Title: LEDGF/p75 TATA-less promoter is driven by the transcription factor Sp1.

Authors: Desfarges S, Abderrahmani A, Hernàndez-Novoa B, Munoz M, Ciuffi A

Journal: Journal of molecular biology

Year: 2011 Nov 25

Volume: 414

Issue: 2

Pages: 177-93

DOI: 10.1016/j.jmb.2011.10.010

Title: LEDGF/p75 TATA-less promoter is driven by the transcription factor Sp1

Authors: Sébastien Desfarges[1], Amar Abderrahmani[2], Beatriz Hernàndez-Novoa[1], Miguel Munoz[1] and Angela Ciuffi[1]

* Corresponding author: e-mail: angela.ciuffi@chuv.ch

Phone: +41 (0) 21 314 40 99; Fax: +41 (0) 21 314 40 95

[1] Institute of Microbiology (IMUL), University Hospital Center and University of Lausanne, Bugnon 48 – CHUV, CH-1011 Lausanne, Switzerland, [2] Department of Cellular and Molecular Biology, University Hospital Center and University of Lausanne, Bugnon 9, CH-1011 Lausanne, Switzerland.

The abbreviations used are: LEDGF, lens epithelium-derived growth factor; PSIP1, PC4 and SFRS1 interacting protein 1; TSS, transcription start site; HEK, human embryonic kidney; 5'RLM RACE, 5' RNA ligase-mediated rapid amplification of cDNA ends; ChIP, Chromatin immunoprecipitation and HIV, human immunodeficiency virus.

ABSTRACT

PC4 and SFRS1 interacting protein 1 (*PSIP1*) encodes two splice variants, lens epithelium-derived growth factor or p75 (LEDGF/p75) and p52. *PSIP1* gene products were shown to be involved in transcriptional regulation, affecting a plethora of cellular processes, including cell proliferation, cell survival, and stress response. Furthermore, LEDGF/p75 has implications for various diseases and infections, including autoimmunity, leukemia, embryo development, psoriasis and HIV integration. Here, we reported the first characterization of the *PSIP1*promoter. By 5' RACE approach, we identified novel transcription start sites in different cell types. Using a luciferase reporter system, we identified regulatory elements controlling LEDGF/p75 and p52 expression. These include i) minimal promoters, -112/+59 and +609/+781, driving the basal expression of LEDGF/p75 and the shorter splice variant p52 respectively, ii) a sequence (+319/+397) that may control the ratio between LEDGF/p75 and p52 expression, and iii) a strong enhancer (-320/-207) implicated in the modulation of LEDGF/p75 transcriptional activity. Computational, biochemical and genetic approaches enabled to identify the transcription factor Sp1 as a key modulator of the *PSIP1* promoter, controlling LEDGF/p75 transcription through two binding sites at -72/-64 and -46/-36.

Overall, our results provide the first data concerning the LEDGF/p75 promoter regulation giving new insights to further understand its biological function as well as opening the door of new therapeutic strategies in which LEDGF/p75 is involved.

Keywords: gene regulation, PSIP1, transcription start site, HIV, cancer.

INTRODUCTION

Lens epithelium-derived growth factor or p75 (LEDGF/p75) and its shorter form p52 are two ubiquitously expressed splice variants encoded by the *PSIP1* gene (*PC4 and SFRS1 interacting protein 1*) (Figure 1a), highly expressed in thymus, testis and brain, with LEDGF/p75 being the most abundant splice variant [1]. They were initially identified through cofractionation with the transcriptional positive co-activator PC4 from HeLa cell extracts [1]. As PC4, both LEDGF/p75 and p52 proteins were shown *in vitro* to be co-activators of the general transcriptional machinery, significantly enhancing transcription although with different intensities [2].

LEDGF/p75 and p52 share the same first 325 amino acids, mostly involved in chromatin association. This N-terminal region contains a conserved PWWP domain, two AT-hook motifs, three charged regions (CR) and a classical nuclear localization signal (NLS) (Figure 1a) [3; 4]. In contrast, the C-terminal portion of the two *PSIP1* gene products is different, thereby contributing to divergent specificities for interacting with different protein partners. Apart from its co-activator role, little is known about p52 [1]. Recent findings suggested that p52 overexpression was able to promote apoptosis in cancer cells [5], and to favor neurite growth and axonal elongation in neural cells [6].

In contrast, the second splice variant LEDGF/p75 has been more thoroughly investigated, however cellular roles of this protein are still incompletely characterized. LEDGF/p75 was described as a growth factor, a transcriptional co-activator, and a chromatin adapter, with implications in cell survival under stress-related conditions, homeobox gene regulation, autoimmunity, leukemia transformation, psoriasis pathology, and integration of the human immunodeficiency virus (HIV) DNA in the host genome [7; 8; 9; 10; 11; 12; 13; 14; 15]. These disparate cellular effects may reflect the variety of interacting protein partners. To date, these include JPO2/RAM2 protein [16; 17], the pogo transposable element with ZNF domain (pogZ) [18],

menin/MLL complex [15] and the Cdc7-activator of S-phase kinase (ASK) [19]. All these proteins bind specifically to the C-terminal domain of LEDGF/p75 that helps their recruitment to the chromosome, thereby implying a role for LEDGF/p75 as a general adaptor protein tethering diverse factors to chromatin. To date, among LEGDF/p75 interacting proteins, the one between LEDGF/p75 and the HIV integrase protein is the best characterized [20; 21; 22; 23; 24]. Upon entry into the host cell, HIV RNA genome is reverse transcribed into a linear double stranded DNA, which is subsequently imported into the nucleus and integrated into the genome of the host cell by the virally encoded integrase. The sites of HIV integration events are not random but display preferences for active transcription units [22]. The current model suggests that the chromatin-associated LEDGF/p75 recruits the incoming HIV preintegration complexes by direct binding to the viral integrase, thereby promoting integration to nearby genomic locations [25; 26; 27; 28].

Despite the increased interest in understanding the biological roles of LEDGF/p75 and p52, little is known about their expression regulation. Recently, Brown-Bryan et *al* showed that LEDGF/p75 and its shorter splice variant p52 are overexpressed, in tumor cells [5]. It has been proposed that LEDGF/p75 and p52 relative amounts could influence survival and cell death decisions under stress. Indeed, stress-related conditions, including serum starvation, oxidative stress, tumor necrosis factor alpha (TNF-α), heat shock and UV irradiation were shown to stimulate LEDGF/p75 expression, thereby improving survival of a wide range of cells types [29; 30]. The opposite effect has been observed in transforming growth factor β (TGF-β)-treated cells reducing LEDGF/p75 mRNA expression and stress-related LEDGF/p75 downstream genes [31]. In contrast to LEDGF/p75, overexpression of p52 induced apoptosis, and caspase-mediated cleavage of p52 generates a shorter fragment that interferes with the transactivation potential of the survival LEDGF/p75 protein in various tumor cell lines [5]. Understanding expression of *PSIP1* splice variants may help modulating the fate of cancer cells, either

survival or death, and therefore may provide attractive strategies to overcome tumor chemoresistance as well as reducing the tumorigenic potential of LEDGF/p75-overexpressing cells [10; 11; 32].

In the present study, *PSIP1* promoter truncations cloned in a luciferase reporter system were used to uncover key genomic regions involved in LEDGF/p75 expression. These include the identification of LEDGF/p75 and p52 minimal promoters, as well as cis-acting regions including the +319/+397 region potentially controlling LEDGF/p75 and p52 expression ratio. Computational and experimental analyses identified the transcription factor Sp1 as a key player modulating LEDGF/p75 expression. Understanding LEDGF/p75 expression regulation may help to characterize the biological function of LEDGF/p75 and provide attractive strategies for LEDGF/p75-related pathologies.

RESULTS

Identification of novel transcriptional start sites for LEDGF/p75 and p52.

According to the NCBI sequence database, the *PSIP1* promoter contains two transcription start sites (TSS), presumably one for each PSIP1 splice variant, LEDGF/p75 and p52 (Figure 1a). TSS positions for LEDGF/p75 (**NM_033222.3** and **NM_001128217.1**; NCBI) and p52 (**NM_021144.3**; NCBI) transcripts were initially determined by Ge and collaborators [1] and subsequently modified by the NCBI staff and collaborators upon further bioinformatic analyses. In addition to these reference transcripts, multiple alternative splice variants were reported, starting at various TSSs.

To characterize the TSS of *PSIP1* splice variants in our experimental system, *i.e.* in HEK cells, we used a 5'- RNA Ligase Mediated Rapid Amplification of cDNA Ends (5'-RLM-RACE) approach. *PSIP1* mRNAs were successfully amplified using specific primers (Table 1 and Figure 1b) located at the beginning of the coding sequence. Upon 5' end sequence analysis of 26 *PSIP1*-specific transcripts, we identified LEDGF/p75 TSS in our system (Figure 1c, +1), 36 bp downstream of the previously reported reference variant 3 (**NM_001128217.1**) (Figure 1c, black triangle), as well as the p52 TSS (Figure 1c, +745), 64 bp downstream of the previously reported reference variant (**NM_021144.3**) (Figure 1c, black triangle).

We also identified some minor TSSs (Figure 1c, arrows) indicating that, for both spliced variants, some sloppiness of the transcription initiation can occur. A similar 5'-RLM-RACE experiment was performed in SupT1 cells (T-cell line) and sequence analysis of 92 *PSIP1* transcripts confirmed the TSS positions determined previously in HEK cells.

Based on these data, we used the LEDGF/p75 5'end position newly identified in our experimental conditions as the major TSS reference, designated +1 (Figure 1c). Compared to this, p52 TSS corresponds to +745, and the ATG start codon of the coding sequence to +782.

Analysis of transcripts starting at +1 in HEK and SupT1 cells confirmed the presence of an intron (+154/+640) consistent with the one described in LEDGF/p75 transcripts (**NM_001128217.1**), (Figure 1c) [1].

In order to investigate the possible existence of endogenous LEDGF/p75 and p52 alternative transcripts starting at both TSSs, we performed reverse transcription using 3' primers specific for each transcript (Figure 1d). LEDGF/p75 specific transcripts were amplified by PCR using a primer close to the LEDGF/p75 TSS (Figure 1d, lane 1), while p52 specific transcripts were not (Figure 1d, lane 3), suggesting that in these conditions, no p52 transcript starting at the LEDGF/p75 TSS was detected, arguing for TSS-specific transcripts. As control, PCR amplification from the ATG to the transcript specific 3' end was also performed (Figure 1d, lanes 2 and 4).

All together, these experiments suggest that the position of the TSS determines the nature of the splice variant.

Identification of the PSIP1 functional promoter.

To uncover essential regions in *PSIP1* promoter sequence, we cloned *PSIP1* promoter truncations upstream of a *firefly* luciferase reporter gene. Promoter activity was assessed by luciferase expression in HEK cells. We first used the *PSIP1* promoter sequence ranging from -2043 to +781 (Figure 2a), thus containing both LEDGF/p75 and p52 TSSs, as well as the intron (+154/+640). Using luciferase-specific primers for 5'-RLM-RACE, we analyzed 49 clones. 32 clones identified a TSS at position +745 (with minor sloppiness), corresponding to the endogenous p52 TSS, and 17 clones identified a TSS at position +1 consistent with the

endogenous LEDGF/p75 TSS. Luciferase transcripts starting at LEDGF/p75 TSS revealed the absence of the +154/+640 sequence, consistent with intron splicing. No additional minor TSS was identified when using PSIP1-driven luciferase constructs. These data suggested that *PSIP1*-firefly luciferase constructs transfected in HEK cells recapitulated *PSIP1* endogenous transcriptional activity and alternative splicing, validating our experimental approach for subsequent analysis of *PSIP1* promoter activity by luciferase reporter assay.

Luciferase expression driven by *PSIP1* -2043/+781 promoter resulted in ~98x activity above background (Figure 2b and supplementary figure S1). Progressive *PSIP1* promoter truncations revealed that deleting sequences between -2043 and -723 did not affect significantly luciferase expression, as the *PSIP1* -723/+781 promoter construct showed a luciferase activity similar to the -2043/+781 one (Figure 2b and supplementary figure S1). In contrast, *PSIP1* -207/+781 promoter construct showed a luciferase activity of ~25x above background, a 4-fold reduction as compared to the -723/+781 construct ($p < 0.01$) (Figure 2b and supplementary figure S1), suggesting that the -723/+781 sequence contained the major determinants of the functional *PSIP1* promoter, and that the -723/-207 region may contain an enhancer.

LEDGF/p75 promoter activity is higher than p52 promoter activity.

To determine separately the expression of the two *PSIP1* splice variants and the minimal promoters driving their transcription, the -723/+781 fragment was split into two smaller fragments: -723/+59 and +140/+781 containing the LEDGF/p75 TSS and the p52 TSS respectively (Figure 2c and supplementary figure S1). Measurement of their respective luciferase activities revealed that promoter activity of the -723/+781 fragment was ~67x above background, while -723/+59 and +140/+781 displayed activities of 59x and 5x above background respectively, suggesting that *PSIP1* -723/+781 resulted from the additive

transcriptional activities initiated at each TSS. In addition, -723/+59 containing the LEDGF/p75 TSS was ~12 fold more efficient than the p52 TSS-containing +140/+781 region in driving luciferase expression, consistent with higher expression of LEDGF/p75 compared to p52 observed *in vivo* [5].

Identification of the LEDGF/p75 minimal promoter (-112/+59) and an enhancer region (-320/-207*).*

To identify regulatory elements driving LEDGF/p75 expression, finer progressive 5'deletions of the -723/+59 promoter fragment (containing only the TSS of LEDGF/p75) were generated and tested for luciferase activity (Figure 2d). Luciferase activity was similar in promoter constructs carrying progressive 5'deletions from -723 to -320, indicating that the -723/-320 fragment did not contain major regulatory elements. In contrast, the luciferase activity of the fragment -207/+59 strongly decreased (~2.5 fold, p value ≤ 0.01), consistent with the presence of a putative enhancer in the -320/-207 region.

A second drop in luciferase activity was observed with the -37/+59 construct below the threshold of detectable luciferase activity (1.5x above background, p value ≤ 0.001) as compared to the -112/+59 construct, suggesting that the -112/+59 fragment contained the minimal determinants required for LEDGF/p75 expression (Figure 2d, left panel). To further assess the role of the -112/+59 region, this region was deleted in the -320/+781 construct and tested for luciferase activity (supplementary figure S1). Deletion of this region totally abrogated LEDGF/p75 expression without affecting p52 expression (supplementary figure S1).

To validate the -320/-207 region as a functional enhancer, the -320/-207 fragment was cloned upstream a consensual TATA-box driving the expression of *firefly* luciferase (pGL4.23 vector) and compared the luciferase activity in presence or in absence of the enhancer (Figure

2d, right panel). Luciferase activity driven by a minimal promoter and the -320/-207 fragment was ~23 fold higher than the one driven by the minimal promoter only pGL4.23, $p < 0.001$), confirming that the -320/-207 region contained a functional enhancer.

Identification of regulatory elements in p52 promoter.

The luciferase activity driven by the *PSIP1* +140/+781 was 5x above background (Figure 2e). Progressive 5'deletions from +140 to +448 of the region containing p52 TSS showed similar levels of luciferase activity, suggesting that this region did not contain major determinants regulating p52 expression. However, deletion of the region +448/+548 and +548/+609 induced a moderate but significant decrease in luciferase activity (p value ≤ 0.01), suggesting that the region +448/+609 contained regulatory elements enhancing the transcription from p52 TSS (figure 2e). Although the luciferase activity driven by regions surrounding the p52 TSS was low, the +609/+781 region exhibited a low but significant activity (~1.7 fold expression above background, p value ≤ 0.001), indicative of the p52 minimal promoter.

Crosstalk between LEDGF/p75 and p52 promoter activities.

Independent analysis of LEDGF/p75 and p52 promoters revealed the presence of regulatory elements affecting the expression of each splice variant. However, *in vivo*, these two promoters overlap and may interfere with one another, probably leading to the observed difference between LEDGF/p75 and p52 expression. To investigate this possibility, we analyzed luciferase activity driven by various LEDGF/p75 promoter constructs (Figure 3). We first assessed the luciferase activity of constructs starting from -320 with progressive 3' deletions (Figure 3, top panel). Luciferase activity driven by the LEDGF/p75 TSS was not reproducibly different between promoter constructs ranging from -320/+59 to -320/+339. In contrast, the presence of the +339/+420 region strongly reduced luciferase activity up to

background level (Figure 3, top panel, construct -320/+420), suggesting that this region had a strong silencer activity on LEDGF/p75 TSS. Addition of region +521/+659 abolished the +339/+420 silencer effect and enhanced the overall luciferase activity of ~2 fold as compared to -320/+59 constructs, suggesting a dominant effect of this enhancer (Figure 3, top panel, construct -320/+659). Finally, the -320/+781 construct, containing both LEDGF/p75 and p52 TSS showed a reduced luciferase activity as compared to the -320/+659 construct, but a similar activity as compared to the -320/+59 construct (Figure 3, top panel), suggesting that the +659/+781 region inhibits the enhancer effect of the +521/+659 region.

We then assessed the luciferase activity of similar constructs but starting at -207, thus without the -320/-207 enhancer region (Figure 3, middle panel). The profile of luciferase activity were similar to the -320 constructs but at lower levels of activity (~2.5 fold reduction) due to the absence of the -320/-207 enhancer region defined previously.

The luciferase activity of constructs starting at -112 (Figure 3, bottom panel) showed again a similar profile of luciferase activities than the previous -320 and -207 constructs, but with higher activities than the -207 series and lower activities than the -320 series, confirming the -207/-112 silencer region and the dominant -320/-207 enhancer.

All together, these data are consistent with binding of regulatory elements to *PSIP1* promoter regions located between LEDGF/p75 and p52 TSS, thereby modulating their relative expression.

Expression of LEDGF/p75 and p52 is mediated by the transcription factor Sp1.

Analysis of the sequences surrounding LEDGF/p75 and p52 TSSs did not reveal any consensual TATA or TATA-like box sequence, thereby classifying *PSIP1* promoter as TATA-less. TATA-less promoters are characterized by the presence of multiple TSSs, CpG

islands and Sp1 transcription factor binding sites [33]. Alibaba2.1 (TRANSFAC) and TESS search tools were used to analyze LEDGF/p75 and p52 minimal promoter sequences and identify transcription factor binding sites. *PSIP1* sequence analysis identified multiple putative Sp1 binding sites, characterized by GC-rich content, and consistent with TATA-less promoters (Figure 4a).

To investigate the role of Sp1 in *PSIP1* promoter activity, luciferase activity driven by LEDGF/p75 -723/+59 and p52 +140/+781 promoters was first assessed in presence or absence of mithramycin (Figure 4b, top panels). Mithramycin binds GC-rich motifs, and is widely used as an inhibitor of Sp1–DNA binding [34; 35]. As expected for Sp1-bound promoters, mithramycin impaired luciferase expression driven by both LEDGF/p75 and p52 promoters, in a dose-dependent manner (Figure 4b, top panels). To confirm the involvement of Sp1 in *PSIP1* promoter activity, luciferase activities were assessed in HEK cells expressing increasing amounts of the transcription factor Sp1 (Figure 4b, bottom panels). Increasing amounts of Sp1 were associated with increased luciferase activities, for both LEDGF/p75 and p52 promoter constructs. Increased luciferase activities also correlated with enhanced expression of endogenous LEDGF/p75 and p52 (Figure 4c), further arguing for a role of Sp1 in modulating *PSIP1* promoter activity.

In order to investigate whether Sp1-mediated *PSIP1* promoter regulation was due to a direct Sp1 binding to *PSIP1* promoter sequences and not to an indirect effect, we performed chromatin immunoprecipitation (ChIP) with anti-Sp1 antibody followed by PCR amplification of -112/+59 *PSIP1* promoter sequence. As shown in Figure 4d, anti-Sp1 ChIP specifically captured the -112/+59 LEDGF/p75 promoter sequence, which could be amplified by specific PCR. In contrast, no *PSIP1*-specific PCR amplification was detected when using a rabbit normal serum for ChIP or in presence of mithramycin, the Sp1-DNA binding inhibitor. All together, these data are consistent with a role of Sp1 in modulating *PSIP1* transcription.

Identification of two Sp1 binding sites in modulating LEDGF/p75 promoter activity.

To identify more specifically which putative Sp1 binding sites are essential for Sp1 activity, mutagenesis was carried out on the -112/+59 region in order to reduce the GC content of these sequences (Figure 4a and Table 1) in the context of -723/+59 and tested for luciferase activity (Figure 5a).

Mutations of site 2 (CAT**TA**C**AA**C for mSp1 site 2 instead of wt CATCCCCCC) and site 3 (G**A**TC**A**C**TA**C**A**C for mSp1 site 3 instead of wt GGTCGCGCCCC), individually or combined, reduced significantly LEDGF/p75 promoter activity whereas site 1 and the multiple sites 4 individually or in combination did not alter significantly luciferase activity (Figure 5a and supplementary figure S2).

To further investigate the role of these Sp1 binding sites 2 and 3, luciferase activity was assessed in presence of increasing amounts of Sp1 (Figure 5b). As expected, increasing amounts of Sp1 stimulated LEDGF/p75 wt promoter activity and thus luciferase expression. In contrast, no similar stimulation was observed in presence of increasing amounts of Sp1 when the *PSIP1* promoter was mutated for one or two Sp1 binding sites (Figure 5b). Furthermore, transcript analysis by 5' RLM-RACE showed that, in addition to lower transcript quantity, transcript quality was also affected, with transcription starting further upstream, in the -320/-207 enhancer region (data not shown).

Thus, Sp1 overexpression required both wt Sp1 binding sites located at -72/-64 and -46/-36 of LEDGF/p75 promoter for complete activity.

DISCUSSION

Although LEDGF/p75 has many implications in cell survival and has been involved in many diseases, such as cancer, psoriasis, autoimmunity and HIV, little is known about its expression regulation. Here, we reported for the first time key elements involved in *PSIP1* gene regulation, more particularly regulation of the PSIP1/LEDGF/p75 splice variant.

By 5' RLM-RACE approach of endogenous LEDGF/p75 and p52 as well as luciferase constructs, we identified the transcription start site of LEDGF/p75 and p52 transcripts to be located at 36 bp and 64 bp downstream of the NCBI reference respectively, in our experimental system. Furthermore, RACE and RT-PCR experiments provided evidence that p52 transcripts initiated at p52 TSS but not at LEDGF/p75 TSS, suggesting that LEDGF/p75 and p52 could be two overlapping genes with two distinct promoters that could impact on each other rather than two splice variants.

Computational analysis of *PSIP1* promoter revealed the absence of TATA box consensus. Consistent with TATA-less promoters, TSS sloppiness can be observed, as implied by the variety of PSIP1 mRNA sequences deposited to NCBI nucleotide sequence database [36; 37].

Transcription of TATA-less promoters are mostly driven by the Sp1 transcription factor [38]. Consistently, we identified putative Sp1 binding sites in both LEDGF/p75 and p52 promoters. Modulation of Sp1 activity by overexpression or using a Sp1–DNA binding inhibitor confirmed the role of Sp1 in regulating *PSIP1* transcription. Further investigation of the essential -112/+59 region on the LEDGF/p75 promoter by mutagenesis revealed two functional Sp1 binding sites, sites 2 (-72/-64) and 3 (-46/-36). The impact on transcription was higher for the mutation on site 3 with a two-fold decrease of LEDGF/p75 promoter activity compared to the WT. Mutations on both sites 2 and 3 decreased at least 5 fold the LEDGF/p75 promoter activity. However, mutation on both Sp1 binding sites did not totally abrogate

LEDGF/p75 promoter activity (in contrast to the deletion of the region -112/+59), suggesting that additional factors may be involved in transcription regulation.

The nuclear protein Sp1 belongs to a growing family of transcription factors that modulate gene expression [39]. Sp1 binds specifically to DNA and to the cofactor required for Sp1 (CRSP) complex in order to initiate transcription, via recruitment of TATA binding protein-associated factors [40]. Sp1-mediated transcription has been implicated in the growth and metastasis of cancer cell lines [41], and has been shown to be overexpressed in tumor cells [42; 43]. Similarly, and consistent with our data of Sp1-mediated transcription of *PSIP1*, several studies reported that LEDGF/p75 is overexpressed in cancer cells compared to normal tissue [5]. Furthermore, in pancreatic cancer cells, Sp1 stimulates expression of the vascular endothelial growth factor (VEGF) [42; 44]. Recently, LEDGF/p75 was shown to modulate the expression of VEGF [7; 45]. Our data may suggest an additional and indirect mechanism of VEGF expression regulation, in which Sp1 increases LEDGF/p75 expression, which in turn stimulates VEGF expression.

PSIP1 promoter contains two overlapping promoters, each driving the expression of *PSIP1* splice variants, LEDGF/p75 and its shorter isoform p52. Artificial dissociation of the overlapping LEDGF/p75 and p52 promoters identified a strong dominant enhancer in the -320/-207 region and a silencer in the -207/-112 region affecting Sp1-mediated transcription of LEDGF/p75, and an enhancer in the +448/+609 region affecting p52 transcription (Figure 6). However, *in vivo*, the two promoters are tightly intertwined and impact each other, thus affecting the expression of both splice variants as observed in previous work [5] and in the present study, and keeping constant the uneven expression ratio between LEDGF/p75 and p52. Consistent with this model, the +339/+420 region was able to completely shut down the expression driven by LEDGF/p75 TSS specifically, effect that could be relieved by the +521/+659 region (that overlaps partially with the +448/+609 p52 enhancer). This +521/+659 region was also able to increase ~2 fold LEDGF/p75 transcriptional activity, providing a

positive loop of regulation. In presence of the +659/+781 region surrounding the p52 TSS, the ~2 fold enhancer effect induced by the +521/+659 region is inhibited. The detailed mechanism of action of these regulatory loops is not yet completely understood but might involve cooperation between Sp1 and other transcription factors bound on regulatory elements of the promoter. Direct action through transcription factor interactions and/or through chromatin rearrangement may explain the complex interplay between the two overlapping promoters of *PSIP1* splice variants, thereby allowing fine regulation of LEDGF/p75 and p52 transcription.
Further studies should provide additional details and help further understanding of the molecular mechanisms regulating LEDGF/p75 and p52 expression, thereby providing new therapeutic strategies to treat LEDGF/p75-related pathologies.

This present study identified the first pieces of the complex mechanism of the LEDGF/p75 and p52 interplay regulation. Further studies should provide additional details and help further understanding these molecular mechanisms that should bring new strategies to modulate LEDGF/p75 and p52 expression for LEDGF/p75-related pathologies.

MATERIALS AND METHODS

Cell culture, reagents and antibodies.

HEK (human embryonic kidney) cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 2 mM glutamax (Invitrogen, CA, USA), 10% heat inactivated fetal calf serum and antibiotics (1% penicillin / streptomycin). SupT1 cells were cultivated in RPMI-1640 medium supplemented with 2 mM glutamax, 10% heat inactivated fetal calf serum and antibiotics (1% penicillin / streptomycin).

Mithramycin (Sigma-Aldrich, MO, USA), a Sp1 binding inhibitor, was directly added to cell culture medium at final concentrations ranging from 50 nM to 800 nM.

Antibodies used for western blot analyses included goat anti-Sp1 (PEP-2) antibody (sc-59G; Santa Cruz Biotechnology, CA, USA), mouse anti-PSIP1 antibody recognizing both LEDGF/p75 and p52 proteins (clone 26; BD Biosciences, NJ, USA), mouse anti-α-tubulin antibody (clone B-5-1-2; Sigma-Aldrich, MO, USA), as well as horseradish peroxidase-conjugated anti-mouse and anti-goat secondary antibodies (P0260 and P0160 respectively; DakoCytomation, Denmark). Working dilutions were 1/300, 1/300, 1/1000 respectively for the primary antibodies and 1/2000 for the secondary antibodies.

5' RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE).

Total RNA was extracted using Illustra RNA Spin mini kit (GE Healthcare, UK) following manufacturer's guidelines. 2 µg of total RNA were used for 5′ RLM-RACE using GeneRacer (Invitrogen, CA, USA), according to manufacturer's instructions with some modifications specified hereafter. Briefly, 5′-uncapped RNAs (immature transcripts) were dephosphorylated by the calf intestinal phosphatase to avoid their subsequent contamination in the procedure. Subsequently, mature mRNAs were treated with Tobacco acid pyrophophatase for 5' cap removal, ligated to the GeneRacer™ RNA oligo and reverse transcribed with the Superscript II

using a gene-specific 5'-biotinylated reverse primer (for LEDGF/p75 and p52: 5'-biotin-CGAGCTGGCCAATGGGGATAACC-3'; for luciferase: 5'-biotin-CGGTCCCGTCTTCGAGTGGGTAGAATGG -3'). Biotinylated cDNAs were diluted in 400 µl of Bind and Wash 2X buffer **(**10 mM Tris-HCl pH7.5, 1mM EDTA, 2M NaCl**)** and incubated at 4°C for 1 hour with 400 µl streptavidin-coupled magnetic beads (Dynal, Invitrogen, CA, USA). After 3 washes with 500 µl of Bind and Wash 1X buffer, cDNAs coupled beads were resuspended in 30 µl of nuclease free water. A first PCR reaction was performed on 3 µl of bead-captured cDNAbeads with 2.5 U PfuTurbo™ polymerase (Stratagene, CA, USA), 600 nM GeneRacer™ 5' forward primer (5'-CGACUGGAGCACGAGGACACUGA-3'), 200 nM gene-specific reverse primer (5'-CGAGCTGGCCAATGGGGATAACC-3' for LEDGF/p75 and p52 or 5'-CGGTCCCGTCTTCGAGTGGGTAGAATGG-3' for luciferase), 200 µM each dNTP, in 50 µl final volume. PCR cycling conditions were: 5 min at 95°C, 35 cycles of 30 sec at 95°C-30 sec at 55°C – 5 min at 68°C, and a final 10 min at 68°C. 1 µl of this first PCR reaction was used for a nested PCR with 5 U Herculase2™ polymerase (Stratagene, CA, USA), 200 nM GeneRacer™ 5' nested primer (5'-GGACACUGACAUGGACUGAAGGAGUA-3'), 200 nM gene-specific nested primer (5'-CATCTTGGCGAAGATGAGGTCTCC-3' for LEDGF/p75 and p52 or 5'-TTTTTTCTCGAGGTTTCGGGGGCGAGACCGGG-3' for luciferase), 200 µM each dNTP, in 50 µl final volume. PCR cycling conditions were: 5 min at 95°C, 35 cycles of 30 sec at 95°C-30 sec at 55°C – 2 min at 72°C, and a final 10 min at 72°C. PCR products were separated by gel electrophoresis and purified using Invisorb gel extraction kit (Invitek, GmbH). Purified PCR products were incubated with 5 U Taq DNA polymerase (Invitrogen, CA, USA) at 72°C for 15 min, allowing the adjunction of a single deoxyadenosine (A) to the 3´ ends, before cloning into pCR4-TOPO TA vectors (TOPO TA Cloning Kit, Invitrogen, CA, USA) and transformation into TOP10 bacteria according to manufacturer's instructions. Plasmid

DNA was isolated from individual colonies and sequenced using the provided M13F and M13R primers.

Construction of PSIP1 promoter truncations in firefly luciferase reporter vectors.

PSIP1 promoter (-2043/+781 bp) was amplified by PCR using BAC clone RP11-211N10 (BACPAC Resources, Children's Hospital Oakland Research Institute) as template, with primers MA.pr-110 and MA.pr-124 (Table 1) and Herculase II Fusion DNA polymerase® (Stratagene, CA, USA) according to instructions, and subcloned into TOPO-blunt vector (Invitrogen, CA, USA). This plasmid was used as template for subsequent PCR reactions, allowing the construction of a collection of *PSIP1* promoter truncations (table 1). PCR reactions were performed with Pfu Turbo polymerase® according to the manufacturer's conditions and supplemented with 5% DMSO (Stratagene, CA, USA). All the primers used to generate the *PSIP1* truncations were tailed with KpnI site (forward primers) or XhoI site (reverse primers). Upon gel purification with Invisorb gel extraction kit (Invitek, Germany), PCR products were digested with KpnI and XhoI (New England Biolabs, MA, USA), cleaned on PCRapace columns (Invitek, Germany) and inserted into the pGL4.10 basic vector (Promega, Madison, WI), upstream of the *firefly luciferase* coding sequence. All the constructions were checked by restriction analysis and sequencing.

Determination of promoter activity by dual-luciferase reporter assay.

HEK cells (100,000 cells) were plated in 500 µl of culture medium in a 24-well plate and allowed to grow for one day to reach approximately 50-60% confluence for transfection. DNA transfection (1 µg) was performed using the cationic polymer jetPEI™ reagent according to the manufacturer's procedure (Polyplus-transfection, France). *Firefly* luciferase reporter constructs, driven by *PSIP1* promoter fragments (250 ng), were co-transfected with 5 ng of *Renilla*

luciferase reporter construct (pGL4.74), driven by the thymidine kinase promoter (Promega, Madison, WI) to normalize for transfection efficiency. An empty vector, pCI, was added to the transfection mix to reach the 1 µg total DNA required for optimal transfection efficiency.

For inhibition assays, mithramicin (0-800 nM) was added to the cells 1h prior transfection and then kept throughout the experiment.

For Sp1 overexpression assays, increasing amounts of pCMV-Sp1 encoding plasmid (0-750 ng) was added to the luciferase transfection mixture, and completed to 1 µg total DNA with pCI as described above.

Cells were collected 40h post-transfection, lysed in 150 µl of 1X Passive Lysis Buffer for 10 min at room temperature, and frozen at -20°C. Luminescence intensity was measured by dual luciferase assay using 5 µl cell lysate and according to manufacturer's recommandations (Promega, Madison, WI). Briefly, *firefly* luciferase activity was first measured by adding 25µl LARII reagent (*firefly* luciferase substrate), followed by the adjunction of 25 µl of Stop&Glo reagent, allowing for measurement of *Renilla* luciferase activity. Luminescence intensity, reflecting luciferase activity, was measured as relative light units (RLU) with the Lumat LB 9507 luminometer (Berthold technologies, Germany).

PSIP1 promoter activity (*i.e. firefly* luciferase activity) was normalized by the transfection efficiency (*i.e. Renilla* luciferase activity). The empty pGL4.10- vector (without promoter) was included in all experiments to determine the background level of *firefly* luciferase activity. *PSIP1* promoter activity was graphed as fold activity above background, and was calculated as following: (ratio *PSIP1* firefly/renilla) / (ratio pGL4.10 firefly/renilla), unless specified otherwise.

For experiments with Sp1 overexpression, normalizations were performed by total protein concentration (Pierce® BCA Protein Assay Kit; Thermo scientific, MA, USA). 25 µl of sample diluted in 0.1X PBS (as recommended by the manufacturer;

http://www.promega.com/enotes/applications/ap0047_tabs.htm) was used for the assay. Measured RLU of each sample was divided with the background level (*firefly* luciferase activity of pGL4.10 / total proteins concentration of pGL4.10-transfected cells).

Detection of LEDGF/p75 transcripts by RT-PCR.

Total RNA from HEK cells ($5x10^6$ cells) was extracted using Illustra RNAspin mini RNA isolation kit (GE Healthcare, UK), followed by selection of polyA+ RNA using OligoTex (Qiagen) according to manufacturers' instructions. Reverse transcription was performed using 150 ng polyA+ RNA, in presence of 500 nM LEDGF/p75 or p52 specific 3' primers, using the High Fidelity cDNA synthesis kit (Applied Biosystems), followed by standard PCR (HotStar Taq master mix, Qiagen) performed with specific primers as described in Figure 1d.

Site-directed mutagenesis of Sp1 putative binding sites.

The plasmid pGL4.10 containing the full length *PSIP1* promoter (*i.e.* from -2043 to +781 bp) was used as template for site-directed mutagenesis. Mutations of putative Sp1 binding sites were performed by GC content modification. Briefly, mutations in each putative Sp1 binding site were performed by PCR using 3.5 U Pfu Turbo polymerase®, 5% DMSO, 200 µM each dNTP and 200nM of each primer. Primers MA.pr-280 and MA.pr-281 were used for the mutagenesis of Sp1 site 1, MA.pr-282 and MA.pr-283 for Sp1 site 2, MA.pr-284 and MA.pr-285 for Sp1 site 3, and MA.pr-365 and MA.pr-366 for Sp1 site 4 (Table 1). PCR cycling conditions were 1 cycle at 95°C for 2 min, 18 cycles at 95°C for 30 sec55°C for 30 sec-68°C for 10 min, and 1 cycle at 68°C for an additional elongation time of 10 min. Upon amplification, PCR reactions were supplemented with 10 U DpnI and incubated at 37°C for 1h to digest the methylated parental plasmid (template), and finally transformed in XL1-blue

supercompetent cells (Stratagene, CA, USA). Mutagenesis efficiency was checked by sequence analysis.

Western blots analysis.

30 µg of total proteins in PLB 1X were mixed with GeBa sample buffer (Gene Bio-Application, Israel)) and heated for 5 min at 95°C. Samples and the Color Plus™ prestained protein ladder (New England Biolabs, MA, USA) were separated by electrophoresis using pre-cast 8-16% polyacrylamide gel (Gene Bio-Application, Israel) for 1h at 160V and transferred onto nitrocellulose membrane (Whatman, Germany) for 1h at 400 mA in TBT-methanol buffer (1.9 M glycine, 0.25 M Tris, 20% methanol, pH 8.6). Immunoblotting was performed using the SNAP-i.d. protein detection system (allowing solutions to pass through the membranes by vacuum aspiration), according to the manufacturer's recommendations (Merck-Millipore, Billerica, MA). Briefly, pre-wet membranes were blocked with 15 ml of 0.1% non-fat milk solution , incubated for 10 min at room temperature in primary antibody diluted in 0.1% non-fat milk, followed by three washes with PBS 0.1% Tween®, incubated similarly with the secondary antibody. The membrane was incubated with the LiteAblot chemiluminescent substrate (Euroclone, Italy) for 2 mins and revealed on autoradiographic Amersham hyperfilm™ MP (GE Healthcare, UK).

Chromatin immunoprecipitation (ChIP).

Chromatin immunoprecipitation was performed using the MAGnify chromatin immunoprecipitation system (Invitrogen, CA, USA), according to the manufacturer's protocol with some modifications. Briefly, HEK cells ($2x10^6$) were transfected with the *firefly* luciferase reporter plasmid containing the *PSIP1* promoter from -723 bp to +59 bp, with or without mithramycin treatment (200 nM). At 30h post-transfection, cells were crosslinked in 1% final

formaldehyde for 10 min at 37°C, and the reaction was stopped by adding 1.25M glycine for 5 min at 4°C. Crosslinked cells were washed twice in ice cold PBS and collected by centrifugation for 5 min at 700g. The cell pellet was resuspended with 500 µl SDS lysis buffer (1% SDS, 10mM EDTA, 50 mM Tris pH 8.1), supplemented with protease inhibitors (protease inhibitor cocktail P8340; Sigma-Aldrich, St Louis, MO), and incubated on ice for 10 min. Cell lysates (100 µl) were diluted 5 fold in PBS and sonicated using a Soniprep 150 (N.Zivy & Co, Switzerland) for 5 cycles of 10 sec at maximum potency followed by 1 min incubation at 4°C), generating chromatin sheared fragments between 250 and 1500 bp. For each condition, 2µg of ChIP-grade Sp1 antibody (Abcam, UK) or normal rabbit IgG (provided in the kit as control) were mixed with 200µl Dilution Buffer allowed to bind to protein A and G conjugated magnetic beads, and used to immunoprecipitate chromatin for 2h at 4°C. After several washes, reverse crosslink and purification, immunoprecipitated DNA (2µl) was amplified using 400nM MA.pr-121 and MA.pr-130 (Table 1) and 22 µl AccuPrime™ *Pfx* SuperMix (Invitrogen, CA, USA). As controls, DNA inputs (before ChIP) were amplified similarly. The PCR reaction was performed under the following conditions: 1 cycle 95°C for 5 min, 30 cycles at 95°C 30sec, 55°C 30sec and 68°C 1min, and 1 cycle 68°C 10 min. PCR amplification was checked by gel electrophoresis using 5µl of the PCR reaction.

Statistical analysis.

Data are presented as mean ± SEM. Differences between groups were assessed by two-sided student t- test. * = p value ≤ 0.1; ** = p value ≤ 0.01; *** = p value ≤ 0.001. One way ANOVA statistical analysis was performed for experiments using mithramycin and Sp1 overexpression. All statistical analyses were performed using GraphPad software Prism™ v5 (GraphPad Software, CA, USA).

Bioinformatic analysis of PSIP1 promoter sequence.

To identify putative transcription factors binding sites on LEDGF/p75 (-112/+59 bp) and p52 (+609/+781 bp) core promoters, the software Alibaba 2.1 (http://www.gene-regulation.com/pub/programs/alibaba2/index.html) was used to screen the TRANSFAC 4.0 database. The analyses were carried out under strict parameters (matrix conservation 80% and matrix width of 10 bp).

AUTHORS' CONTRIBUTIONS

SD and AC designed the experiments and wrote the paper. SD performed the experiments. AC conceived and led the project, MM provided technical help and AA edited and proofread the manuscript.

ACKOWLEDGEMENTS

We are grateful to Dr. Thierry Roger of the University Hospital Center and University of Lausanne for the generous gift of the plasmid coded for the transcription factor Sp1. We thank Damien Nicolas for his technical help during this study. This work was supported by the Swiss National Science Foundation (3100A0-120553), the federal program "Egalité" and the Novartis Foundation.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

1. Ge, H., Si, Y. & Roeder, R. G. (1998). Isolation of cDNAs encoding novel transcription coactivators p52 and p75 reveals an alternate regulatory mechanism of transcriptional activation. *Embo J* **17**, 6723-9.
2. Ge, H., Si, Y. & Wolffe, A. P. (1998). A novel transcriptional coactivator, p52, functionally interacts with the essential splicing factor ASF/SF2. *Mol Cell* **2**, 751-9.
3. Llano, M., Morrison, J. & Poeschla, E. M. (2009). Virological and cellular roles of the transcriptional coactivator LEDGF/p75. *Curr Top Microbiol Immunol* **339**, 125-46.
4. Engelman, A. & Cherepanov, P. (2008). The Lentiviral Integrase Binding Protein LEDGF/p75 and HIV-1 Replication. *PLoS Pathog* **4**, e1000046.
5. Brown-Bryan, T. A., Leoh, L. S., Ganapathy, V., Pacheco, F. J., Mediavilla-Varela, M., Filippova, M., Linkhart, T. A., Gijsbers, R., Debyser, Z. & Casiano, C. A. (2008). Alternative splicing and caspase-mediated cleavage generate antagonistic variants of the stress oncoprotein LEDGF/p75. *Mol Cancer Res* **6**, 1293-307.
6. Zhao, H. S., Chen, S. J., Wu, N., Wang, X. Q., Yin, Z. Q. & Wang, Y. (2008). LEDGFp52 controls rat retinal ganglion cell neurite growth in culture and regulates specific neuronal growth-associated genes and protein production. *J Int Med Res* **36**, 815-29.
7. Cohen, B., Addadi, Y., Sapoznik, S., Meir, G., Kalchenko, V., Harmelin, A., Ben-Dor, S. & Neeman, M. (2009). Transcriptional regulation of vascular endothelial growth factor C by oxidative and thermal stress is mediated by lens epithelium-derived growth factor/p75. *Neoplasia* **11**, 921-33.
8. Ganapathy, V. & Casiano, C. A. (2004). Autoimmunity to the nuclear autoantigen DFS70 (LEDGF): what exactly are the autoantibodies trying to tell us? *Arthritis Rheum* **50**, 684-8.

9. Matsui, H., Lin, L. R., Singh, D. P., Shinohara, T. & Reddy, V. N. (2001). Lens epithelium-derived growth factor: increased survival and decreased DNA breakage of human RPE cells induced by oxidative stress. *Invest Ophthalmol Vis Sci* **42**, 2935-41.

10. Mediavilla-Varela, M., Pacheco, F. J., Almaguel, F., Perez, J., Sahakian, E., Daniels, T. R., Leoh, L. S., Padilla, A., Wall, N. R., Lilly, M. B., De Leon, M. & Casiano, C. A. (2009). Docetaxel-induced prostate cancer cell death involves concomitant activation of caspase and lysosomal pathways and is attenuated by LEDGF/p75. *Mol Cancer* **8**, 68.

11. Rammer, P., Groth-Pedersen, L., Kirkegaard, T., Daugaard, M., Rytter, A., Szyniarowski, P., Hoyer-Hansen, M., Povlsen, L. K., Nylandsted, J., Larsen, J. E. & Jaattela, M. (2010). BAMLET activates a lysosomal cell death program in cancer cells. *Mol Cancer Ther* **9**, 24-32.

12. Singh, D. P., Ohguro, N., Chylack, L. T., Jr. & Shinohara, T. (1999). Lens epithelium-derived growth factor: increased resistance to thermal and oxidative stresses. *Invest Ophthalmol Vis Sci* **40**, 1444-51.

13. Sutherland, H. G., Newton, K., Brownstein, D. G., Holmes, M. C., Kress, C., Semple, C. A. & Bickmore, W. A. (2006). Disruption of Ledgf/Psip1 results in perinatal mortality and homeotic skeletal transformations. *Mol Cell Biol* **26**, 7201-10.

14. Takeichi, T., Sugiura, K., Muro, Y., Matsumoto, K., Ogawa, Y., Futamura, K., Kaminuma, O., Hashimoto, N., Shimoyama, Y., Saito, H. & Tomita, Y. (2010). Overexpression of LEDGF/DFS70 Induces IL-6 via p38 Activation in HaCaT Cells, Similar to that Seen in the Psoriatic Condition. *J Invest Dermatol*.

15. Yokoyama, A. & Cleary, M. L. (2008). Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell* **14**, 36-46.

16. Bartholomeeusen, K., De Rijck, J., Busschots, K., Desender, L., Gijsbers, R., Emiliani, S., Benarous, R., Debyser, Z. & Christ, F. (2007). Differential Interaction of HIV-1 Integrase and JPO2 with the C Terminus of LEDGF/p75. *J Mol Biol* **372**, 407-21.

17. Maertens, G. N., Cherepanov, P. & Engelman, A. (2006). Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. *J Cell Sci* **119**, 2563-71.

18. Bartholomeeusen, K., Christ, F., Hendrix, J., Rain, J. C., Emiliani, S., Benarous, R., Debyser, Z., Gijsbers, R. & De Rijck, J. (2009). Lens epithelium-derived growth factor/p75 interacts with the transposase-derived DDE domain of PogZ. *J Biol Chem* **284**, 11467-77.

19. Hughes, S., Jenkins, V., Dar, M. J., Engelman, A. & Cherepanov, P. (2010). Transcriptional co-activator LEDGF interacts with Cdc7-activator of S-phase kinase (ASK) and stimulates its enzymatic activity. *J Biol Chem* **285**, 541-54.

20. Cherepanov, P., Devroe, E., Silver, P. A. & Engelman, A. (2004). Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J Biol Chem* **279**, 48883-92.

21. Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J. R. & Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* **11**, 1287-9.

22. Desfarges, S. & Ciuffi, A. (2010). Retroviral Integration Site Selection. *Viruses* **2**, 111-130.

23. Llano, M., Saenz, D. T., Meehan, A., Wongthida, P., Peretz, M., Walker, W. H., Teo, W. & Poeschla, E. M. (2006). An essential role for LEDGF/p75 in HIV integration. *Science* **314**, 461-4.

24. Maertens, G., Cherepanov, P., Debyser, Z., Engelborghs, Y. & Engelman, A. (2004). Identification and characterization of a functional nuclear localization signal in the HIV-1 integrase interactor LEDGF/p75. *J Biol Chem* **279**, 33421-9.

25. Ferris, A. L., Wu, X., Hughes, C. M., Stewart, C., Smith, S. J., Milne, T. A., Wang, G. G., Shun, M. C., Allis, C. D., Engelman, A. & Hughes, S. H. (2010). Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration. *Proc Natl Acad Sci U S A* **107**, 3135-40.

26. Gijsbers, R., Ronen, K., Vets, S., Malani, N., De Rijck, J., McNeely, M., Bushman, F. D. & Debyser, Z. (2010). LEDGF Hybrids Efficiently Retarget Lentiviral Integration Into Heterochromatin. *Mol Ther* **18**, 552-60.

27. Meehan, A. M., Saenz, D. T., Morrison, J. H., Garcia-Rivera, J. A., Peretz, M., Llano, M. & Poeschla, E. M. (2009). LEDGF/p75 proteins with alternative chromatin tethers are functional HIV-1 cofactors. *PLoS Pathog* **5**, e1000522.

28. Silvers, R. M., Smith, J. A., Schowalter, M., Litwin, S., Liang, Z., Geary, K. & Daniel, R. (2010). Modification of integration site preferences of an HIV-1-based vector by expression of a novel synthetic protein. *Hum Gene Ther* **21**, 337-49.

29. Shinohara, T., Singh, D. P. & Fatma, N. (2002). LEDGF, a survival factor, activates stress-related genes. *Prog Retin Eye Res* **21**, 341-58.

30. Takamura, Y., Fatma, N., Kubo, E. & Singh, D. P. (2006). Regulation of heavy subunit chain of gamma-glutamylcysteine synthetase by tumor necrosis factor-alpha in lens epithelial cells: role of LEDGF/p75. *Am J Physiol Cell Physiol* **290**, C554-66.

31. Sharma, P., Fatma, N., Kubo, E., Shinohara, T., Chylack, L. T., Jr. & Singh, D. P. (2003). Lens epithelium-derived growth factor relieves transforming growth factor-beta1-induced transcription repression of heat shock proteins in human lens epithelial cells. *J Biol Chem* **278**, 20037-46.

32. Daugaard, M., Kirkegaard-Sorensen, T., Ostenfeld, M. S., Aaboe, M., Hoyer-Hansen, M., Orntoft, T. F., Rohde, M. & Jaattela, M. (2007). Lens epithelium-derived growth factor is an Hsp70-2 regulated guardian of lysosomal stability in human cancer. *Cancer Res* **67**, 2559-67.

33. Anish, R., Hossain, M. B., Jacobson, R. H. & Takada, S. (2009). Characterization of transcription from TATA-less promoters: identification of a new core promoter element XCPE2 and analysis of factor requirements. *PLoS One* **4**, e5103.

34. Arinze, I. J. & Kawai, Y. (2003). Sp family of transcription factors is involved in valproic acid-induced expression of Galphai2. *J Biol Chem* **278**, 17785-91.

35. Zhu, G. H., Lenzi, M. & Schwartz, E. L. (2002). The Sp1 transcription factor contributes to the tumor necrosis factor-induced expression of the angiogenic factor thymidine phosphorylase in human colon carcinoma cells. *Oncogene* **21**, 8477-85.

36. Yang, C., Bolotin, E., Jiang, T., Sladek, F. M. & Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene* **389**, 52-65.

37. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engstrom, P. G., Frith, M. C., Forrest, A. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A. & Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature genetics* **38**, 626-35.

38. Kadonaga, J. T., Courey, A. J., Ladika, J. & Tjian, R. (1988). Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science* **242**, 1566-70.

39. Safe, S. & Abdelrahim, M. (2005). Sp transcription factor family and its role in cancer. *Eur J Cancer* **41**, 2438-48.

40. Ryu, S., Zhou, S., Ladurner, A. G. & Tjian, R. (1999). The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* **397**, 446-50.

41. Black, A. R., Black, J. D. & Azizkhan-Clifford, J. (2001). Sp1 and kruppel-like factor family of transcription factors in cell growth regulation and cancer. *J Cell Physiol* **188**, 143-60.

42. Shi, Q., Le, X., Abbruzzese, J. L., Peng, Z., Qian, C. N., Tang, H., Xiong, Q., Wang, B., Li, X. C. & Xie, K. (2001). Constitutive Sp1 activity is essential for differential constitutive expression of vascular endothelial growth factor in human pancreatic adenocarcinoma. *Cancer Res* **61**, 4143-54.

43. Wang, L., Wei, D., Huang, S., Peng, Z., Le, X., Wu, T. T., Yao, J., Ajani, J. & Xie, K. (2003). Transcription factor Sp1 expression is a significant predictor of survival in human gastric cancer. *Clin Cancer Res* **9**, 6371-80.

44. Abdelrahim, M., Smith, R., 3rd, Burghardt, R. & Safe, S. (2004). Role of Sp proteins in regulation of vascular endothelial growth factor expression and proliferation of pancreatic cancer cells. *Cancer Res* **64**, 6740-9.

45. Sapoznik, S., Cohen, B., Tzuman, Y., Meir, G., Ben-Dor, S., Harmelin, A. & Neeman, M. (2009). Gonadotropin-regulated lymphangiogenesis in ovarian cancer is mediated by LEDGF-induced expression of VEGF-C. *Cancer Res* **69**, 9306-14.

TABLE AND FIGURE LEGENDS

Table 1. List of primers used in this study.

Figure 1. Identification of *PSIP1* transcription start sites (TSSs) using 5'- RNA Ligase Mediated Rapid Amplification of cDNA Ends (5' RLM-RACE). a) Schematic representation of *PSIP1* genomic DNA (gDNA), with the two major RNA splice variants and encoded proteins. Exons were represented with boxes according to their length and their positions. **b)** Scheme of 5'RLM-RACE amplification on HEK cell total RNA (right panel). Reverse transcription and first PCR were performed with primers 1 (GeneRacer™ 5' primer) and 1' (gene specific primer); followed by nested PCR with primers 2 and 2'. PCR products were separated by gel electrophoresis (left panel), highlighting the two major variants, LEDGF/p75 and p52. **c)** Genomic DNA sequence of *PSIP1* (as deposited in Genbank), showing the TSS identified in our experimental system: LEDGF/p75 TSS (+1) and p52 TSS (+745), respectively at 36 bp and 64 bp downstream of the annotated NCBI TSS (triangle). Additional minor TSS positions are marked with arrows. The underlined sequence corresponds to the first intron of the *PSIP1* gene (+154/+640) and asterisks to the start codon of both LEDGF/p75 and p52. **d)** Specific RT-PCR of LEDGF/p75 and p52 transcripts. Reverse transcription of endogenous LEDGF/p75 (lanes 1 and 2) and p52 (lanes 3 and 4) transcripts was performed using specific 3' primers (right scheme) and used for PCR with p75-5'/Exon2a/2c-PSIP1 (lane 1), ATG-PSIP1/p75-3' (lane 2), p75-5'/Exon2a/2c-PSIP1 (lane 3), ATG-PSIP1/p52-3' (lane 4).

Figure 2. Identification of *PSIP1* promoter regulatory elements.

a) Schematic representation of the *PSIP1* promoter showing the progressive 5' end truncations used to identify *PSIP1* promoter regulatory elements . **b)** Isolation of the functional *PSIP1* promoter. Luciferase activity driven by 5' end PSIP1 promoter truncations at -2048, -1195, -723 and -207. **c)** Comparison of the -723/+59 and +140/+781 PSIP1 regions showed that LEDGF/p75 (~59 fold) and p52 (~ 5 fold) promoters displayed additive transcription activity. As control, minimal promoter activity containing consensual TATA box was measured in luciferase assay (blue rectangle). **d)** Left graph: Luciferase activity driven by finer progressive 5' end truncations of the LEDGF/p75 promoter. Right graph: luciferase activity driven by a minimal TATA box in presence or absence of the upstream -320/-207 enhancer region.. **e)** Identification of p52 promoter regulatory regions.

PSIP1 promoter activity was graphed as fold activity above background determined by the empty vector (pGL4.10), and was calculated as: RLU *PSIP1* construct / RLU background. Data presented as mean ± SEM. Differences between groups were assessed by student t- test. * = p value ≤ 0.1; ** = p value ≤ 0.01; *** = p value ≤ 0.001.

Figure 3. Crosstalk between LEDGF/p75 and p52 promoter activities.

Analysis of luciferase activity with 3' progressive deletions starting at -320 (top panel), -207 (middle panel) and -112 (bottom panel). *PSIP1* promoter activity was graphed as fold activity above background (determined by luciferase activity driven by the empty pGL4.10 vector), and was calculated as following: RLU of the *PSIP1* construction / RLU of the background. Data are presented as mean ± SEM.

Figure 4. *PSIP1* transcription is modulated by the Sp1 transcription factor.

a) Mapping of putative Sp1 binding sites on *PSIP1* promoter using Alibaba2.1 querying the TRANSFAC database **b)** Effect of the Sp1 binding site inhibitor mithramycin (top graphs)

and Sp1 overexpression (bottom graphs) on LEDGF/p75 (left graphs) and p52 (right graphs) promoters.

Firefly luciferase activity was normalized with the total protein concentration and then compared with the normaliyed background level (*firefly* luciferase activity of pGL4.10 / total protein concentration of pGL4.10-transfected cells). Data are presented as mean ± SEM. Differences between groups were assessed by ANOVA test. (mithramycin assay: n=6, p value ≤ 0.0001; Sp1 overexpression assay: n=5, p value ≤ 0.001). **c)** Endogenous levels of LEDGF/p75 and p52 were modulated by the transcription factor Sp1. Western blot analysis of cells transfected with increasing amounts of Sp1 expression vector, for Sp1 expression (top panel), LEDGF/p75 and p52 expression (middle panel) and tubulin expression as loading control (bottom panel). **d)** Sp1 binding to LEDGF/p75 promoter was investigated by chromatin immunoprecipitation followed by specific PCR. ChIP was performed using anti-Sp1 or normal rabbit IgG antibody (negative control). The immunoprecipitated DNA was amplified with primers specific for LEDGF/p75 minimal promoter (-112/+59). To check for ChIP specificity, 200 nM of the Sp1 binding inhibitor mithramycin was added. Amplification of the LEDGF/p75 minimal promoter (-112bp/+59) was performed by PCR. The sheared and crosslinked chromatin prior to ChIP was used as a positive control for PCR (input). (*) corresponded to the primers used in the PCR reaction.

Figure 5. Sp1 binding sites -72/-64 and -46/-36 are responsible for Sp1-mediated PSIP1 transcription regulation.

a) Mutagenesis of putative Sp1 sites of LEDGF/p75 promoter, individually or in combination. Putative Sp1 site 1 corresponds to the -101/-91 region of the promoter, putative Sp1 site 2 to -72/-64 and putative Sp1 site 3 to-46/-36 (see also Figure 4a). *PSIP1* promoter activity was graphed as fold activity above background determined by the empty vector (pGL4.10), and

was calculated as following: RLU of the *PSIP1* construct / RLU of the background. Data are presented as mean ± SEM. Differences between groups were assessed by students t- test. ** = p value ≤ 0.01. **b)** Effect of Sp1 overexpression on promoter constructs with mutated Sp1 binding sites. WT promoter -723/+59 containing the LEDGF/p75 TSS was stimulated with increasing amounts of Sp1, whereas Sp1 mutated promoters did not reveal any dose-dependent activity in presence of Sp1 overexpression. Western blots (bottom panels) were performed using anti-Sp1 and anti-tubulin to check Sp1 overexpression in all the samples tested in luciferase assay.

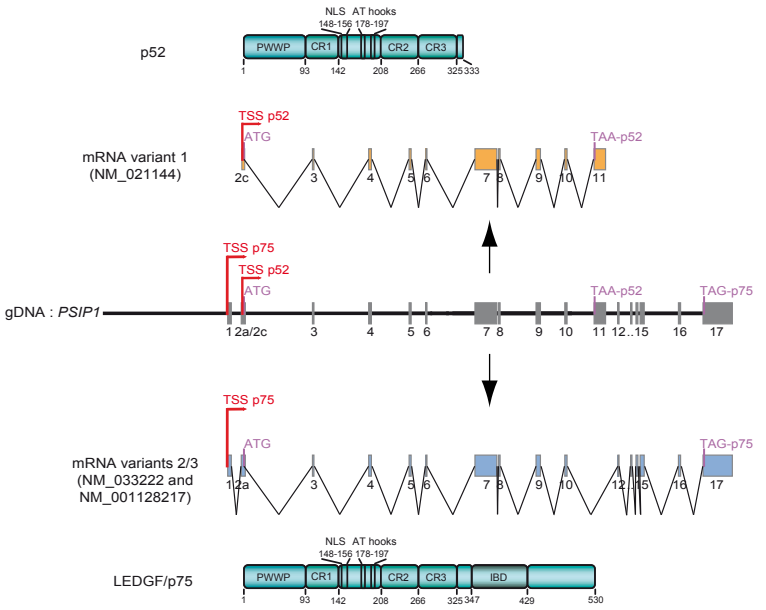
Figure 6. Model of *PSIP1* gene regulation. Grey boxes represent enhancers of the downstream TSS. The red box indicates the upstream silencer of LEDGF/p75 TSS. Green lines indicate mechanisms decreasing LEDGF/p75 transcription. Blue lines show mechanisms resulting in an increase of LEDGF/p75 promoter activity. Arrows indicate stimulation of the regulatory element whereas bars show inhibition of the regulatory element. The Sp1 transcription factor is represented with the black hexagon which can bind the functional binding sites -72/-64 and -46/-36 to induce transcription of the LEDGF/p75 promoter.

name	sequence	position	Used for
MA-pr 110	TTTTTTGGTACCGCAGCAGGAGCTAAGTATGG	-2043/-2023	5’ deletions of LEDGF/p75 promoter
MA.pr-111	TTTTTTGGTACCTTGCTCCACTTGGAACTCTC	-1195/-1175	5’ deletions of LEDGF/p75 promoter
MA.pr-115	TTTTTTGGTACCCCTGTCATTGTTTTCTACC	-723/-704	5’ deletions of LEDGF/p75 promoter
MA.pr-116	TTTTTTGGTACCATACTACAATTTCAAGGAAAAGG	-618/-595	5’ deletions of LEDGF/p75 promoter
MA.pr-117	TTTTTTGGTACCCCGCCGCATGCTCCAATTTCATC	-538/-515	5’ deletions of LEDGF/p75 promoter
MA.pr-118	TTTTTTGGTACCAATCTTTACTGCCACTTTCTCC	-420/-398	5’ deletions of LEDGF/p75 promoter
MA.pr-119	TTTTTTGGTACCGCCTTTTACATACAGTACAC	-320/-300	5’and 3’ deletions of LEDGF/p75 promoter and effect of the +59/+659 region
MA.pr-120	TTTTTTGGTACCTTTCGCCCAGTCCTTTCTTC	-207/-187	5’ deletions of LEDGF/p75 promoter
MA-pr 121	TTTTTTGGTACCAATCCGCTTCGGAGCCACAC	-112/-92	5’ deletions of LEDGF/p75 promoter
MA.pr-122	TTTTTTGGTACCCAGTGCTAGCGGGCGCCGAG	-37/-17	5’ deletions of LEDGF/p75 promoter
MA.pr-123	TTTTTTGGTACCTTCGCTTTAACCGCCCTCGGTG	+609/+631	5’ deletions of p52 promoter
MA.pr-124	TTTTTTCTCGAGGTTTCGGGGGCGAGACCGGG	+781/+761	5’ deletions of p52 promoter and 3’ deletion of the LEDGF/p75 promoter
MA.pr-127	TTTTTTCTCGAGAAAGGCAGGGATTCCGAGAAG	+339/+318	3’ deletions of LEDGF/p75 promoter
MA.pr-128	TTTTTTCTCGAGCCCGGGCGGGCCGCGTCCAC	+240/+220	3’ deletions of LEDGF/p75 promoter
MA.pr-129	TTTTTTCTCGAGGGGCGCCGACGCTGCGGTTG	+140/+120	3’ deletions of LEDGF/p75 promoter
MA.pr-130	TTTTTTCTCGAGCGTCTCAACGGCTCGGAATC	+59/+39	5’ deletions of LEDGF/p75 promoter
MA.pr-226	TTTTTTCTCGAGCGGGCGGGGGAGGATGCCTCGG	+659/+647	Effect of the +59/+659 region
MA.pr-265	TTTTTTGGTACCCGCGGCCCCGGCAGGTGAGC	+140/+160	5’ deletions of p52 promoter
MA.pr-266	TTTTTTGGTACCGGCCCGCCCGGGAGCCGAGG	+229/+249	5’ deletions of p52 promoter
MA.pr-267	TTTTTTGGTACCTCCCATGCACCCCCTCCCTTTTG	+345/+368	5’ deletions of p52 promoter
MA.pr-268	TTTTTTGGTACCTCAGGTGGTGGCGTCTCTTCGGTGG	+448/+473	5’ deletions of p52 promoter
MA.pr-269	TTTTTTGGTACCCTCCCACCCCCACCCGCCGGTTCC	+548/+572	5’ deletions of p52 promoter
MA.pr-270	TTTTTTCTCGAGGCCCGTCTGCCCGCCCCATCTTTC	+420/+396	3’ deletions of LEDGF/p75 promoter
MA.pr-271	TTTTTTCTCGAGCTCCCCCGCCAGTGCGCTGCCTCCG	+521/+496	3’ deletions of LEDGF/p75 promoter
MA.pr-280	ATCCGCTTCG**T**AG**TA**ACA**T**AGCTTCGCCGGGTGC	-111/-80	Sp1 binding site mutagenesis (site 1)
MA.pr-281	GGCGAAGCT**A**TGT**TA**CT**A**CGAAGCGGATTTTCTGG	-119/-86	Sp1 binding site mutagenesis (site 1)
MA.pr-282	CGGGTGCTGCAACAT**TA**C**AA**CTCTCTCGGTAAAC	-86/-53	Sp1 binding site mutagenesis (site 2)
MA.pr-283	ACCGAGAGAG**TT**G**TA**ATGTTGCAGCACCCGGCGAAGC	-92/-56	Sp1 binding site mutagenesis (site 2)
MA.pr-284	TAAACAGTTG**A**TC**A**C**TA**C**A**CAGTGCTAGCGGGC	-56/-32	Sp1 binding site mutagenesis (site 3)
MA.pr-285	GCACTG**T**G**TA**G**T**GA**T**CAACTGTTTACCGAGAGAGG	-66/-37	Sp1 binding site mutagenesis (site 3)
MA.pr-346	CAGTGCTAGCGGGCGCCGA**TA**G**T**GA**T**C**T**GCG	-37/-7	Sp1 binding site mutagenesis (site 4.1)
MA.pr-347	GCGC**A**G**A**TC**A**C**TA**TCGGCGCCCGCTAGCACTG	-6/-37	Sp1 binding site mutagenesis (site 4.1)
MA.pr-348	CGAGCGGGAGCC**T**C**T**C**A**G**A**AG**T**AGCGCAGC	-21/+10	Sp1 binding site mutagenesis (site 4.2)
MA.pr-349	GTAGCTGCGCT**A**CT**T**C**T**G**A**G**A**GGCTCCCGCTCG	+13/-21	Sp1 binding site mutagenesis (site 4.2)
MA.pr-350	CGA**TA**G**T**GA**T**C**T**GCGCGGGAGCAGCGCAGC	-21/-9	Sp1 binding site mutagenesis (site 4.3)
MA.pr-351	GTAGCTGCGCTGCTCCCGCGC**A**G**A**TC**A**C**TA**TCG	-13/-21	Sp1 binding site mutagenesis (site 4.3)
MA.pr-365	AGTGCTA**T**C**AT**G**T**G**A**CGA**TA**G**T**GA**T**C**TT**C**T**C**A**G**A**AG**T**AGCGCAGC	-36/+10	Sp1 binding site mutagenesis (full site 4)
MA.pr-366	GCTGCGCT**A**CT**T**C**T**G**A**G**AA**G**A**TC**A**C**TA**TCG**T**C**A**C**AT**G**A**TAGCACT	+10/-36	Sp1 binding site mutagenesis (full site 4)

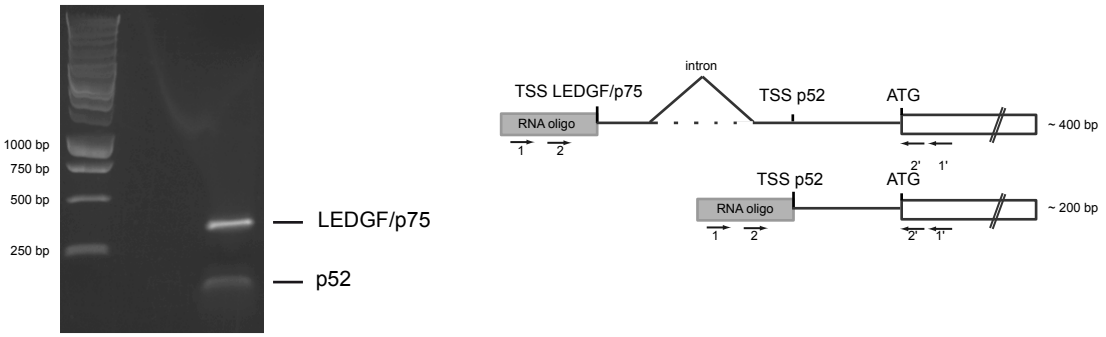
Bold: mismatch for mutagenesis of Sp1 binding sites
Underlined: restriction site of XhoI or KpnI

Figure 1

a)



b)



c)

LEDGF/p75

ACAGTTGGTCGCGCCCCAGTGCTAGCGGGCGCCGAGCGGGAGCCGCGCGGGAGCAGCGCAGCTACGGCGGCGGCAGCGGCGG

-36 +1

CGCGGTTGCGATTCCGAGCCGTTGAGACGCCTCTGCGGCAGCTGGTGGCGCAGGTGGCTTGCGTGGACGCGGGTAGAGGCGAC

CGGCCAGCAACCGCAGCGTCGGCGCCCGCGGCCCCGGCAGCGTGAGCGGGCGACGGGGACTCGCCCCGCCCCCCCCTCGGTGG

CGTCGCTCCTGGCTGTAGCGGGGAAGTGGACGCGGCCCGCCCGGGAGCCGAGGGGGCCGAGGCTGCAGCCGCTGGGCAGCAG

GTTGCGGGGGAGGCCTTGCGTGGGCGGAAGGGGGCCGCTCGCTTCTCGGAATCCCTGCCTTTTTCTCTCCCATGCACCCCCTC

CCTTTTGGGGCTTAAGCTCGGTGGAAAAAGCCCGGAGAAAGATGGGGCGGGCAGACGGGCGATTTCTGGCCGGGCAGCATCC

CTGCTTCAGGTGGTGGCGTCTCTTCGGTGGCGGGGTGGAGGACCTGCTCCAGCGCGGAGGCAGCGCACTGGCGGGGGAGGAG

CGGGAGGCGGCCGTCCCCTCGCTCTCCCACCCCCACCCGCCGGTTCCCCTCCCCCTCCCCTTGCGGCCCCTCCCTTCGGGGTT

CTTCGCTTTAACCGCCCTCGGTGCTCTCCCTAGCAGGCGCGTCGGGACGCCCCGAGGCATCCTCCCCCGCCCGCGGGCCCGGT

+681

p52

AGCTGGGCCCGCGTCCGCCGCCCGCATCCCCGCGCCGCCGCATCTCCTCGCCGCCTCCCGGGCTTCGGACCCCCGGTCTCGCC

+745 ***

CCCGAAACATG

d)

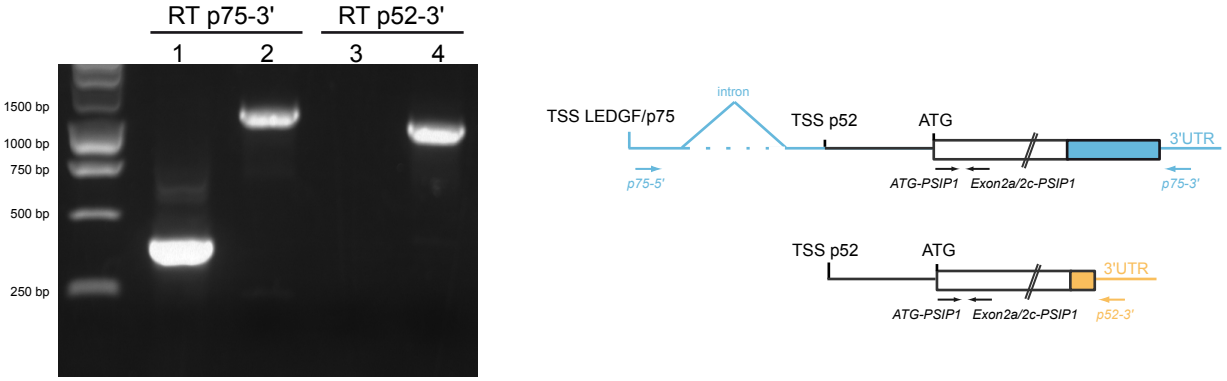


Figure 2

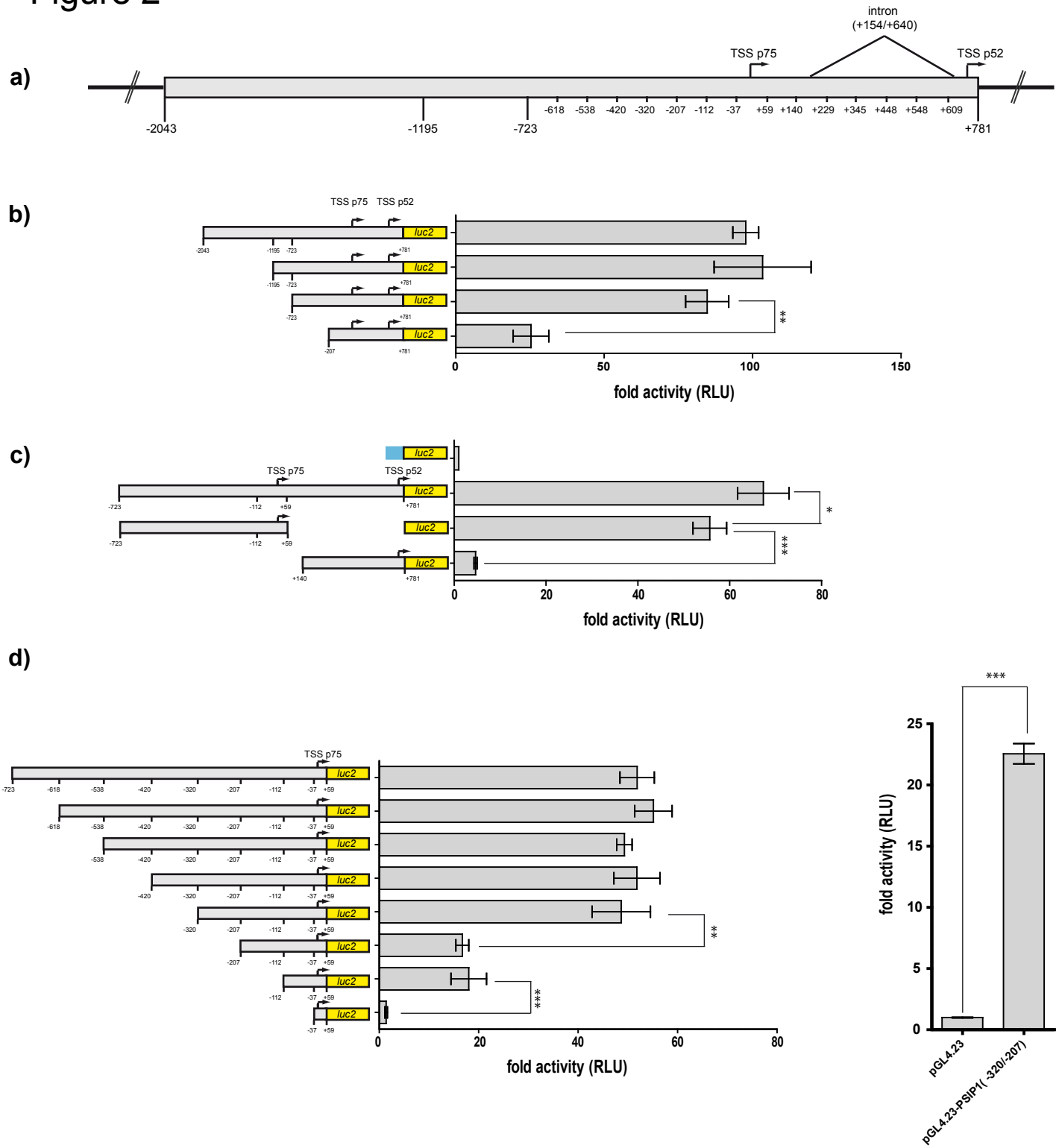

a)
intron
(+154/+640)
TSS p75
TSS p52
-2043
-1195
-723
-618 -538 -420 -320 -207 -112 -37 +59 +140 +229 +345 +448 +548 +609
+781
b)
TSS p75 TSS p52
luc2
fold activity (RLU)
c)
TSS p75
TSS p52
luc2
fold activity (RLU)
d)
TSS p75
luc2
fold activity (RLU)
pGL4.23
pGL4.23-PSIP1(-320/-207)
e)
TSS p52
luc2
fold activity (RLU)

Figure 3

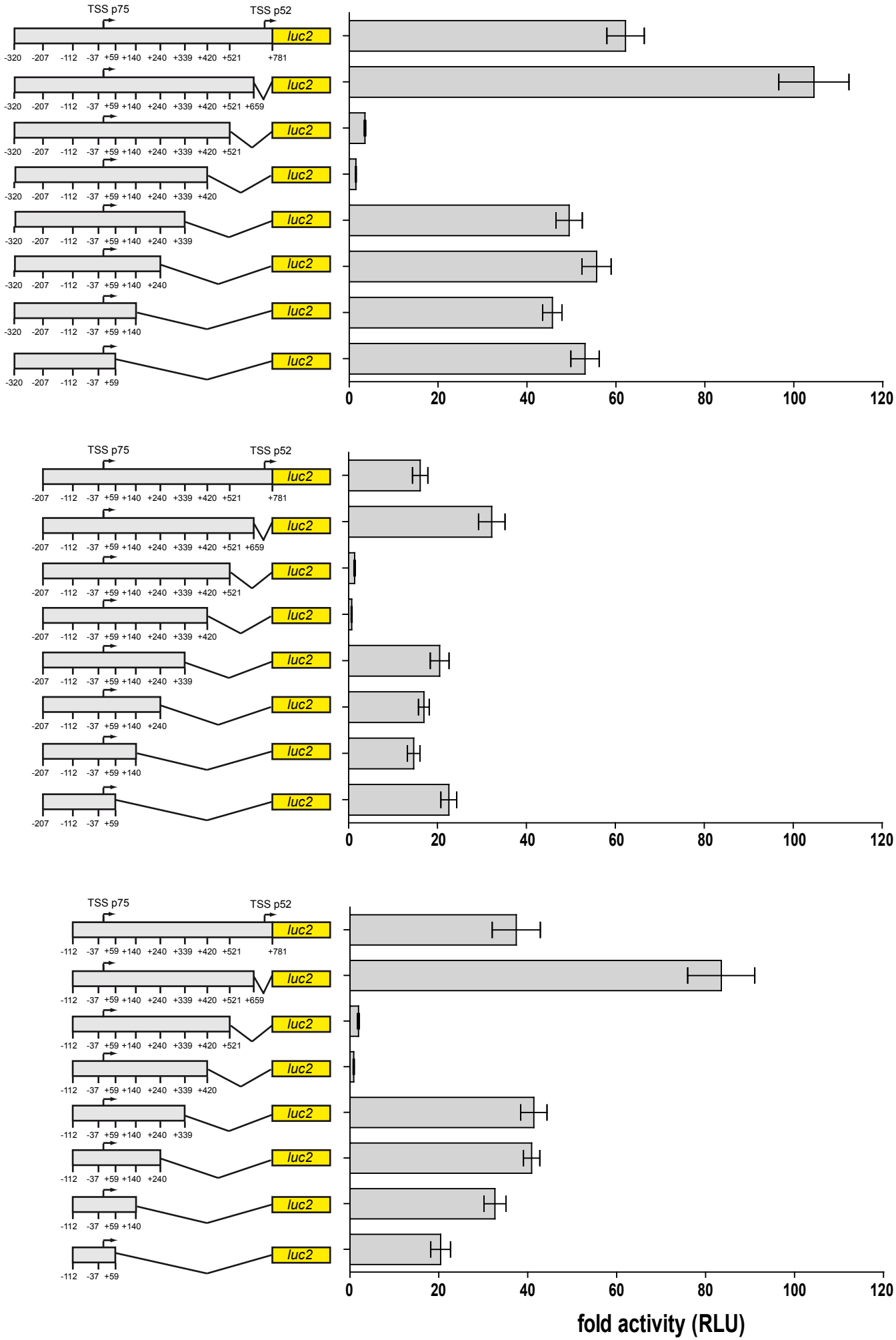
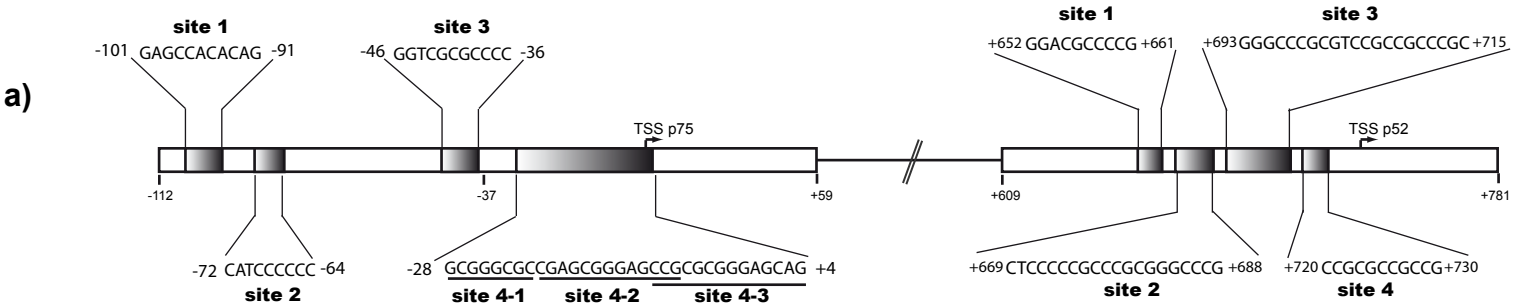
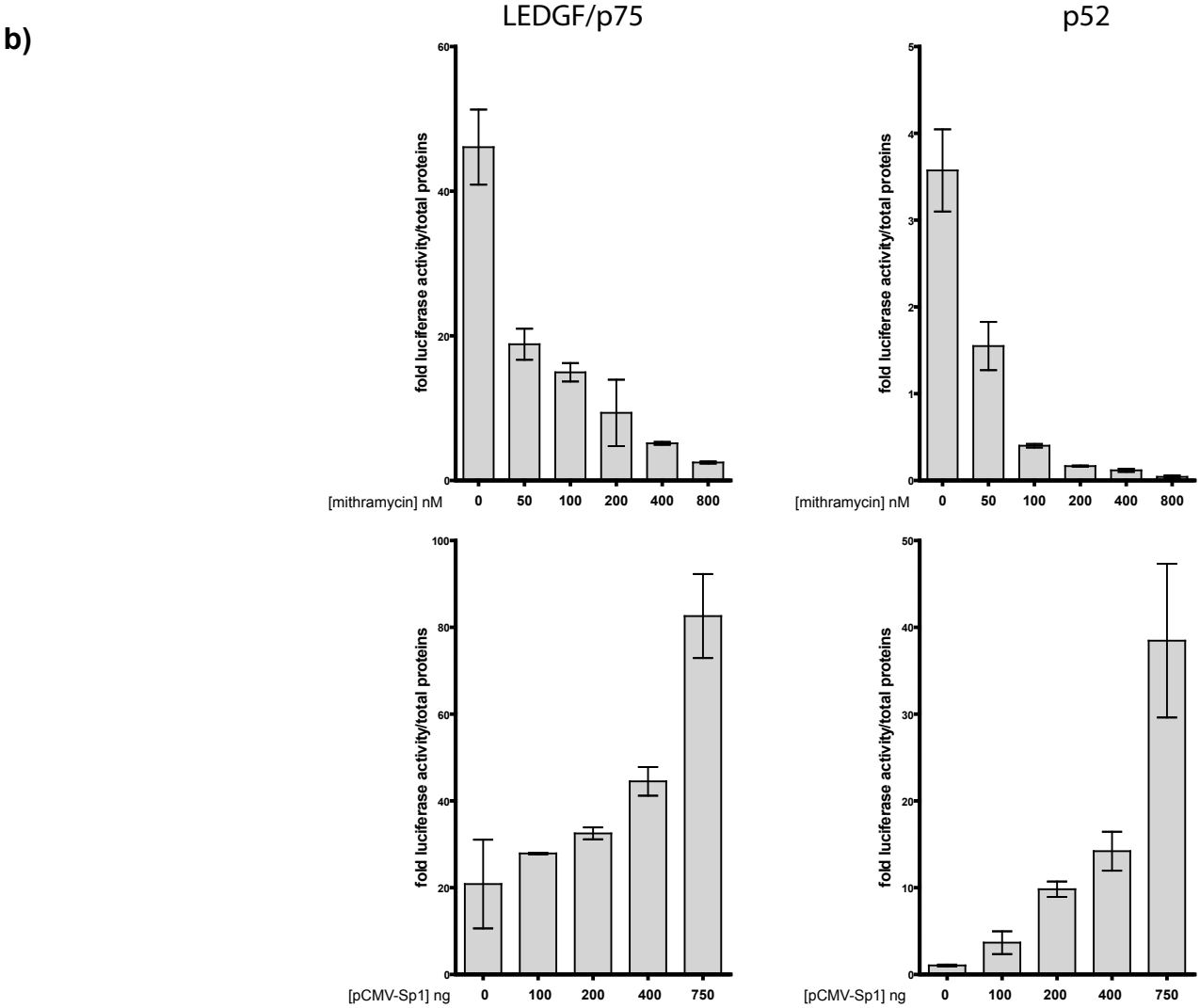
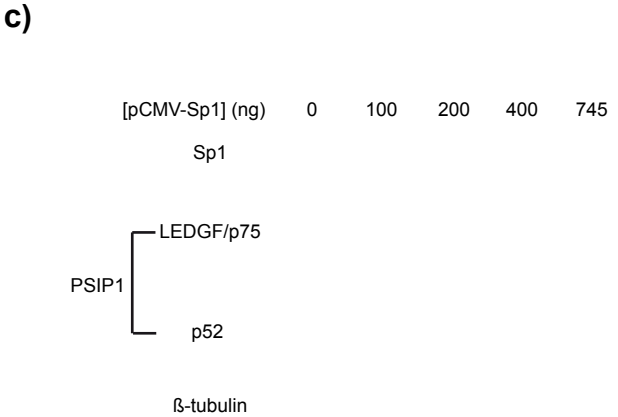

TSS p75
TSS p52
luc2
-320 -207 -112 -37 +59 +140 +240 +339 +420 +521 +781
-320 -207 -112 -37 +59 +140 +240 +339 +420 +521 +659
-320 -207 -112 -37 +59 +140 +240 +339 +420 +521
-320 -207 -112 -37 +59 +140 +240 +339 +420
-320 -207 -112 -37 +59 +140 +240 +339
-320 -207 -112 -37 +59 +140 +240
-320 -207 -112 -37 +59 +140
-320 -207 -112 -37 +59
-207 -112 -37 +59 +140 +240 +339 +420 +521 +781
-207 -112 -37 +59 +140 +240 +339 +420 +521 +659
-207 -112 -37 +59 +140 +240 +339 +420 +521
-207 -112 -37 +59 +140 +240 +339 +420
-207 -112 -37 +59 +140 +240 +339
-207 -112 -37 +59 +140 +240
-207 -112 -37 +59 +140
-207 -112 -37 +59
-112 -37 +59 +140 +240 +339 +420 +521 +781
-112 -37 +59 +140 +240 +339 +420 +521 +659
-112 -37 +59 +140 +240 +339 +420 +521
-112 -37 +59 +140 +240 +339 +420
-112 -37 +59 +140 +240 +339
-112 -37 +59 +140 +240
-112 -37 +59 +140
-112 -37 +59
0 20 40 60 80 100 120
fold activity (RLU)

Figure 4


a)
site 1
-101 GAGCCACACAG -91
site 3
-46 GGTCGCGCCCC -36
TSS p75
-112
-37
+59
-72 CATCCCCCC -64
site 2
-28 GCGGGCGCCGAGCGGGAGCCGCGCGGGAGCAG +4
site 4-1
site 4-2
site 4-3
site 1
+652 GGACGCCCCG +661
site 3
+693 GGGCCCGCGTCCGCCGCCCGC +715
TSS p52
+609
+781
+669 CTCCCCCGCCCGCGGGCCCG +688
site 2
+720 CCGCGCCGCCG +730
site 4


b)
LEDGF/p75
p52
fold luciferase activity/total proteins
[mithramycin] nM 0 50 100 200 400 800
[pCMV-Sp1] ng 0 100 200 400 750


c)
[pCMV-Sp1] (ng) 0 100 200 400 745
Sp1
PSIP1
LEDGF/p75
p52
ß-tubulin

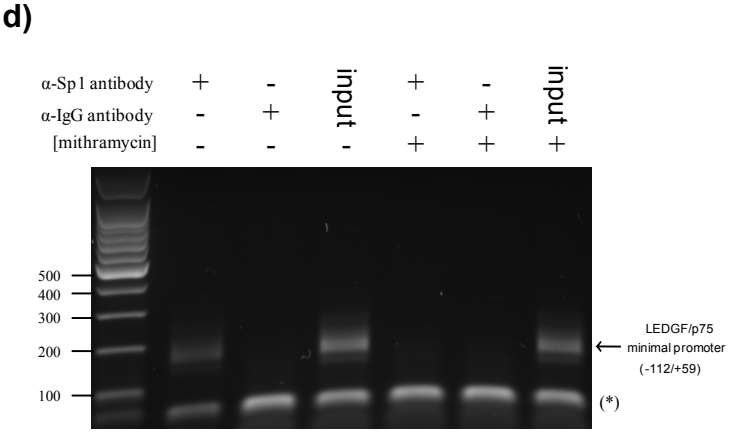
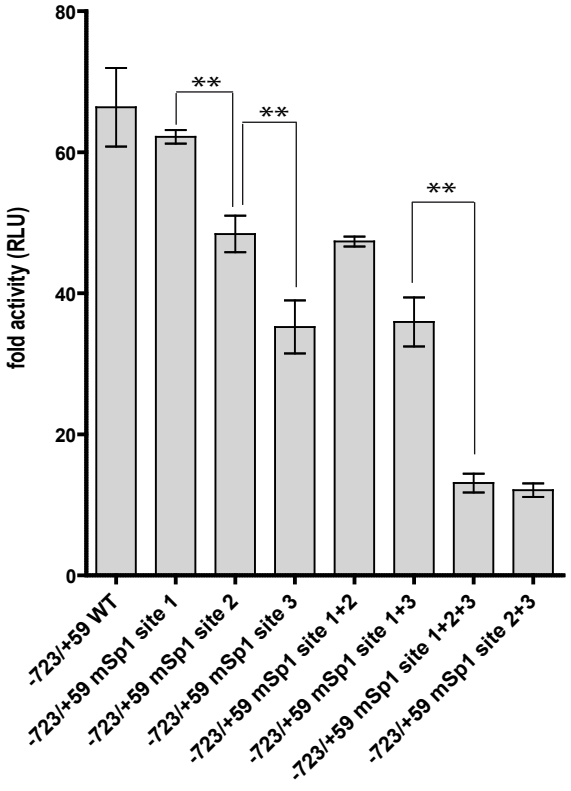

d)
α-Sp1 antibody + - input + - input
α-IgG antibody - + - + -
[mithramycin] - - - + + +
500
400
300
200
100
LEDGF/p75
minimal promoter
(-112/+59)
(*)

Figure 5

a)


fold activity (RLU)
80
60
40
20
0
**
**
**
-723/+59 WT
-723/+59 mSp1 site 1
-723/+59 mSp1 site 2
-723/+59 mSp1 site 3
-723/+59 mSp1 site 1+2
-723/+59 mSp1 site 1+3
-723/+59 mSp1 site 1+2+3
-723/+59 mSp1 site 2+3

b)

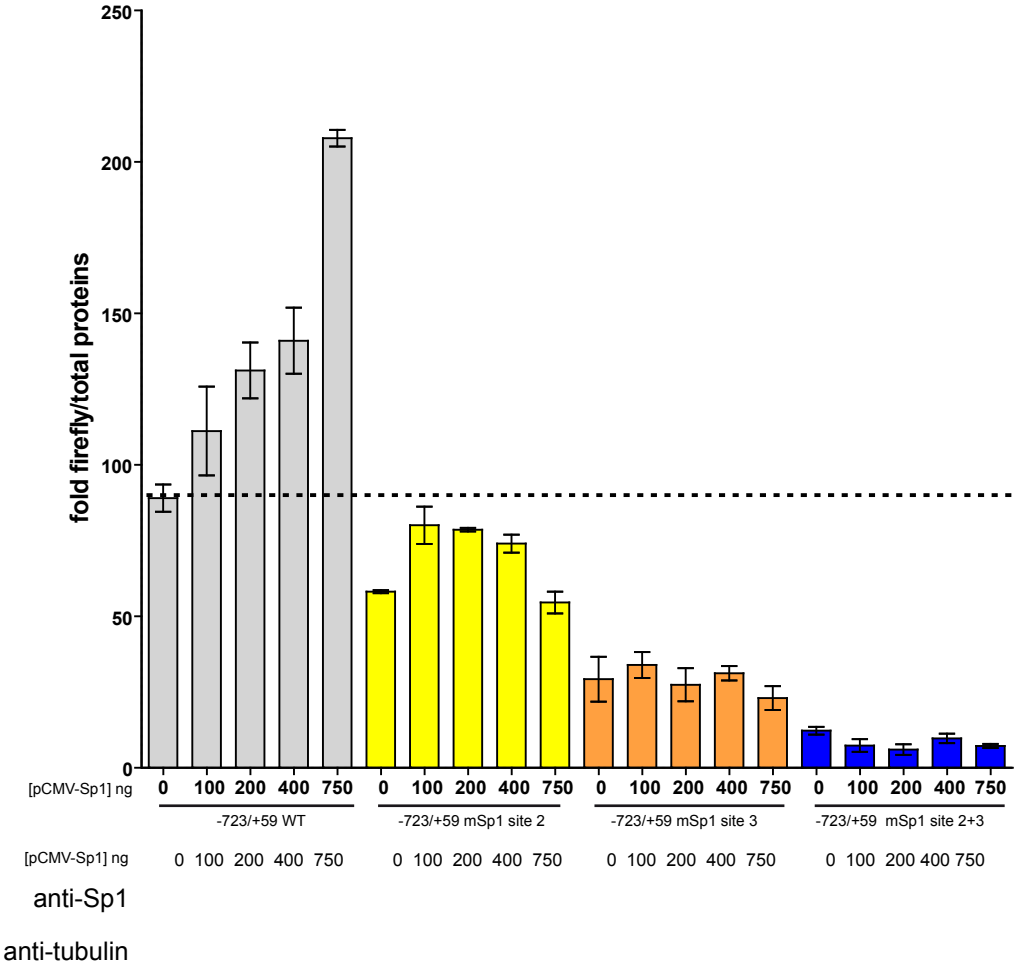
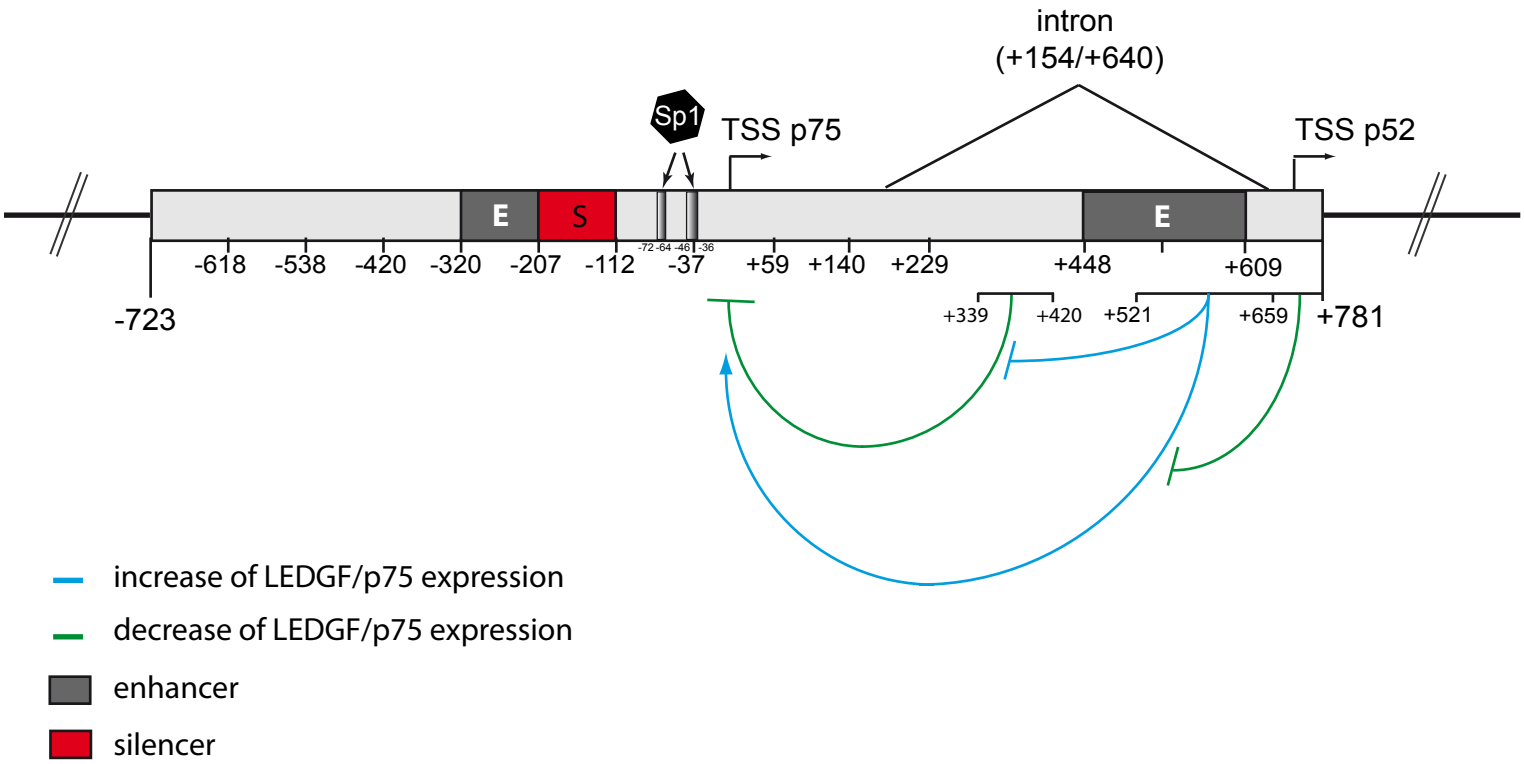

fold firefly/total proteins
250
200
150
100
50
0
[pCMV-Sp1] ng
0 100 200 400 750 0 100 200 400 750 0 100 200 400 750 0 100 200 400 750
-723/+59 WT
-723/+59 mSp1 site 2
-723/+59 mSp1 site 3
-723/+59 mSp1 site 2+3
[pCMV-Sp1] ng
0 100 200 400 750 0 100 200 400 750 0 100 200 400 750 0 100 200 400 750
anti-Sp1
anti-tubulin

Figure 6


intron
(+154/+640)
Sp1
TSS p75
TSS p52
E
S
E
-618
-538
-420
-320
-207
-112
-37
+59
+140
+229
+448
+609
-723
+339
+420
+521
+659
+781
increase of LEDGF/p75 expression
decrease of LEDGF/p75 expression
enhancer
silencer

S1)

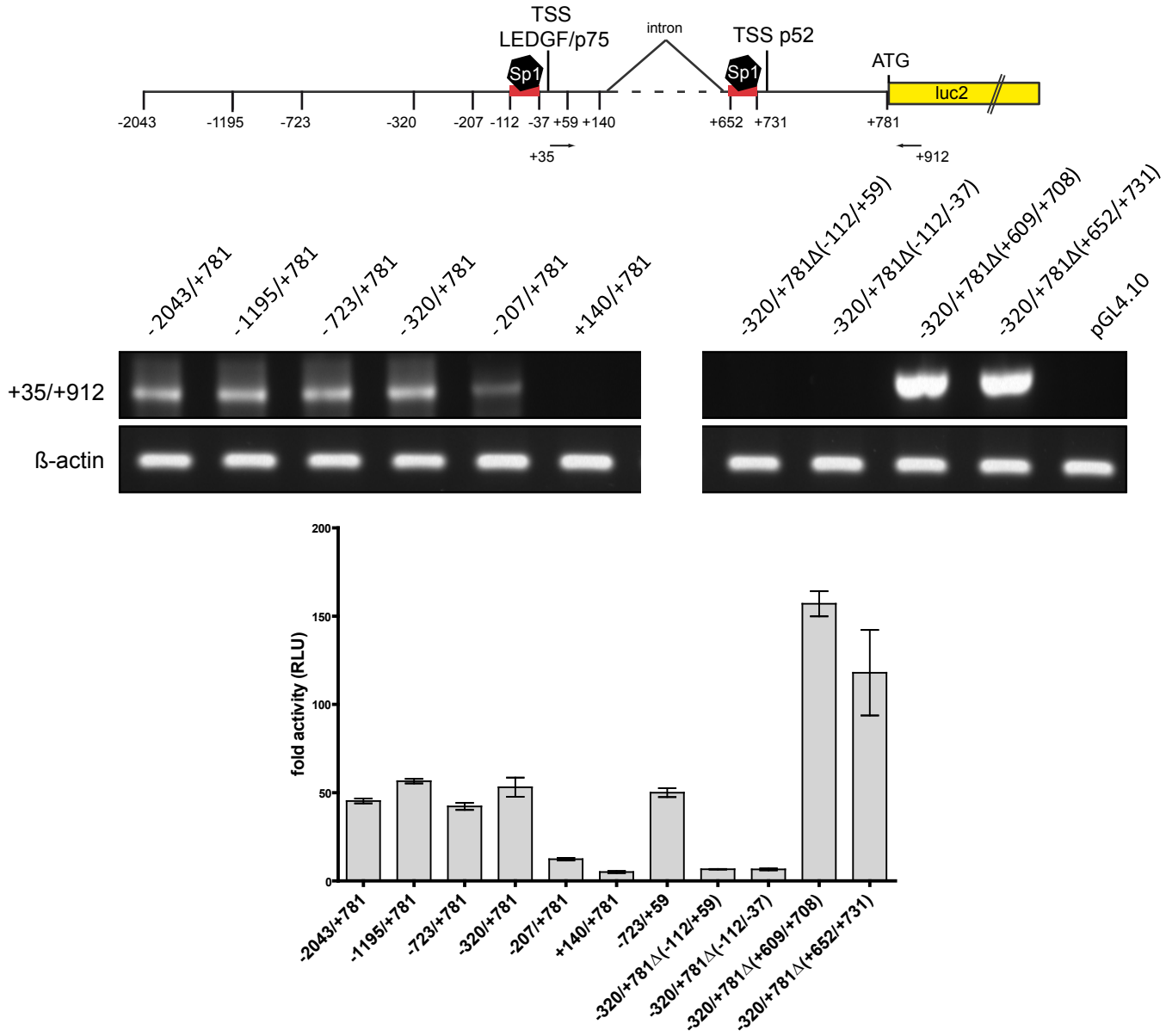


Figure S1. Identification of luciferase transcripts starting at TSS p75 (+1). PolyA+ RNA was extracted from HEK cells 40 h post transfection with luciferase constructs driven by diverse *PSIP1* promoter regions. Reverse transcription was performed using 150 ng of polyA+ RNA, hexamers and MultiScribe RT (Applied Biosystem). PCR was performed using primers that amplified the +35/+912 product, and ß-actin as control. Luciferase assay was performed at 40h post transfection as described in the Materials and Methods section.

S2)

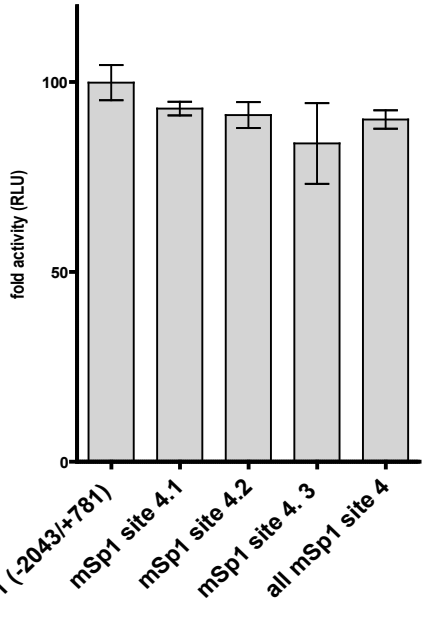


Figure S2. Mutagenesis of the putative Sp1 binding site 4. Computational analysis of the promoter showed three putative overlapping Sp1 binding sites (named 4.1, 4-2 and 4-3) in the region -28/+4 encompassing the LEDGF/p75 TSS. Individual or combined mutations in Sp1 site 4 did not affect LEDGF/p75 promoter activity. PSIP1 promoter activity was graphed as fold activity above background determined by the empty vector (pGL4.10), and was calculated as following: RLU of the PSIP1 construction / RLU of the background

LEDGF/p75 TATA-less promoter is driven by the transcription factor Sp1

Highlights

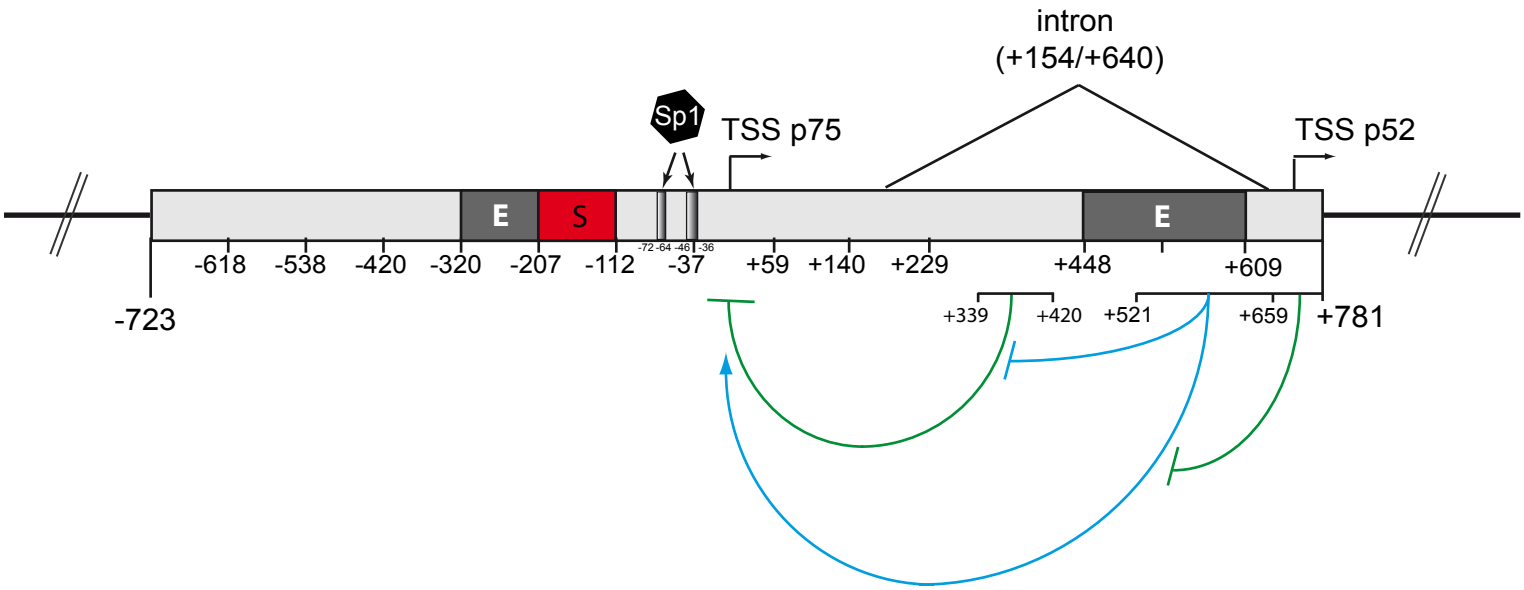
> LEDGF/p75 is involved in many cellular processes and diseases

> LEDGF/p75 gene regulation was analyzed by luciferase assay

> LEDGF/p75 is driven by a TATA-less promoter

> Promoter regions necessary for LEDGF/p75 regulation were also identified

> Sp1 is the key regulator for LEDGF/p75 expression


intron
(+154/+640)
Sp1
TSS p75
TSS p52
E
S
E
-618
-538
-420
-320
-207
-112
-37
+59
+140
+229
+448
+609
-723
+339
+420
+521
+659
+781