did not contain AEP. Therefore, these authors hypothesized that the MLL complex probably requires other, as yet unidentified factors or signals for specific recruitment of AEP, and concluded that AEP recruitment, a downstream event in physiologic MLL-dependent transcriptional pathways, is regulated in a context-dependent manner. This hypothesis suggested that our data may be integrated in the Yokoyama model. Specifically, we suggest that the Mediator complex, depending on its subunit composition, plays a crucial role in the

recruitment of the AEP complex on specific MLL target promoters. Indeed, the wild-type MLL complex may initiate the activation pathway by binding the regulatory region of target genes. It would then recruit some specific Mediator components, in a context-dependent manner. For example, MLL could recruit the head subunit MED17/TRAP80, a known direct interactor of ASH2L that, together with RBBP4, an AF4 partner (Supplemental Table S4), forms the MLL complex [50, 51]. Alternatively, some Mediator components may recruit MLL directly to the target chromatin. Indeed, PPARBP/MED1 directly interacts with various ligands of nuclear receptors (thyroid hormone receptor, vitamin D receptor, peroxisome proliferator-activated receptor- γ , hepatocyte nuclear factor 4 α , glucocorticoid receptor and oestrogen receptor), as well as with non-receptor type factors such as GATA1. After these initial recruitments, assembly of the pre-initiation complex (PIC), which entails the various general transcription factors and Pol II, and transcription initiation then ensue together with the concomitant recruitment of other *ad hoc* components of Mediator. After Pol II clears the promoter, the process can proceed directly to the elongation phase thanks to the Mediator-dependent specific recruitment of AF4 and the other components of the AEP complex described by Yokoyama et al. [16].

Lastly, we found that YWHAQ and YWHAE, two members of the 14-3-3 protein family, interact with AF4. 14-3-3s are ubiquitous proteins, usually cytosolic, that exert an extraordinarily wide-ranging influence on cellular functions, including cell-cycle control and apoptosis. They operate by binding to specific phosphorylated sites on such diverse target proteins as oncogene products, tumour suppressor proteins and regulators of cell survival, proliferation and growth [52]. 14-3-3s are often associated with dynamic nucleo-cytoplasmic shuttling. Upon phosphorylation, many nuclear proteins, including transcription factors, bind to 14-3-3s, which control their rate of nuclear import/export thereby modulating transcriptional processes [52]. We suggest that 14-3-3s bind phosphorylated AF4 and contribute to the regulation of its movement into and out of the nucleus. In the cytosol, AF4 undergoes rapid proteasomal degradation via its well-known interaction with SIAH-1a and SIAH-2 ubiquitin ligases [10, 11]. This mechanism could closely control AF4 turnover and, consequently, AF4-dependent

transcriptional elongation. The phosphorylation sites that we have identified in AF4 might play a key role in the putative 14-3-3-mediated AF4 regulation/degradation pathway.

The characterization of AF4-interacting proteins reported herein supports the growing body of evidence that AF4 is a crucial activator in a multimeric complex that promotes transcription in human cells, and that this multiprotein complex is functionally and structurally regulated (i.e., via the availability of components, phosphorylation, proteome degradation, cellular localization, conformational shifts). Indeed, our data greatly increases the number of putative components participating in the multiprotein complex formed by AF4 as well as by other MLL fusion partner (ENL ELL, AF5q31, AF9 etc.). The information reported herein is useful for future studies aimed at unravelling the AF4-dependent molecular mechanisms thereby giving insights into the molecular basis of AF4-mediated leukaemia and eventually leading to novel therapeutic targets. We previously reported that even when MLL-AF4 chimeras lack part of the AF4 transactivation domain, they continue to give rise to acute lymphoblastic leukaemia [53]. Given previous [16, 40] and the present results, it is now clear that also this altered MLL-AF4 chimera, irrespective of the transactivation domain, is able to recruit all the protein components that are necessary for the transcription of genes that enhance and sustain cell transformation.

ACKNOWLEDGEMENTS. We are grateful to Jean Ann Gilder for text editing.

FUNDING: This work was supported by the Ministero della Salute (Roma, Italy), by MIUR [PS 35-126 IND] and Progetto SCoPE; by PRIN 2007 to F.S.; by a CEINGE-Regione Campania Convention [DGRC 1901/2009].

SUPPORTING INFORMATION AVAILABLE: Identification of AF4-1 interacting proteins by nanoLC/ESI-MS/MS, along with identified peptide sequences, precursor mass, charge state, mass errors, single peptide MASCOT score and sequence coverage (Supplementary Table S1). Analysis of control

lane by nano LC/ESI-MS/MS, along with identified peptide sequences, precursor mass, charge state, mass errors, single peptide MASCOT score, sequence coverage (Supplementary Table S2). Confirmation of identified phosphorylation sites (Supplementary Table S3). In silico analysis of AF4 interactors by the Human Protein Reference Database (<http://www.hprd.org>) (Supplementary Table S4).

REFERENCES

- 1 Hess, J. L. (2004) MLL: a histone methyltransferase disrupted in leukemia. *Trends Mol. Med.* **10**, 500-507.
- 2 Chen, C. S., Hilden, J. M., Frestedt, J., Domer, P. H., Moore, R., Korsmeyer, S. J., Kersey, J. H. (1993) The chromosome 4q21 gene (AF-4/FEL) is widely expressed in normal tissues and shows breakpoint diversity in t(4;11)(q21;q23) acute leukemia. *Blood.* **82**, 1080-1085.
- 3 Frestedt, J. L., Hilden, J. M., Moore, R. O., Kersey, J. H. (1996) Differential expression of AF4/FEL mRNA in human tissues. *Genet. Anal.* **12**, 147-149.
- 4 Nilson, I., Reichel, M., Ennas, M. G., Greim, R., Knorr, C., Siegler, G., Greil, J., Fey, G. H., Marschalek, R. (1997) Exon/intron structure of the human AF-4 gene a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia. *Br. J. Haematol.* **98**, 157-169.
- 5 Erfurth, F., Hemenway, C. S., de Erkenez, A. C., Domer, P. H. (2004) MLL fusion partners AF4 and AF9 interact at subnuclear foci. *Lukemia.* **18**, 92-102.
- 6 Bensaid, M., Melko, M., Bechara, E. G., Davidovic, L., Berretta, A., Catania, M. V., Gecz, J., Lalli, E., Bardoni, B. (2009) FRAXE-associated mental retardation protein (FMR2) is an RNA-binding protein with high affinity for G-quartet RNA forming structure. *Nucleic Acids Res.* **37**, 1269-1279.
- 7 Ma, C., Staudt, L. M. (1996) LAF-4 encodes a lymphoid nuclear protein with transactivation potential that is homologous to AF-4 the gene fused to MLL in t(4;11) leukemias. *Blood.* **87**, 734-745.

- 8 Taki, T., Kano, H., Taniwaki, M., Sako, M., Yanagisawa, M., Hayashi, Y. (1999) AF5q31, a newly identified AF4-related gene, is fused to MLL in infant acute lymphoblastic leukemia with *ins(5;11)(q31;q13q23)*. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14535-14540.
- 9 Prasad, R., Yano, T., Sorio, C., Nakamura, T., Rallapalli, R., Gu, Y., Leshkowitz, D., Croce, C. M., Canaani, E. (1995) Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 12160-12164.
- 10 Oliver, P. L., Bitoun, E., Clark, J., Jones, E. L., Davies, K. E. (2004) Mediation of Af4 protein function in the cerebellum by Siah proteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14901-14906.
- 11 Bursen, A., Moritz, S., Gaussmann, A., Moritz, S., Dingermann, T., Marschalek, R. (2004) Interaction of AF4 wild-type and AF4•MLL fusion protein with SIAH proteins: indication for t(4;11) pathobiology? *Oncogene.* **23**, 6237-6249.
- 12 Isnard, P., Core, N., Naquet, P., Djabali, M. (2000) Altered lymphoid development in mice deficient for the mAF4 proto-oncogene. *Blood.* **96**, 705-10.
- 13 Isaacs, A. M., Oliver, P. L., Jones, E. L., Jeans, A., Potter, A., Hovik B. H., Nolan, P. M., Vizor, L., Glenister, P., Simon, A. K., Gray, I. C., Spurr, N. K., Brown, S. D., Hunter, A. J., Davies, K. E. (2003) A mutation in Af4 is predicted to cause cerebellar ataxia and cataracts in the robotic mouse. *J. Neurosci.* **23**, 1631-1637.
- 14 Bitoun, E., Davies, K. E. (2005) The robotic mouse: unravelling the function of AF4 in the cerebellum. *Cerebellum*, **4**, 250-260.
- 15 Bitoun, E., Oliver, P. L., Davies, K. E. (2007) The mixed-lineage leukemia fusion partner AF4 stimulates RNA polymerase II transcriptional elongation and mediates coordinated chromatin remodeling. *Hum. Mol. Genet.* **16**, 92-106.
- 16 Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I., Cleary, M. L. (2010) A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription. *Cancer Cell.* **17**, 198-212.

- 17 Jiang, L., He, L., Fountoulakis, M. (2004) Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *J. Chromatogr. A.* **1023**, 317-320.
- 18 Orru, S., Aspesi, A., Armiraglio, M., Caterino, M., Loreni, F., Ruoppolo, M., Santoro, C., Dianzani, I. (2007) Analysis of RPS19's interactome. *Mol. Cell. Proteomics.* **6**, 382-393.
- 19 Yin, H., Killeen, K., Brennen, R., Sobek, D., Werlich, M., Van De Goor, T. (2005) Microfluidic chip for peptide analysis with an integrated HPLC column sample enrichment column and nanoelectrospray tip. *Anal. Chem.* **77**, 527-533.
- 20 Perkins, D. N., Pappin, D. J., Creasy, D. M., Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* **20**, 3551-3567.
- 21 Mathivanan, S., Periaswamy, B., Gandhi, T., Kandasamy, K., Suresh, S., Mohmood, R., Ramachandra, Y., Pandey, A. (2006) An evaluation of human protein-protein interaction data in the public domain. *BMC Bioinformatic.* **7**, S19.
- 22 Krivtsov, A. V., Armstrong, S. A. (2007) MLL translocations, histone modifications and leukaemia stem-cell development. *Nat. Rev. Cancer.* **7**, 823-833.
- 23 Zeisig, D. T., Bittner, C. B., Zeisig, B. B., Garcia-Cuellar, M. P., Hess, J. L., Slany, R. K. (2005) The eleven-nineteen-leukemia protein ENL connects nuclear MLL fusion partners with chromatin. *Oncogene.* **24**, 525-532.
- 24 Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., Gygi, S. P. (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12130-12135.
- 25 Colland, F., Jacq, X., Trouplin, V., Mouglin, C., Groizelau, C., Hamburger, A., Meil, A., Wojcik, J., Legrain, P., Gauthier, J. M. (2006) Functional proteomics mapping of a human signalling pathway. *Genome Res.* **14**, 1324-1332

- 26 Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M. B., Price, D. H. (1997) Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev.* **11**, 2622-2632.
- 27 Ahn, S. H., Kim, M., Buratowski, S. (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol. Cell.* **13**, 67-76
- 28 Yang, Z., Yik, J. H., Chen, R., He, N., Jang, M. K., Ozato, K., Zhou, Q. (2005) Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell.* **19**, 535-45.
- 29 Lin, C., Smith, E. R., Takahashi, H., Lai, K. C., Martin-Brown, S., Florens, L., Washburn, M. P., Conaway, J. W., Conaway, R. C., Shilatifard, A. (2010) AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol. Cell.* **37**, 429-437.
- 30 Hartzog, G. A. (2003) Transcription elongation by RNA polymerase II. *Curr. Opin. Genet. Dev.* **13**, 119-126.
- 31 Hirose, Y., Ohkuma, Y. (2007) Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J. Biochem.* **141**, 601-608.
- 32 Mongelard, F., Bouvet, P. (2007) Nucleolin: a multiFACeTed protein. *Trends Cell. Biol.* **17**, 80-86.
- 33 Guermah, M., Palhan, V. B., Tackett, A. J., Chait, B. T., Roeder, R. G. (2006) Synergistic functions of SII and p300 in productive activator-dependent transcription of chromatin templates. *Cell.* **125**, 275-86.
- 34 Yankulov, K., Todorov, I., Romanowski, P., Licatalosi, D., Cilli, K., McCracken, S., Laskey, R., Bentley, D. L. (1999) MCM proteins are associated with RNA polymerase II holoenzyme. *Mol. Cell. Biol.* **19**, 6154-6163.

- 35 Thirman, M. J., Levitan, D. A., Kobayashi, H., Simon, M. C., Rowley, J. D. (1994) Cloning of ELL a gene that fuses to MLL in a t(11;19)(q23;p13.1) in acute myeloid leukemia. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12110-12114.
- 36 Shilatifard, A., Lane, W. S., Jackson, K. W., Conaway, R. C., Conaway, J. W. (1996) An RNA polymerase II elongation factor encoded by the human ELL gene. *Science*. **271**, 1873-1876.
- 37 Zhou, J., Feng, X., Ban, B., Liu, J., Wang, Z., Xiao, W. (2009) Elongation factor ELL (Eleven-Nineteen Lysine-rich Leukemia) acts as a transcription factor for direct thrombospondin-1 regulation. *J. Biol. Chem.* **284**, 19142-19152.
- 38 Simone, F., Polak, P. E., Kaberlein, J. J., Luo, R. T., Levitan, D. A., Thirman, M. J. (2001) EAF1 a novel ELL-associated factor that is delocalized by expression of the MLL-ELL fusion protein. *Blood*. **98**, 201-209.
- 39 Luo, R. T., Lavau, C., Du, C., Simone, F., Polak, P. E., Kawamata, S., Thirman, M. J. (2001) The elongation domain of ELL is dispensable but its ELL-associated factor 1 interaction domain is essential for MLL-ELL-induced leukemogenesis. *Mol. Cell. Biol.* **21**, 5678-87.
- 40 Smith, E., Lin, C., Shilatifard, A. (2011) The super elongation complex (SEC) and MLL in development and disease. *Genes Dev.* **25**, 661-72.
- 41 Malik, S., Roeder, R. G. (2005) Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem. Sci.* **30**, 256-263.
- 42 Conaway, R. C., Sato, S., Tomomori-Sato, C., Yao, T., Conaway, J. W. (2005) The mammalian Mediator complex and its role in transcriptional regulation. *Trends Biochem. Sci.* **30**, 250-5.
- 43 Asturias, F. J., Craighead, J. L. (2003) RNA polymerase II at initiation. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6893-6895.

- 44 Knuesel, M. T., Taatjes, D. J. (2011) Mediator and post-recruitment regulation of RNA polymerase II. *Transcr.* **2**, 28-31.
- 45 Meyer, K. D., Lin, S. C., Bernecky, C., Gao, Y., Taatjes, D. J. (2010) p53 activates transcription by directing structural shifts in Mediator. *Nat. Struct. Mol. Biol.* **17**, 753-760.
- 46 Malik, S., Roeder, R. G. (2010) The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* **11**, 761-72.
- 47 Bourbon, H. M. (2008) Comparative genomics supports a deep evolutionary origin for the large four-module transcriptional mediator complex. *Nucleic Acids Res.* **36**, 3993-4008.
- 48 Kim, S., Xu, X., Hecht, A., Boyer, T. G. (2006) Mediator is a transducer of Wnt/beta-catenin signaling. *J. Biol. Chem.* **281**, 14066-75.
- 49 Yeung, J., Esposito, M. T., Gandillet, A., Zeisig, B. B., Griessinger, E., Bonnet, D., So, C. W. (2010) β -Catenin mediates the establishment and drug resistance of MLL leukemic stem cells. *Cancer Cell.* **18**, 606-18.
- 50 Garapaty, S., Xu, C. F., Trojer, P., Mahajan, M. A., Neubert, T. A., Samuels, H. H. (2009) Identification and characterization of a novel nuclear protein complex involved in nuclear hormone receptor-mediated gene regulation. *J. Biol. Chem.* **284**, 7542-52.
- 51 Mahajan, M. A., Murray, A., Levy, D., Samuels, H. H. (2007) Nuclear receptor coregulator (NRC): mapping of the dimerization domain, activation of p53 and STAT-2, and identification of the activation domain AD2 necessary for nuclear receptor signaling. *Mol. Endocrinol.* **21**, 1822-34.
- 52 Mackintosh, C. (2004) Dynamic interactions between 14-3-3 proteins and phosphoproteins regulate diverse cellular processes. *Biochem. J.* **15**, 329-342.
- 53 Pane, F., Intrieri, M., Izzo, B., Quintarelli, C., Vitale, D., Migliorati, R., Sebastio, L., Salvatore, F. (2002) A novel MLL/AF4 fusion gene lacking the AF4 transactivating domain in infant acute lymphoblastic leukemia. *Blood.* **100**, 4247-4248.

FIGURE LEGENDS

Figure 1. Schematic representation of the flagged AF4 constructs used in this study. The various known protein domains are shown.

Figure 2. Expression of flagged proteins and affinity co-immunoprecipitations. **A.** Western blot analysis of recombinant AF4 polypeptides expressed in whole cellular extract (WCE) from HEK-293 cells. An extract from cells transfected with the empty 3xFlag vector was used as mock control. **B.** Proteins from total lysate of HEK-293 transfected respectively with Flag-AF4-1 were immunoprecipitated using anti-Flag agarose beads (IP FLAG). Flag-AF4-1 immunocomplex was in part separated on 10% SDS-PAGE gel (left panel, colloidal Coomassie staining) and in part immunoblotted with anti-Flag monoclonal antibody (right panel). In both cases, an immunoprecipitate from cells transfected with empty 3xFlag vector was used as mock control.

Figure 3. Classification of the identified proteins. **A.** According to Gene Ontology Cellular Localization and **B.** Biological Processes.

Figure 4. Flag-AF4 associates with CRSP3, MED7, ELL, CDK9, SIAH1, YWHAQ and YWHAE. **A.** Proteins from total lysate of HEK-293 cell lines transfected respectively with Flag-AF4, Flag-AF4-1 and empty vector (mock control) were immunoprecipitated (IP) using anti-Flag agarose beads, then resolved on 10% SDS-PAGE gel and analysed by western blotting with specific antibodies. **B.** Proteins from whole cell extract of HEK-293 cells transfected respectively with Flag-AF4 and empty vector (mock control) were immunoprecipitated (IP) using specific antibodies, then resolved on 8% SDS-PAGE gel and analysed by western blotting with anti-Flag antibody.

Figure 5. Endogenous AF4 associates with CRSP3, MED7, ELL, CDK9, SIAH1, YWHAQ and YWHAE. Proteins from total lysates of 697 cell lines were immunoprecipitated (IP) using anti-AF4, anti-CDK9, anti-YWHAQ, anti-CRSP3, anti-MED7, anti-ELL and anti-YWHAE. Immunocomplexes were resolved on 10% SDS-PAGE gel and analysed by western blotting with specific antibodies. WCE

represents control and IgG the mock. On the left of each WB panel, the molecular weight of the analyzed proteins is shown.

Figure 6. MS/MS spectra of modified and unmodified peptides of AF4.

A. MS/MS spectrum of ELSPLISLPSPVPPLSPIHSNQQpTLPR; **B.** MSMS spectrum of ELSPLISLPSPVPPLpSPIHSNQQTLPR; **C.** MSMS spectrum of ELSPLISLPSPVPPLSPIHSNQQTLPR; \diamond indicates the m/z signal of the parent ion.

Figure 7. Map of protein-protein interactions. All proteins identified in this study are reported in the map that was obtained by using the open-source software STRING. AF4 protein is not included. The purple lines indicates interactions proved by *in vitro* and/or *in vivo* assay and previously reported in the literature (PubMed); the blue lines show proteins that participate in the same complex, as reported in the well-annotated databases HPRD, Database of Interacting Proteins (DIP) and BioGrid (<http://www.thebiogrid.org/>). The physical distance between two proteins (nodes) along an edge in a graph has no meaning.

Figure 8. CDK9 interacts with the N-terminal region of AF4 (aa 2-277). Proteins from total lysate (WCE) of HEK-293 cell lines transfected respectively with three constructs coding for different regions of AF4-1 fragment (see Figure 1) and with empty vector (mock control) were immunoprecipitated (IP) using anti-Flag antibodies, resolved on 12% SDS-PAGE gel and analysed by western blotting with anti-CDK9 antibody.

Table 1. AF4-1 interacting proteins identified by nanoLC/ESI-MS/MS and their GO classification

Gene Symbol	Accession Number	Protein
-------------	------------------	---------

Transcription regulatory proteins:

PPARBP, MED1	gij2765322	Activator recruited cofactor 205kDa component
MED12	gij4827042	Mediator of RNA polymerase II transcription, subunit 12 homolog
CRSP2, MED14	gij4580326	Cofactor required for Sp1 transcriptional activation, subunit 2, 150kDa
CRSP3, MED23	gij28558969	Cofactor required for Sp1 transcriptional activation, subunit 3, 130kDa
THRAP4, MED24	gij8699628	Vitamin D receptor-interacting protein complex component DRIP100
CRSP6, MED17	gij28558975	Cofactor required for Sp1 transcriptional activation, subunit 6, 77kDa
CRSP7, MED26	gij28558977	Cofactor required for Sp1 transcriptional activation, subunit 7, 70kDa
RBBP4	gij13111851	Retinoblastoma binding protein 4
RUVBL2	gij13111851	48 kDa TATA box binding protein interacting protein
EAF1	gij27370592	ELL associated factor 1
MED4	gij7141320	p36 TRAP/SMCC/PC2 subunit (Mediator of RNA polymerase II transcription, subunit 4 homolog)
CRSP8, MED27	gij7141322	Cofactor required for Sp1 transcriptional activation, subunit 8, 34kDa
MED7	gij13528909	Cofactor required for Sp1 transcriptional activation, subunit 9
MED6	gij3329506	RNA polymerase transcriptional regulation mediator
MED8	gij33988564	Mediator of RNA polymerase II transcription subunit 8 Activator-recruited cofactor 32 kDa component
TRFP, MED20	gij4323033	Trf (TATA binding protein-related factor)-proximal homolog
MED21	gij1515377	RNA polymerase II holoenzyme component SRB7
ELL	gij10130023	RNA polymerase II elongation factor ELL
PCQAP	gij14043091	MED15
TCEA1	gij313223	Transcription elongation factor A (SII), 1
DNA-directed RNA polymerase proteins:		
POLR2A	gij36124	RNA polymerase II largest subunit
POLR2B	gij4505941	RNA polymerase II second largest subunit
POLR2C	gij2920711	Polymerase (RNA) II (DNA directed) polypeptide C, 33kDa
DNA binding proteins:		
MCM3	gij1552242	DNA polymerase alpha holoenzyme-associated protein P1 (p102 protein)
RUVBL1	gij15277588	49-kDa TATA box-binding protein-interacting protein
RNA binding proteins:		
NCL	gij189306	Nucleolin
PRPF31	gij40254869	PRP31 pre-mRNA processing factor 31 homolog
SNRPD2	gij29294624	Small nuclear ribonucleoprotein D2 polypeptide 16.5kDa
Serine/threonine kinase protein:		
CDK9	gij12805029	Cyclin-dependent kinase 9 (CDC2-related kinase)
Serine/threonine phosphatase protein:		
PPP2R1A	gij178663	Protein phosphatase 2 (formerly 2A) regulatory subunit A (PR 65) alpha isoform
Translation regulator protein:		
EEF1B2	gij12652911	Eukaryotic translation elongation factor 1 beta 1
Auxiliary transport protein:		
GDI2	gij285975	Rab GDP dissociation inhibitor beta

Receptor signaling complex scaffold proteins:

YWHAE	gij12655169	14-3-3 protein, epsilon isoform (Protein kinase C inhibitor protein 1)
YWHAQ	gij55594676	14-3-3 protein theta (14-3-3 protein T-cell)
SPIN	gij5730065	SPINL
AP3S1	gij4502861	Adapter-related protein complex 3 sigma 1 subunit

Signal transduction proteins

CCNT1	gij2981196	CDK9 associated C type cyclin
S100A8	gij21614544	S100 calcium-binding protein A8
S100A9	gij4506773	S100 calcium-binding protein A9

Transmembrane receptor protein tyrosine kinase protein:

FGFR2	gij29432	Fibroblast growth factor (FGR) receptor
-------	----------	---

Lipid kinase proteins:

PIP5K2B	gij1857637	Phosphatidylinositol-4-phosphate 5-kinase type II beta
PIP5K2C	gij21322230	Phosphatidylinositol-4-phosphate 5-kinase, type II, gamma

Heterotrimeric G-protein GTPase protein:

GNB4	gij12654119	Guanine nucleotide binding protein (G protein), beta polypeptide 4
------	-------------	--

Guanyl-nucleotide exchange factor protein:

ARHGEF4	gij8809845	Rho guanine nucleotide exchange factor (GEF) 4
---------	------------	--

Enzyme: Dehydrogenase

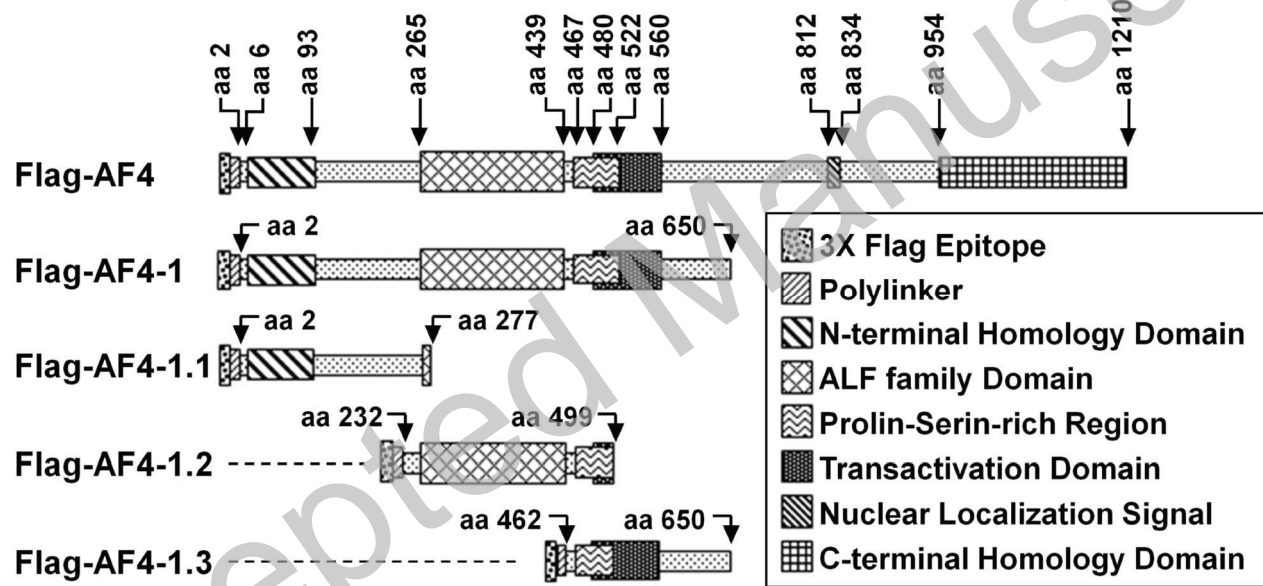
GAPDH	gij31645	Glyceraldehyde-3-phosphate dehydrogenase
-------	----------	--

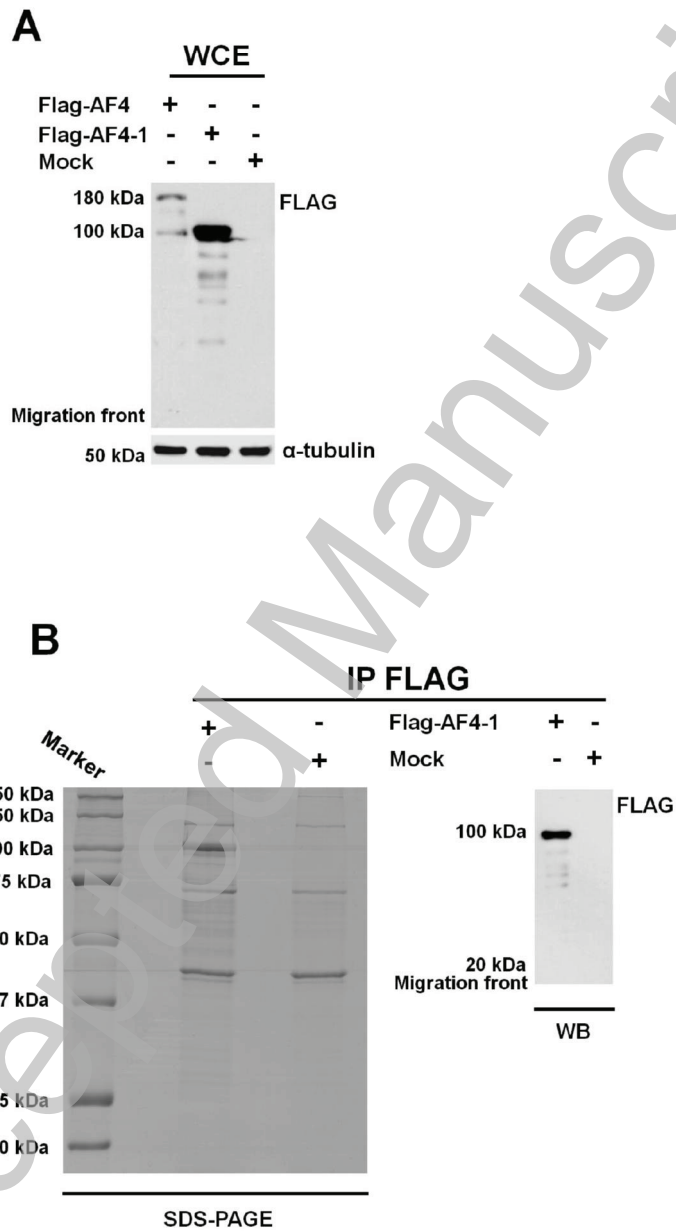
Enzyme: Synthase

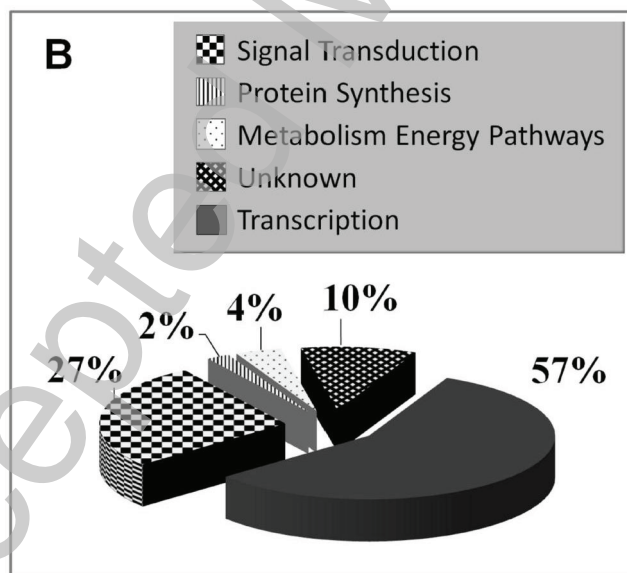
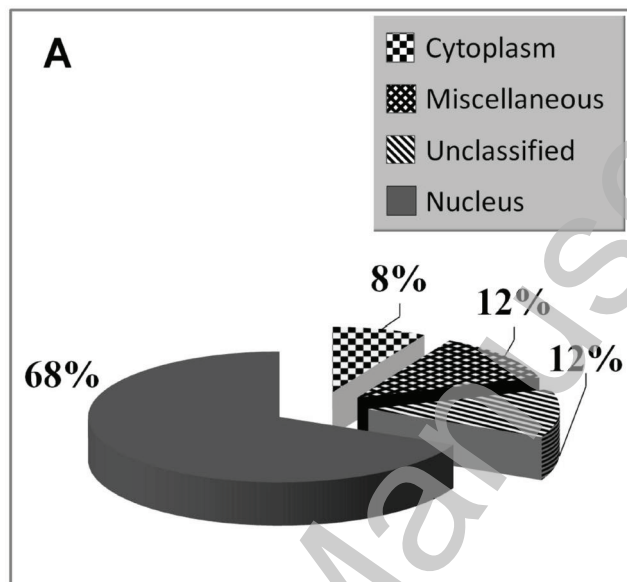
PTS	gij4506331	6-pyruvoyltetrahydropterin synthase
-----	------------	-------------------------------------

Unknown:

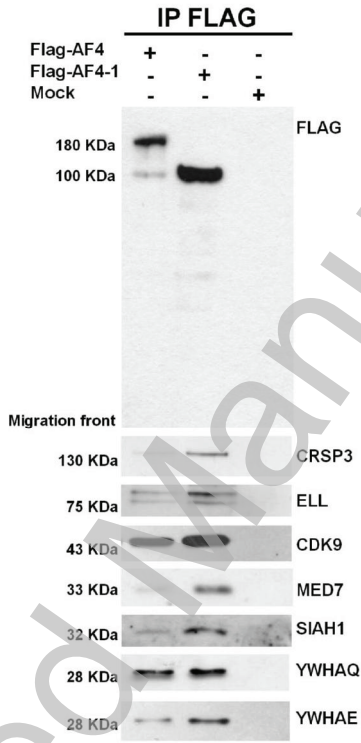
SMAD9	gij13959539	MAD mothers against decapentaplegic homolog 9
LSM14A	gij16877144	LSM14 homolog A (RNA-associated protein 55)
DRG1	gij17939479	Developmentally regulated GTP binding protein 1 (Neural precursor cell expressed developmentally down regulated 3)
HCCA2	gij55249549	HCCA2 protein (MOB2)
C20orf11	gij21594655	Chromosome 20 open reading frame 11







A



B

