UNIVERSITA' DEGLI STUDI DI MILANO
FACOLTA' DI MEDICINA E CHIRURGIA

DIPARTIMENTO DI BIOLOGIA E GENETICA PER LE SCIENZE MEDICHE AA 2007-09

UCSD SCHOOL OF MEDICINE AA 2009-10

CORSO DI DOTTORATO DI RICERCA IN:

BIOTECNOLOGIE APPLICATE ALLE SCIENZE MEDICHE – CICLO XXIII-

TESI DI DOTTORATO DI RICERCA:

Alternative splicing and phosphorylation provide a mechanistic basis for the functional specialization of LSD1/KDM1 histone demethylase in the central nervous system

BIO/13 – Biologia Applicata

Dottorando:

D.ssa ZIBETTI Cristina

Matricola: R07806

Relatore:

Prof. GINELLI Enrico

Correlatore:

D.ssa. BATTAGLIOLI Elena

Supervisore Anno Accademico 2009-2010:

Prof. ROSENFELD M.G.

Anno Accademico 2009-2010

Seize the moment of excited curiosity on any subject to solve your doubts, for if you let it pass, the desire may never return, and you may remain in ignorance

-William Wirt (1772-1834)-

I have never let my schooling interfere with my education

-Mark Twain (1835-1910)-

The important thing is not to stop questioning

-Albert Einstein (1879-1955)-

Imagination is more important than knowledge

-Albert Einstein (1879-1955)-

Genius is one percent inspiration, ninety-nine percent perspiration

-Thomas A. Edison(1847-1931)-

I have not failed. I've just found 10.000 ways that won't work

-Thomas A. Edison(1847-1931)-

INTRODUCTION

Chromatin represents the physiological substrate of epigenetic regulation, underlying several biological processes, from replication and transcriptional activity to cell lineage commitment and adaptation in response to specific cues (Kouzarides et al. 2007; Wu et al., 2009). In the central nervous system, chromatin integrates a plethora of converging signaling pathways, leading to short- and long-term changes in gene expression (Flavell et al. 2008, Kim et al. 2010) that are crucial for neuronal commitment, terminal differentiation (Ballas et al. 2005) and neuroplasticity throughout life (Walton et al., 1999; Tartaglia et al., 2001; Borrelli et al. 2008, Dulac et al. 2010).
Among the first characterized ubiquitous factors regulating the neuronal phenotype is NRSF (*Neuro Restricted Silencing Factor*), also known as REST (*Repressor Element 1 Silencing Transcription Factor*), which regulates neuronal genes within the CNS, preventing their ectopic expression in extra nervous tissues (Lunyak et al.2002). REST deletion or its functional inactivation in extra nervous system relieves genes from repression resulting in embryonic lethality whereas its overexpression in the CNS determines functional impairment in ionic conductance (Nadeau et al. 2002). REST expression is dynamically regulated during the transition from embryonic stem cells to neuronal progenitors when its proteasome degradation predisposes target genes to the subsequent induction (Ballas et al.2005). During neuronal terminal differentiation retinoic acid signaling leads to REST downregulation and target genes can be induced at different extent: while the expression of class I genes relies exclusively on REST, class II genes like calbindin, BDNF, synaptotagmin IV are variably induced, depending on the coordinated action of REST, CoREST and MeCP2 that can be dismissed upon depolarization (Zhou et al. 2006). REST function in the CNS is also regulated by alternative splicing (Palm et al. 1999) resulting in a dominant negative truncated isoform which exhibits lower affinity for DNA consensus RE1 element but still retains interaction with molecular partners, mSin3a,b and CoREST competing with the full length REST for transcriptional repressors such as HDAC1,2 and ATP-ase components of Swi-SNF complexes.
Although a variety of ubiquitously expressed chromatin-remodeling complexes assist tissue-specific transcription factors in mediating histotype-restricted transcriptional regulation (Visel et al. 2009, Xi et al. 2007), neuro restricted chromatin-remodeling factors have just been recently described (Olave et al., 2002; Barak et al., 2004; Lessard et al., 2007).

Here, I report the identification and functional characterization of four mammal-specific LSD1 variants arising from single or double inclusion of two alternatively spliced exons, resulting in either ubiquitous isoforms or neuro-restricted ones, whose pattern of expression is dynamically regulated during perinatal brain development and early synaptic establishment, contributes to the acquisition of neurite morphology and whose functional diversification partly relies on exon specific phosphorylation. Notably, alternative splicing implements evolutionary complexity without a corresponding increase in gene number (Xing et al. 2006), occurring most frequently on singleton genes for which and additional member may result in the acquisition of a non-redundant function, whereas the frequency decreases as new members are generated (Kopelman et al. 2005, Talavera et al. 2007). Noteworthy is the fact that LSD1 displays a unique paralogue in the whole human genome, *AOF-1*, also known as LSD2 (Karytinos et al.2009, van Essen et al. 2010, Fang et al.2010, Binda et al. 2010) also endowed of demethylase activity, and that alternative splicing occurs massively in the CNS (Jin et al.2008, Lin et al.2010), expanding molecular variability of the resultant proteins that participate to fundamental nervous processes, as axon guidance and synapse formation (Lee et al. 2010, Sala et al. 2003).
Initially identified as common component of multiple co-repressor complexes (Hakimi et al., 2002; Humphrey et al., 2001; Shi et al., 2003; Tong et al., 1998; You et al., 2001) the Lysine Specific Demethylase 1 [LSD1/AOF2/BHC110/KDM1A/SU(VAR)3-3], the first demethylase to be characterized (Shi et al.2004), was found to specifically demethylate mono- and di-methylated H3K4 (H3K4me and H3K4me2, respectively), through a flavin-dependent oxidative reaction similar to that catalyzed by polyamine oxidases and monoamine oxidases (Shi et al., 2004; Forneris et al., 2005a). LSD1 catalyzes the disruption of the α-carbon bond in the histone substrate to generate an unstable imine intermediate which undergoes spontaneous hydrolysis resulting in aldehyde and amine formation. During the process, the FAD^+ prosthetic group is reduced to $FADH_2$ and its oxidative state is rescued by molecular oxygen which is reduced to hydrogen peroxide, acting as electron acceptor. Terminal products are the demethylated lysine K4 residue of histone H3 and formaldehyde and since one methyl group per time can be removed, different K4H3 methylation states can be generated. Given that the oxidation of the methylated aminic groups requires the nucleophilic attach of the substrate by means of protonated nitrogen, LSD1 activity is restricted to the mono and di-methylated forms. Despite the absence of LSD1 activity on the tri-methylated histone substrates, other histone demethylases have been identified,

catalyzing protein demethylation through a hydroxylation reaction that requires iron and 2-oxoglutarate as cofactors (Binda et al.2003). Unlike LSD1, JmJc-containing enzymes can act on all three methylated states of a lysine side chain and some Jumonji enzymes can act on methylated arginines (Chang et al. 2007).

LSD1 structure and function is conserved from yeast to human (Dallman et al., 2004; Lakowski et al., 2006), and it is typically associated to CoREST, a corepressor protein, and histone deacetylases HDAC1 and HDAC2 (Ballas et al., 2001; Humphrey et al., 2001; You et al., 2001; Hakimi et al., 2002; Shi et al., 2003). Human LSD1 consists of 852 amino acids and comprises an N-terminal SWIRM domain, involved in protein interactions and a C-terminal amine oxidase domain which contains an insertion that forms the CoREST interacting site, an extended helical region termed the "Tower" domain (Chen et al., 2006; Stavropoulos et al., 2006). The three-dimensional structure of LSD1 in a ternary complex with its histone peptide substrate and CoREST has been solved, providing an explanation for the biological relevance of CoREST in mediating LSD1 substrate binding and recognition (Yang et al., 2006; Forneris et al., 2007): structural studies have shown that the C-terminal SANT domain within CoREST facilitates the association with chromatin by interacting directly with DNA (Yang et al. 2006). Other LSD1 complex members include the corepressor CtBP (Shi et al. 2003), HMG domain containing protein, BRAF35 (Hakimi et al. 2003, Lee et al. 2005), and BHC80, which contains a PHD finger that specifically recognizes unmodified H3K4 (Lan et al. 2007). In addition to these canonical functions, LSD1 was recently shown to be recruited to the NuRD complex, via interaction of the Tower domain with MTA1-3 in breast cancer cells (Wang et al.2009). Interaction with CoREST prevents LSD1 proteasome degradation and is required for the recognition and demethylation of nucleosomal substrates (Lee et al. 2005). The presence of HDAC1/2 suggests a coordinate modification of histone tails, which is supported by evidence that hypo acetylated histone H3 tails are the preferred substrate for LSD1. Mechanisms that tune LSD1 activity have been extensively investigated, including interactions with protein partners (Lee et al., 2005; Shi et al., 2005) and concomitant modifications on histone tail residues surrounding the demethylation site (Forneris et al., 2005b; Forneris et al., 2006, Metzger et al. 2010). Consistent with H3K4me2, an active marker of transcription, as a substrate, LSD1 is found in cells as part of a core complex with the corepressor, CoREST, and histone deacetylase enzymes 1 and 2 (HDAC1 and -2) which repress transcription by deacetylating histone tails. Nonetheless, the association of LSD1 with the androgen

receptor switches its substrate specificity from H3K4me/me2 to H3K9me/me2 (Metzger et al.2005, Wissmann et al. 2007) resulting in gene activation (Garcia-Bassetts et al.2007). Likewise, repression relief of myogenic genes can be observed upon LSD1 recruitment at regulated promoters in association with Mef2 and MyoD, further supporting its activatory role (Choi et al. 2010). Furthermore, long noncoding RNAs restricted to specific histotypes proved to act as scaffold molecules mediating LSD1 recruitment at *HoxD* proximal promoters (Tsai et al. 2010). These overall evidences indicate that the composition of LSD1-containing complexes has the potential to alter its recruitment at target genes and its substrate specificity.

LSD1 mediated histone demethylation has been related so far to a wide range of biological processes, such as embryonic epiblast development (Foster et al. 2010), myoblast differentiation (Choi et al.2010), adipogenesis (Musri et al.2010), regulation of NfKB signaling (van Essen et al. 2010) and cancer (Kahl et al. 2006, Shi, 2007; Forneris et al., 2008; Nottke et al., 2009, Shulte et al.2009). The LSD1 heterodimeric partner CoREST, is a corepressor for the RE1 silencing transcription factor (REST), which represses neuronal genes in nonneuronal cells (Ballas et al.2001). Inhibition of LSD1 function causes increased expression of CoREST targets such as the acetylcholine receptor (AchR), synapsin, and sodium channels (SCN1A, SCN2A, and SCN3A) in nonneuronal cells (Shi et al.2005). The regulation of hematopoietic differentiation via the growth factor-independent (Gfi) transcription factors partially relies on their interaction with the LSD1/ CoREST/HDAC complex through an N-terminal SNAG domain (Saleque et al. 2007). Pituitary development and the appropriate expression of pituitary specific hormones are also dependent on LSD1 in the mouse (Wang et al.2007). Moreover, LSD1 participation to the Nurr1/CoREST pathway in microglia and astrocytes has been shown to protect dopaminergic neurons from inflammation-induced death (Saijo et al.2009).

Each of these roles involves a direct recruitment to target genes and the manipulation of histone substrates. In contrast, LSD1 function has been implicated in the DNA damage response by demethylating p53, restricting the interaction of p53 with its cofactor p53BP1 (Huang et al.2007). The maintenance DNA methyltransferase Dnmt1 is another nonhistone substrate (Wang et al. 2009a) and because methylation of Dnmt1 by Set7/9 increases protein turnover, the loss of LSD1 demethylase activity results in reduced levels of Dnmt1 and global DNA methylation. Genetic ablation of LSD1 also causes early embryonic lethality at approximately embryonic day 6.5 (Wang et al. 2009b), whereas conditional

ablation of LSD1 in ES cells determines a reduction of CoREST levels and associated HDAC activity with consequent induction of genes related to anterior-posterior patterning and limb formation (Foster et al. 2010), indicating LSD1 regulatory role on the transcriptome during embryonic development.

MATERIALS AND METHODS

Mammalian and bacterial expression vectors

pGal4-LSD1: the full length LSD1 cDNA (NM015013) (aa-35/852) has been obtained from pET-LSD1 digested with NdeI (blunted)/XhoI and cloned in pGal4-vector digested with EcoRI (blunted)/SalI.
pGal4-LSD1-2a/8a: the full length LSD1 cDNA (BC040194) (aa-35/876) containing both the alternative spliced exons has been obtained from the IMAGE clone 4815528 (pBSIIKS vector) digested with ApaI/SpeI. The fragment was substituted in pGal4-hLSD1 digested with ApaI/SpeI.
pGal4-LSD1-2a: a fragment containing Ex2a was excided from pGal4-hLSD1-Ex2a/Ex8a with PvuII and substituted in pGal4-hLSD1-wt digested with PvuII
pGal4-LSD1-8a: a fragment containing E8a was obtained from pGal4-LSD1-2a/8a digested with PvuII and replaced in pGal4-hLSD1 digested with PvuII.
pEGFP-C1-LSD1: a PCR fragment containing LSD1 cDNA (aa1-852) was cloned in EcoRI/KpnI of pEGFP-C1 vector.
pCGN-HA-LSD1: the full LSD1 sequence (aa1-852) was obtained from pEGFP-C1-LSD1 digested with EcoRI/KpnI and cloned after bluntization in pCGN-HA vector digested with BamHI and blunted.
pCGN-HA-LSD1-2a/8a: a fragment containing E2a/E8a was obtained from pGal4-hLSD1-2a/8a digested with FseI/SpeI and substituted in pCGN-HA-LSD1 digested with FseI/SpeI.
pCGN-HA-LSD1-2a: a fragment containing E2a was obtained from pGal4-LSD1-2a digested with FseI/SpeI and cloned in pCGN-HA-LSD1 digested with FseI/SpeI.
pCGN-HA-LSD1-8a: a fragment containing E8a was obtained from pGal4-LSD1-8a digested with FseI/SpeI and replaced in pCGN-HA-LSD1 digested with FseI/SpeI.
pSuperGFPNeo knocker: the siRNAs overhanging precursors were annealed into BglII-XhoI cleaved pSuperGFPNeo, after the H1 RNA PolIII specific promoter to generate scramble, nsLSD1 or uLSD1 hairpins reservoirs.
pCGN-HA-LSD1 wt rat : the LSD1 rat sequence (aa1-852) was obtained by mutagenesis of the pCGN-HA-LSD1 wt human template, limitedly to exons flanking the neurospecific one.
pCGN-HA-T371D: a single T≥D aminoacidic substitution was introduced onto pCGN-HA-LSD1-8a (corresponding to residue Thr 371 in the LSD1 ref. sequence) onto the pCGN-HA-LSD1-8a template

pCGN-HA-T371A: a single T≥A aminoacidic substitution was introduced onto pCGN-HA-LSD1-8a (corresponding to residue Thr 371 in the LSD1 ref. sequence) onto the pCGN-HA-LSD1-8a template pLVTHM knocker: : the siRNAs overhanging precursors were annealed into MluI-ClaI cleaved vector, after the H1 RNA PolIII specific promoter to generate scramble, nsLSD1 or uLSD1 hairpins reservoirs.

Bacterial expression vectors

pET28b-TEV-LSD1 (aa157-852) (Forneris et al. 2005)

pGEX-6P-CoREST (305-482) (Forneris et al., 2007)

pET28b-TEV-LSD1-2a/8a (aa157-876): a fragment containing E2a and E8a was obtained from the IMAGE clone 4815528 digested with HindIII/SpeI and substituted in pET28b-TEV-LSD1 opened with HindIII/SpeI.

pET28b-TEV-LSD1-E2a (aa157-872): a restriction fragment containing E2a was obtained from pET28b-TEV-LSD1-E2a/E8a cutting with NcoI and substituted in pET28b-TEV-LSD1 digested with NcoI.

pET28b-TEV-LSD1-8a (aa157-856): The fragment containing the wild type form of hLSD1 was obtained from pET28b-TEV-LSD1 digested with NcoI and substituted in pET28b-TEV-LSD1- 2a/8a digested with NcoI.

Total RNA extraction, RT-PCR, and QRT-PCR analysis

Total RNAfrom seeded cortical neurons was isolated using the RNeasy Mini Kit(Qiagen). Tissue samples were either obtained from TriReagent (Ambion Cat.9738.) RNA extraction kit or Qiashredder (Qiagen 79654) and RNeasy Mini Kit (Qiagen Cat.74104) according to the sample type and starting material. Briefly, tissues were lysed in TRI Reagent solution, mechanically homogenized and RNA was obtained by acid phenol-chlorophorm organic extraction, followed by DNA digestion. All extracted RNAs were DNA purified by the use of DNAse upon Qiagen RNAeasy Columns, with no regards of the extraction method. RNA purity was further assessed on corresponding cDNAs by specific PCR on the elective housekeeping gene, which is expected to amplify one extra amplicon in case of genomic contamination. Quantitative reverse transcription (qRT)-PCR analysis was performed with Reverse Transcriptase system (Promega Cat A3500 and A3802) or SuperScript III system (Invitrogen). QPCR were performed on either an iQ5 Real-Time PCR Detection System (Bio-Rad) using the iScriptTM two-Step RT-PCR Kit with SYBR Green (Bio-Rad) or a Stratagene MXqPro software. The relative expression of the investigated genes was normalized against β-actin and double checked with Rpsa protein. Primer pairs are reported in the following table. A first screening for the expression of LSD1 splice variants was performed on tissue derived human samples (Ambion First Choice Human Total RNA Survey Panel, Cat #6000 Lot#08608142) which were reverse transcribed by use of oligodT primers in order to select mRNA species only, excluding any possible truncated pseudogene or non polyadenylated transcript, that were predicted *in silico* for LSD1 locus (UCSD Genome browser). Starting material was set as 1 μg of total extracted RNA and quantified by Biophotometer (Biorad) or Nanodrop 1000 3.6.0. Specificity of PCRs reactions was validated on BGal4 vectors containing either BC048134 or BC040194 construct related to the RefSeq NM_015013 .

h, human; r, rat; m, mouse

Primer pairs used for PCR amplifications.

Primer pair	Sequence 5'-3'	Application
hPCR Ex1-Ex18	Fw-GAAACTGGAATAGCAGAGACTCC Rw-TAGTCATTTCCAGATGATCCTGC	To detect entire hLSD1 ORF
hPCR1	Fw-AGTGAGCCTGAAGAACCATC Rw-GGAACCTTGACAGTGTCAGC	To detect 2 hLSD1 isoforms (±2a) including E8a
hPCR2	Fw-AGTGAGCCTGAAGAACCATC Rw-TTTCTCTTTAGGAACAGCTTG	To detect 2 hLSD1 isoforms (±2a) excluding E8a
hBeta actin	Fw-GCGGGAAATCGTGCGTGACATT Rw-CTAGAAGCATTTGCGGTGGA	Control cDNA
mPCR1	Fw-AGTGAGCCGGAAGAGCCGTCTG Rw-AACCTTGACAGTGTCAGCT	To detect 2 mLSD1 isoforms (±2a) including E8a
mPCR2	Fw-AGTGAGCCGGAAGAGCCGTCTG Rw-ATCTTTTTCTTTTGGAACAGC	To detect 2 mLSD1 isoforms (±2a) excluding E8a
mLSD1	Fw-AGTGAGCCGGAAGAGCCGTCTG Rw-CTACCATTTCATCTTTTTCTTTTGG	Rqf-PCR
mBeta actin	Fw-GCCTCAGCAGACACAGAAGG Rw-AATGCCTGGGTACATGGTGG	Control cDNA
rBeta actin	Fw-GGCCGTCTTCCCCTCCATCG Rw-CCAGTTGGTGACAATGCCGTGT	Control cDNA and Real Time qPCR
rLSD1	Fw-GCCTCAGCAGACACAGAAGG Rw-GTTATAAGGTGCTTCTAACTGC	Real Time qPCR
rLSD1	Fw-GAAAATGAAAGTGAGCCCGAGG Rw-CTACCATTTCATCTTTCTCTTTTGG	Rqf-PCR
r GFAP	GFAP Fw-AACTGCAGGCCTTGACCTGC GFAP Rw-TGGTAACTCGCCGACTCCC	Real Time qPCR

adapted from Zibetti et al. JNS, Feb 17, 2010 30(7):2521–2532

Supplementary Table 1. Intron-exon boundaries of the human LSD1 gene

Intron	Relative Start	Relative End	Absolute Start	Absolute End	Length	Donor and Acceptor
2-2a	11183	24974	1:23229715	1:23243506	13792	GT-AG
2-3	11183	30934	1:23229715	1:23249466	19752	GT-AG
2a-3	25035	30934	1:23243567	1:23249466	5900	GT-AG
8-8a	39716	46607	1:23258248	1:23265139	6892	GT-AG
8-9	39716	49086	1:23258248	1:23267618	9371	GT-AG
8a-9	46620	49086	1:23265152	1:23267618	2467	GT-AG

Table 1: Analysis of intron-exon boundaries was performed on human sequence at locus Chr 1: 23214930-23284038 by ASPIC tools according to the current release ASPic v2.1 (http://t.caspur.it/ASPIC). Splice donors and acceptors sites are indicated at predicted constitutive and alternative splice sites.

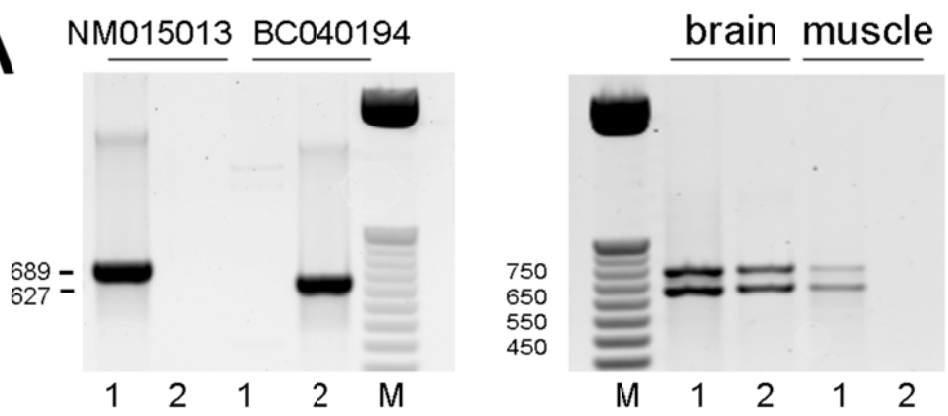
Supplementary Table 2. LSD1 splice isoforms related ESTs among mammals

Organism	Clone number	Tissue	Exon 2a	Exon 8a
Homo sapiens	mRNA BC040194	Hippocampus	+	+
Macaca fascicularis	CJ434064	Cerebellum cortex		+
Mus musculus	BG078465	Embrionic	+	
	CA526880	E14.5 Retina		+
Sus Scrofa	mRNA AJ656444	Embrionic	+	
Bos taurus	EE910679	Fetal brain	+	
	BF868311	Mammary gland	+	
	DV833797	Talamus		+
Gallus gallus	mRNA CR524286		+	-

Table 2. Mammal ESTs containing either or both exon E2a and E8a are indicated. ESTs were retrieved from UCSC database by means of *mRNA and ESTs tracks* tool and aligned against the genome using blat.

adapted from Zibetti et al. JNS, Feb 17, 2010 30(7):2521–2532

A



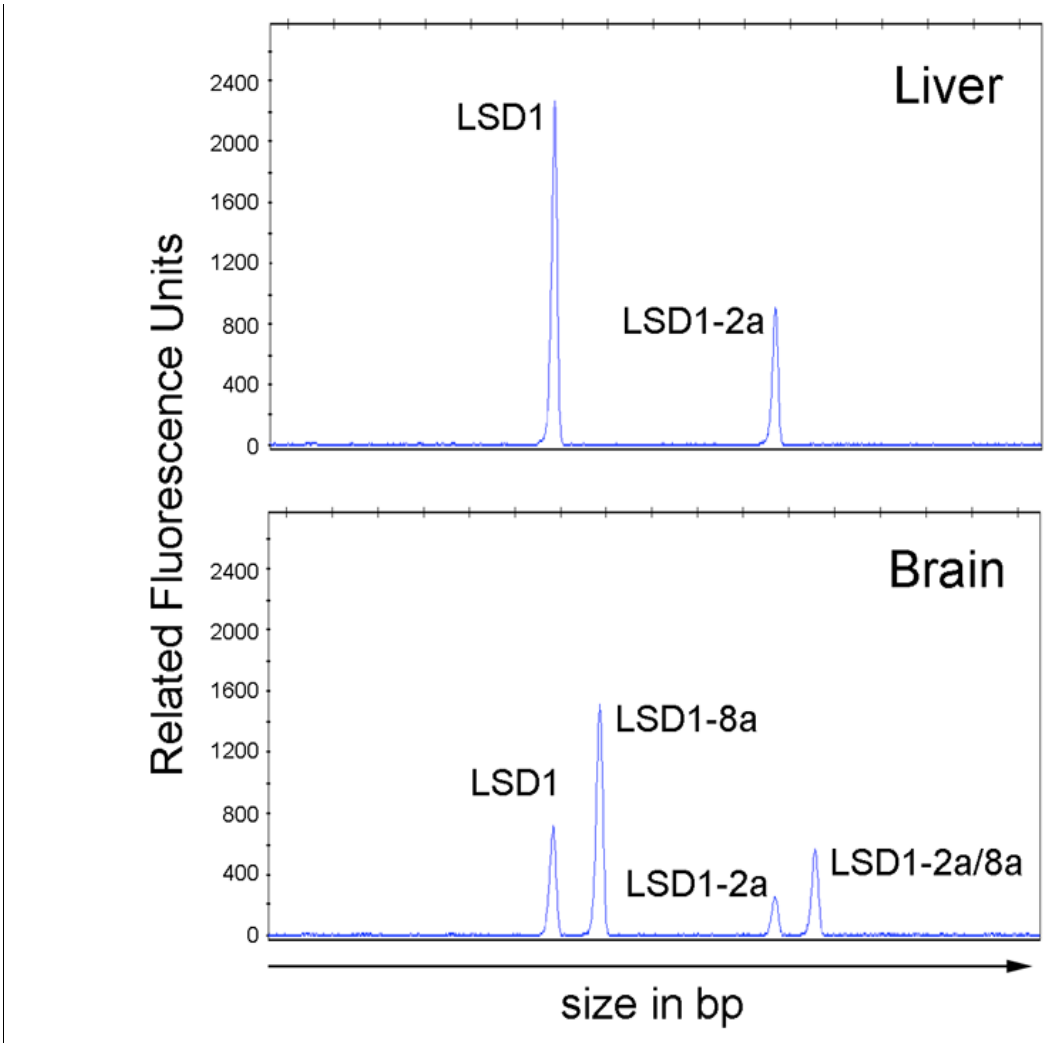
B

Expression of LSD1 splicing variants in human cell lines				
	-2a -8a	+2a -8a	-2a +8a	+2a +8a
IMR32	+	+	-	-
SH-SY5Y	+	+	-	-
HEPG2	+	+	-	-
HEK-293	+	+	-	-
HeLa	+	+	-	-
GRANTA2	+	+	-	-
SUDHL4	+	+	-	-
NTERA2	+	+	-	-
Linphoblastoid	+	+	-	-

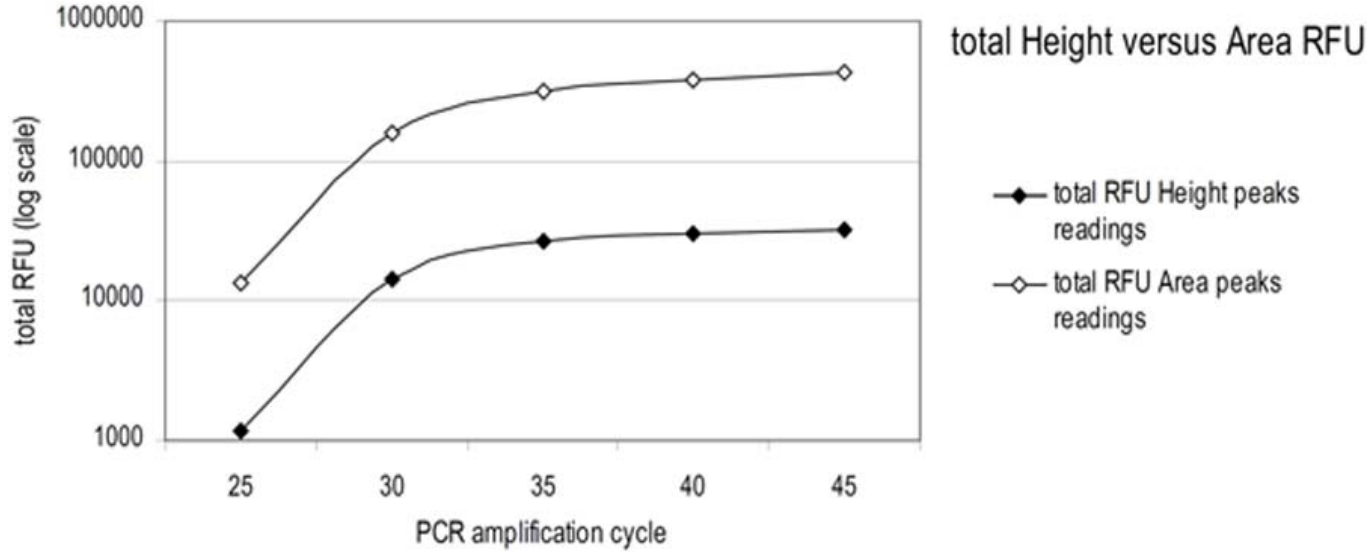
Supplementary figure 1. (A) PCR specificity to detect LSD1 splice variants was assessed on vector templates relative to human RefSeq NM015013, referred to as conventional LSD1 isoform and BC040194 clone, referred to as LSD1-2a/8a. Each PCR reaction allows selective amplification of either E8a including (here indicated as PCR2) or excluding (here indicated as PCR1) isoforms, with no regards for the presence of E2a. Inclusion of E2a results in the generation of differently sized amplicons in tissue samples. As shown, both PCRs reactions gave two products in brain tissue derived samples, which display the full pattern of LSD1 isoforms. Conversely, muscle tissue is devoid of E8a and display conventional LSD1 isoform and LSD1-2a. (B) Expression of LSD1 splice variants was probed on several human cell lines. Adapted from Zibetti et al. JNS, Feb 17, 2010 30(7):2521–2532

Relative quantity fluorescent PCR (Rqf-PCR) detailed protocol

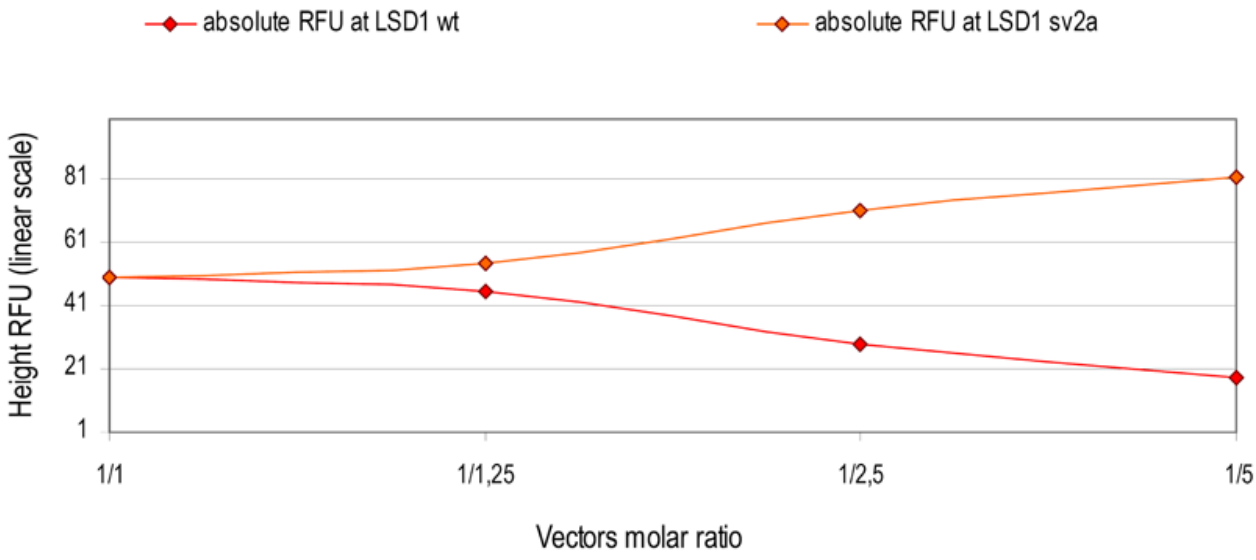
A fluorochrome-conjugated primer forward (Sigma Genosys Oligos) was used to quantify relative proportion of LSD1 isoforms in murine and rat samples. The oligo was modified at 5' termini (6-FAM conjugation) HPLC purified, resuspended in H20 upon shipping and always handled in the dark. PCR reactions were performed with GoTaqFlexi kit (Promega cat. M8305), with colourless 5X Flexibuffer (w/o MgCl2 added). PCR reaction was set up in order to amplify the whole extension of LSD1 transcript encompassing all the four identified isoforms: the extent of PCR products (within a range of 640-710 bp) and the reduced size difference among the generated amplicons (either 12 or 60 or 72 bp) complexively allow co-amplification of splice isoforms in non-competitive conditions, so that all LSD1 variants can be amplified with comparable PCR efficiency and the observed differences reflect the relative proportion of each analysed isoform. PCR was carried out within a linear range before saturation of the reaction could occur, as experimentally verified, always handled in dark, diluted and loaded with a suitable internal lane ROX-conjugated size standard (500ROX^{TM} Standard Applied Biosystem Cat 401734, Carlsbad, CA), denatured with formamide at 95°C for 5 min and ice cooled until electrophoretic separation. The sensitivity of the method was assessed on pCGN-HA LSD1 wt and pCGN-HA LSD1 2a vectors, by verifying the correspondence between the molar ratio of the templates and the associated RFU value (related fluorescence units). Amplicons were run on an AbiPrism Gene Analyser 3100 sequencer, separated by capillary electrophoresis on polymer POP 7^{TM} (Applied Biosystem). Electropherograms were displayed by use of software GeneMapper v4.0 (Applied Biosystem). Current settings are GS500(-250) as size standard, microsatellite default as analysis method with advanced peak detection method, analysis range from 0 to 10.000 and sizing range (bp) 1000. The baseline window is 51, with an rfu (related fluorescence units) peak threshold 50 for blu, red, yellow and green. The electrophoresis was carried on a 36 cm capillary at 60^{o} C with injection duration of 23 min at 1.2 KV. The overall run was carried for 32 min at 15 KV and laser power of 15 mW. Fluorescence threshold was adjusted for each experimental series. Each series was analysed at increasing PCR endpoints to ascertain anticipation of the plateau and signal saturation, with consequent loss of quantitative difference. Each isoform was expressed as percentage of the sum of all detectable variants. Percentage of expression was calculated on the electropherogram displayed height and compared with the area, obtaining comparable results.



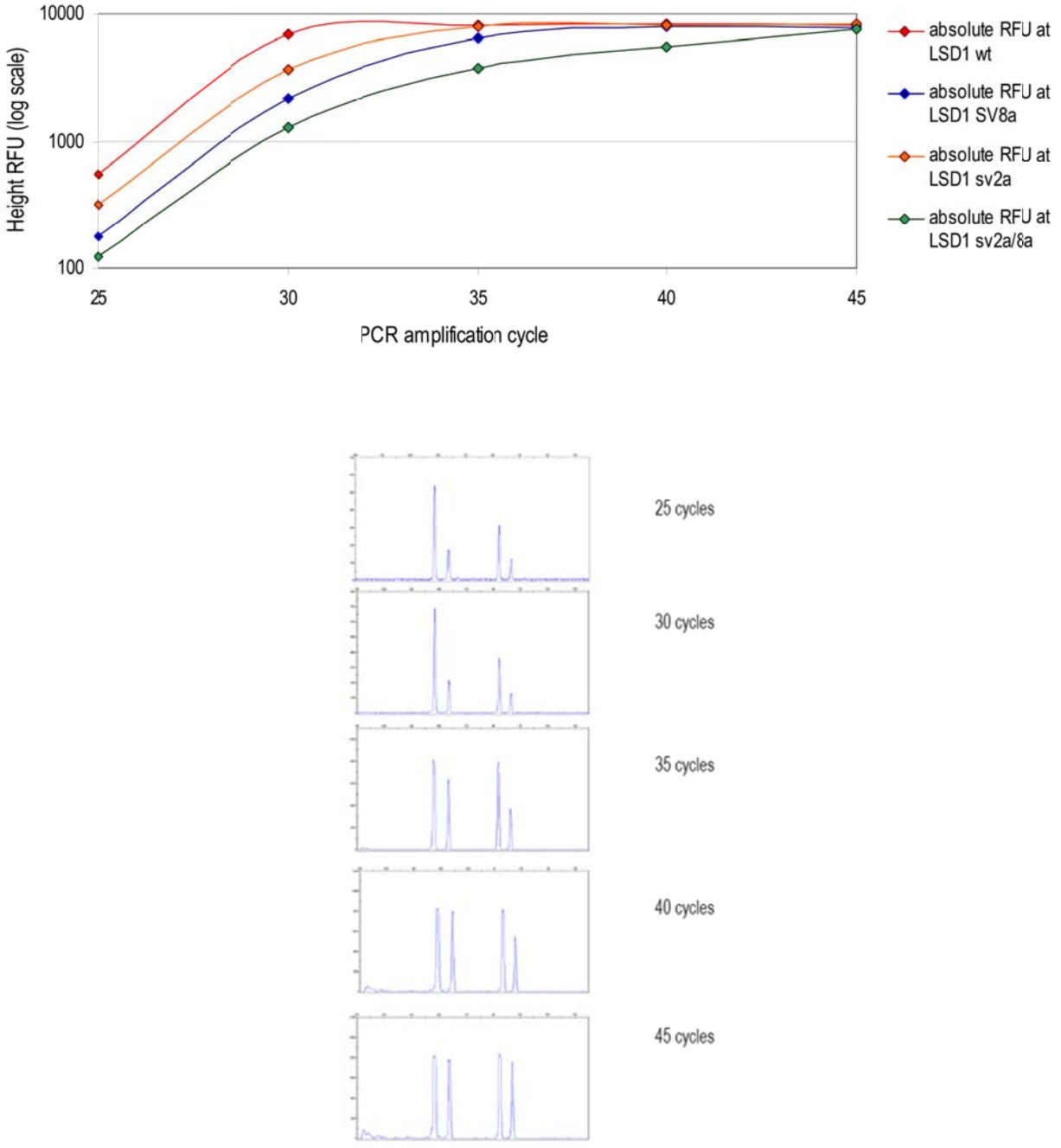
Supplementary figure 2.1 representative example of electropherograms related to rat liver and brain is indicated above



Supplementary figure 2.2 Cumulative RFU height and area (GeneMapper v4, Applied Biosystem) are plotted over PCR cycles with comparable increase within the PCR linear range. RFU height values were applied for calculations and statistics purpose.



Supplementary figure 2.3 Vectors related to the LSD1 splice variants were used to assess PCR efficiency and possible competition among molecular species being co-amplified. Spliced templates display no difference in PCR efficiency (see RFU raw units at the equimolar ratio). RFUs keep proportional at any assayed molar ratio (pCGN-HA-LSD1 wt and pCGN-HA-LSD1-2a), with no regard for the size difference (60 bp). The overall extension of templates and the narrow difference in size (bp) among the co-amplified products allow noncompetitive PCR conditions.



Supplementary figure 2.4 Definition of the linear range of the PCR. For each experimental series PCR was carried at subsequent endpoints to ascertain analysis within a linear amplification range. Here indicated a representative sample from brain tissue for which LSD1 isoforms were analyzed over subsequent PCR endpoints.

Reporter gene assays

Reporter readouts were performed at 48 hrs after transfection for HeLA, SH-Sy5y and cortical neurons. pGL3.1 vector was used as reservoir of 5xUAS -TK- Firefly luciferase gene (Photinus Pyralis) (Chen et al., 1998)(Image 1) providing reporter activity by default whereas pRL-TK-Ren (Image 2) was used as normalizer for variability factors, such as cell number and transfection efficency. LSD1 isoforms were conjugated to the Gal4-DBD domain, mediating their recruitment at the luciferase promoter. Control experiments were performed by using equivalent molar amounts of empty vectors (unconjugated Gal-DBD was assumed as 100% default luciferase activity). Different reporter: LSD1 repressor molar ratios (MR) were assayed, ranging from 1:2 to 1:0,03 to assess a dose-effect response, while keeping the total amount of transfected DNA constant by pBSII-KS (Stratagene). RLU (related luminescence units) values were acquired at Berthold Lumat LB 9501 luminometer.

Firefly RLU values were normalized to Renilla RLU upon background subtraction and averaged on at least three indipendent experiments. Graphed values refer to the Firefly luciferase activity resilient to transfection with LSD1 isoforms, with related error bars as S.E.M. 95% confidence intervals. Bilateral t-Student test (l Stat t I >= T ά/2) was performed for each experimental series, unless differently specified.

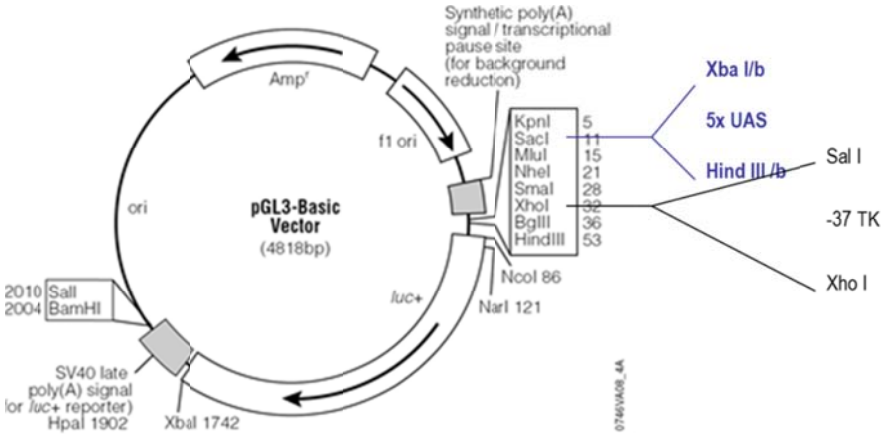


Image 1. G5-UAS-Tk Luc: pGL3-Basic Vector was used as reservoir of the Fairfly Luicferase gene (Promega) with the following modifications: 5x UAS sequences and a TK promoter were introduced to improve the default basal expression of the reporter gene(Chen et al. 1998)

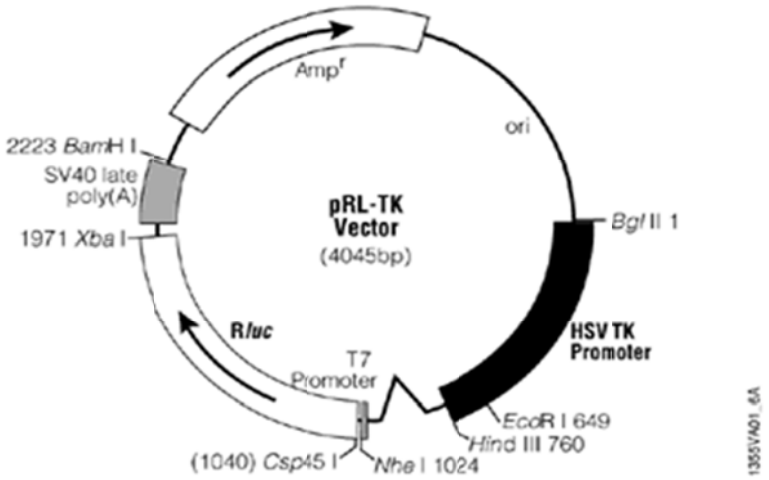


Image 2. pRL-HSV-Luc: Renilla Luciferase expression driven by HSV promoter was used to normalize for differences in transfection efficiency

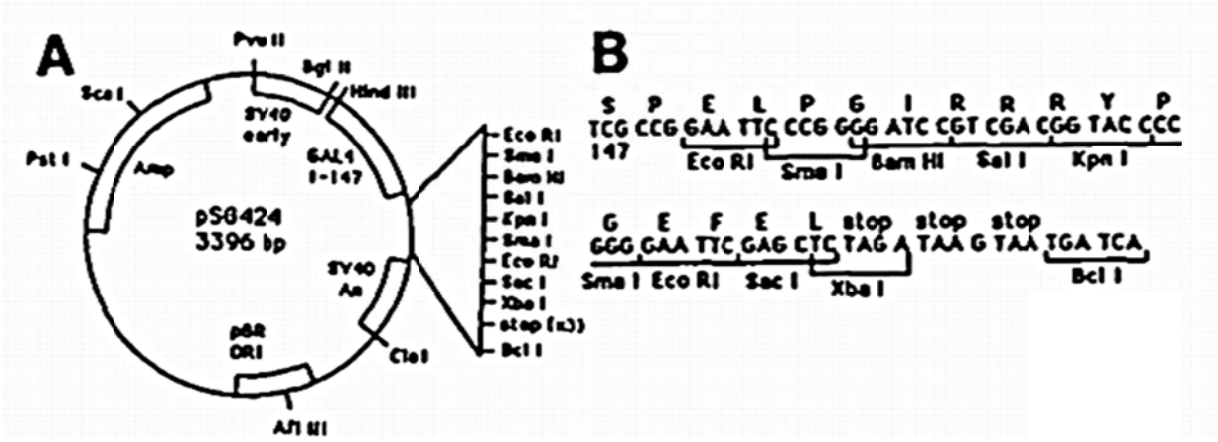
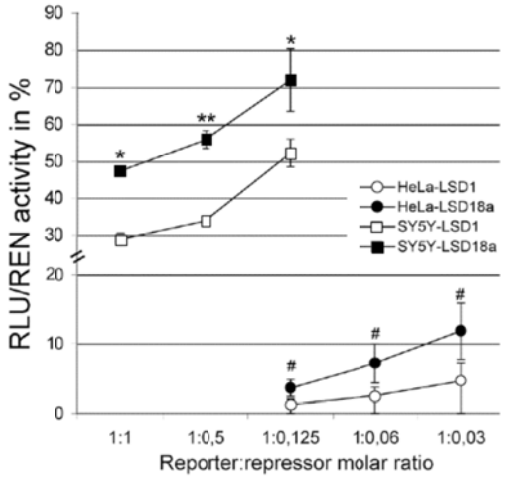


Image 3. pSG414 vector was used for the cobstruction and expression of GAL4 (1-147) fusions in mammilan cells. The plasmid contains the SV40 ori early prooter region fused to the coding seuqence for Gal4(1-147), followed immediately by a polylinker and translational stop codons. Transcripts are terminated within an SV40 DNA segment containg a polyA signal. B) Sequence of the polylinker region from aminoacid codon 147 of Gal4 and the peptide encoded by the polylinker (Webster et al.1998). For generation of the pGal4 conjugated LSD1 variants and pCGN-HA-ΔThr variants, refer to the Mammalian and bacterial expression section.



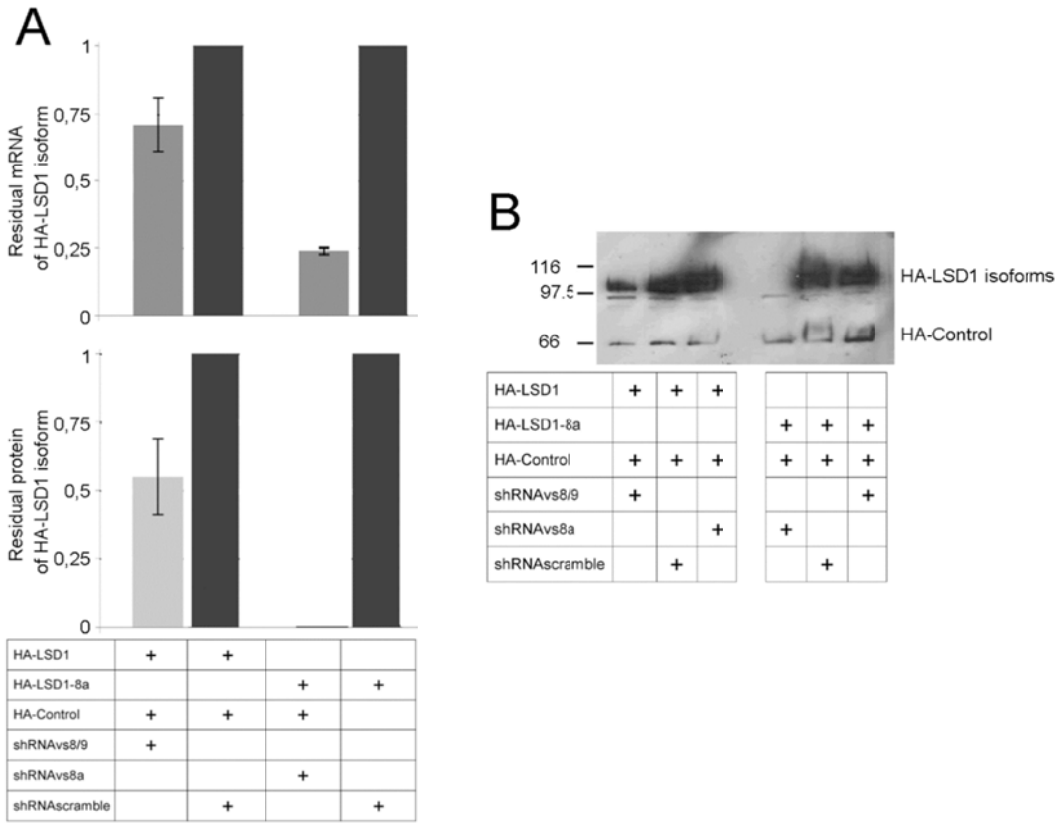
Supplementary figure 3. Repressive activity of LSD1 and LSD1-8a isoforms was compared in Hela and SH-SY5Y cell lines by the luciferase reporter assay. Here indicated normalized Firefly RLU/ Renilla RLU, derived from three independent experiments and indicated as percentage of Gal4 vector alone (default activity: 100% not shown) at different reporter: repressor molar ratios ranging for HeLA cells (circle series) and SH-SY5Y (square series). T-Student test was applied to percentage values between LSD1-8a series and LSD1. P-values (l Stat t I >= T ά/2) T-Student test on transfected series are indicated as wickets (Hela) or asterisks (SH-SY5Y).

Generation of shRNAs targeting LSD1 splice isoforms for transient expression assays

siRNA double stranded precursors were generated against LSD1 splice junctions and delivered by means of hairpins. Sequences were designed of 21nt length avoiding stretches of 4 or more bases, keeping the GC content within the range of 30% and 60% and avoiding SNPs, repeats and low complex sequences. Eight features associated with siRNA functionality (Reynolds et al. 2004) were assumed for a rational siRNA design algorithm: 30% to 52% GC content, at least 3 A/Us at 15-19 position in the sense stretch, lack of internal repeats (Tmelt* ≤ 20 °C), A at 19th position, A at 3rd, U at the 10th, no G/C at 19th, no G at the 13th, each compliant requirement being scored as +1. The siRNAs were scored and selected accordingly, assuming a value of six as probability cutoff for acceptable candidates: A BLASTn homology search was performed to predict possible off target effects.

*Tm = 79.8 + 18.5*log<sub>10([Na$^+$]) + (58.4 * GC%/100) + (11.8 * (GC%/100)^2) - (820/Length)

Oligos were matched onto the target sequence (inferred from RGD1562975) from *Rattus Norvegicus* species, for which the experiments were intended. Three types of hairpins were generated for the scramble condition, the knock down of the neurospecific variant and the knock down of the ubiquitous one. Oligos were annealed and cloned into into pSuper EGFP Neo vectors and assayed for specificity and efficacy against each representative LSD1 target sequence. Specificity of the hairpins was proven in a cell line heterologous to the rat species, to prevent cross-reactivity of the hairpins onto endogenously expressed targets and subsequent underestimation of the efficacy. The knock down was assayed both by QPCR on the residual expressed HA tagged isoforms and by immunoblot detection of the related proteins.



Supplementary figure 6. Efficacy and specificity of the shRNA vs 8a and shRNA vs 8/9 was evaluated in COS-1 cell line. pSuperGFP/Neo vectors were engineered to contain hairpins specifically targeted versus LSD1 isoforms; shRNAs were raised against LSD1 isoforms either with (LSD1-8a) or without exon 8a (LSD1), mediating silencing of LSD1 neuro-specific isoforms or ubiquitous ones, respectively.

Efficacy can be inferred from the reduced expression of HA tagged - LSD1 isoforms, normalized on a co-transfected HA-tagged Control, by a transcriptional (A) and proteomic (B and C) analysis. Knock down effect was evaluated by co-transfecting each hairpin with the related LSD1 isoform (i.e. shRNAvsLSD1 with LSD1; shRNAvs8a with LSD1-8a), while specificity was assessed by co-transfecting the same hairpin with the unrelated LSD1 isoform (i.e. shRNAvsLSD1 with LSD1-8a; shRNAvs8a with LSD1). The rat LSD1-wt isoform was generated by mutagenesis (Quick single site mutagenesis, Stratagene) of the related human sequence (pCGN-HA -LSD1 -wt) providing the exact target towards which the siRNA sequence has been designed. Mutagenesis was applied limitedly to the exons flanking the neurospecific ones. mRNAs expression was evaluated by Real Time PCR performed between the HA tag and ORF of the tagged construct, using the following primers :HA-Fw TATGACGTGCCTGACTATGCC and LSD1 Rw CCGCCTTCTTCCCAGATAAC or Control cDNA primer Rw 5' AAGTAGTAGCCTCTGATAATCC 3'. Here indicated each targeted isoform by using the corresponding shRNA (light grey bars) and compared to a scramble (dark bars). (A) shRNA vs 8/9 reduces by one-third the expression of the corresponding isoforms (0,7 ± 0,13 s.e.m.; n=3) whereas shRNA vs 8a suppresses most of the expression of the neuro-specific isoforms (0,23 ± 0,01 s.e.m; n=3). (B) LSD1 splice-specific knock down was also assessed by densitometric analysis of HA-tagged residual proteins. shRNA vs 8/9 reduces by half the expression of the corresponding protein (0,54± 0,14 S.D; n=3) whereas shRNAvs 8a completely abrogates the neuro-specific isoforms (0,0015 ± 0,01 S.D; n=3).

Immunoblot assay and antibodies

Western Blot was performed following "Molecular Cloning" by Sambrook-Fritsh-Maniatis method.

Briefly, samples were electrophoretically run by SDS-PAGE, transferred on a pre-equilibrated nytrocell membrane (Schleiher, Schuell), blocked in milk 5% w/v with 1‰Tween20, blotted with the suitable antibody in blocking solution either 4 hrs room T°C or overnight 4°C following the manufacturers' instruction, washed three times in TBS- 1‰Tween20, and blotted for 1 hrs room T°C with a secondary HRP-conjugated antibody. SigmaMarker High range (Sigma-Aldrich, 36,000 to 200,000 kDa) and PageRuler™ Plus Prestained Protein Ladder (Fermentas, 10-250 KDa) were used to characterize the migration of LSD1 doublet Antibody detection was performed with Supersignal West Pico (cat 34080 Thermo Scientific) or "Chemiluminescent Probe HRP" (PIERCE) and detection was performed on an autoradiographic lastra (GE). All protein extraction methods require pre-chilled buffers and fresh added protease inhibitors (Leupeptin, Pepstatin A, Aprotinin, Bestatin), such as protease inhibitors and PMSF 0.2mM.

Protein extraction was performed according to the sample type and starting material: adherent cultured cells were first washed with pre-cooled PBS, scraped and gathered at 1000 rpm for 10 min 4°C. Pellets were resuspended in 5 volumes RIPA (50 mM Tris pH 8, 150 mM NaCl, 1 mM EDTA, 1 % NP40, 0,5 mM DTT,PIC, PMSF) or low stringency lysis buffer (50 mM Tris pH 8, 0.5% TritonX100, 150 mM NaCl, 10 mM Imidazol, 0.5 mM EDTA, 10% glycerol, 0.5 mM DTT , Pic, PMSF) for co-immunoprecipitation purpose and rotated 30 min 4°C. Debris were gathered 15 min 14000 rpm 4°C and surnatant was retrieved and quantified following Bradford method, with Biorad Protein assay (cat 500-0006 BioRad) and a BSA calibrated standard curve.

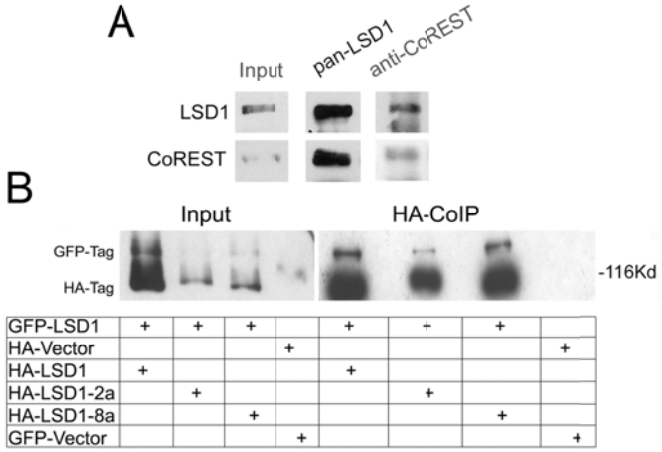
Cell cultures neurons were rinsed and directly lysed in 100 µl sample buffer (Laemmli buffer) and a volume of 12 µl was loaded on SDS PAGE for LSD1 expression analysis during in vitro maturation. Tissues were mechanically disrupted with scissors, pottered and resuspended within a suitable volume of RIPA /Low stringency lysis buffer (1 ml / g of tissue) and spinned at 14000 rpm 4°C for 30 min to pellet debris.

Antibodies used included the following: CoREST (Millipore), HA (Santa Cruz Biotechnology), HDAC2 (Millipore), panLSD1 antibody (Diagenode), Shank (NeuroMab), and C-terminus-binding protein (Cell Signaling Technology). Guanylate kinase domain-associated protein (GKAP), glutamate receptor-

interacting protein (GRIP), and calcium/calmodulin-dependent serine kinase (CASK) (rabbit; gifts from M. Sheng), neuronal nitric oxide synthase (nNOS) (Santa Cruz Biotechnology), GluR1 (Oncogene), and GluR2/3 (Millipore Bioscience Research Reagents).

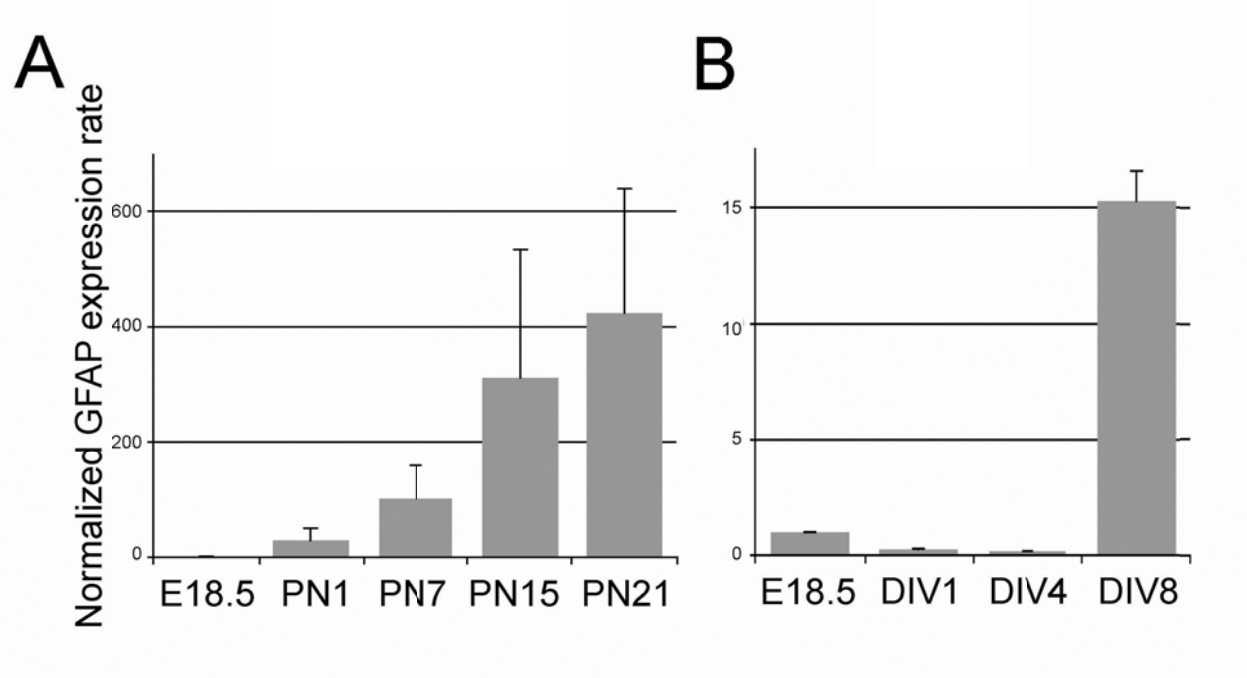
The antibody against E8a-containing isoforms was generated as follows: a peptide containing exon E8a GQADTVKVPKE was synthesized, injected into rabbits with a double boost and the resulting sera were purifies on a peptide affinity columns. Specificity and efficacy were tested on ELISA assay.

Further test was performed on heart, brain tissue samples and cortical seeded neurons over subsequent stages of development. The endogenous pattern of LSD1 variants was compared to total protein extracts from HeLA cells that were either transfected with the ubiquitous or neurospecific LSD1 variants.



Supplementary figure 4: (A) LSD1 and CoREST take part to the same complex in the central nervous system. LSD1 and CoREST immunocomplexes were isolated from adult rat brain whole extracts and resolved by SDS-PAGE. Western blots were probed with the indicated antibodies.

(B) The CoREST co-repressor complex can encompass two identical or different LSD1 isoforms. Whole cell immunocomplexes from HeLa cells overexpressing GFP-LSD1 together with pCGN vector, or HA tagged LSD1, LSD1-8a or LSD1-2a were isolated with GFP antibodies and resolved by SDS-PAGE. Western blot were probed with antibodies to HA. Input represents 5% of total cell extracts used in immunoprecipitation.



Supplementary figure 5: (**A)** GFAP expression was assessed by quantitative Real Time PCR on rat cortical tissues, from E18.5 embryonic stage to postnatal ones and (**B)** on cultured cortical neurons from E18.5 on along subsequent stages of development. GFAP was normalized on beta actin and expressed as average fold increase over E18.5. Bars indicate SD.

Coimmunoprecipitation assay

Coimmunoprecipitation (IP) experiments were performed on total protein extracts obtained from HeLA cells 48 hrs after transfection or on homogenized rat brain tissue in IP buffer (50 mM Tris pH 8, 0.5% TritonX100, 150 mM NaCl, 10 mM Imidazol, 0.5 mM EDTA, 10% glycerol, 0.5 mM DTT , Pic [Sigma Aldrich} , PMSF). Cellular extract (0.5-2 mg was reacted with 1.25 -3 μg of rabbit polyclonal anti-CoREST antibody (Millipore), mouse monoclonal anti-hemagglutinin (HA) (Santa Cruz Biotechnology), or 1.25 μg of mouse monoclonal anti-green fluorescent protein (GFP) (Roche Molecular Biochemicals) or rabbit polyclonal anti-LSD1 (Abcam) with overnight rocking at 4°C. The immunoprecipitates were collected with rProtein G Agarose (Invitrogen). After incubation, the beads were washed four times with the IP buffer containing 5% glycerol and 0.1% Triton X-100. The immunoprecipitates were then eluted with 2X SDS sample buffer (Laemmli buffer) and analyzed by Western blot.

Acidic silver staining for MALDI-TOF analysis of the LSD1 splice variants.

Coimmunoprecipitation of the LSD1 splice variants was performed on rat brain cortical tissues as described above, by use of a rabbit polyclonal anti-LSD1 (Abcam) antibody recognizing a common epitope. Samples were loaded on a 7% SDS-PAGE linear gel and silver stained as previously described (Dunn et al.1994) with acidic modifications compatible to MALDI-TOF analysis. All reagents were prepared with MilliQ water and stored at 4 C. LSD1 bands were excised at the correspondent molecular weight, in gel trypsinized and the resultant peptides were volatilized by MALDI –TOF. The full length LSD1 protein spanning 876 AA was identified by the broad coverage of the sequence. The four splice variants were identified by presence of peptides matching the spliced exons. Further fragmentation of the peptides was performed to detect the phosphorylation of the neuro specific exon at T371 residue of nLSD1-E8a (T391 residue in nLSD1-E2aE8a) , as predicted *in silico* by use of NetPhos v2.0 software (http://www.cbs.dtu.dk/services/NetPhos/), with a *m/z* shift from 1565,996 to 1725,022 attributable to a double phosphorylation status at Val and Thr residues (CPL**Y**EANGQAD**T**VKPKEK)

Expression and purification of recombinant LSD1 splice variants

Recombinant human LSD1 splice variants were expressed as truncated proteins lacking the N-terminal 122 residues, following the same protocols used for the conventional isoform (Forneris et al., 2005a). Proteins were purified by affinity chromatography, taking advantage of the N-terminal His6 purification tag. The tag was then removed by cleavage with tobacco etch virus protease. During protein purification, SDS-PAGE analysis showed that the LSD1 proteins were subject to rapid precipitation and unrecoverable degradation. The problem was solved by copurification with a C-terminal fragment (residues 305-482) of recombinant CoREST by tandem-affinity chromatography as described previously (Forneris et al., 2007). The resulting LSD1–CoREST complex is stable at 4°C for several weeks as judged by SDS-PAGE and activity assays (see next section).

Crystallization and structure determination

Samples of purified human recombinant LSD1 splice variants in complex with CoREST in 25 mM potassium phosphate, pH 7.2, and 5% (w/v) glycerol were concentrated using an Amicon Ultra concentration device (Millipore) to a final concentration of 8 mg/ml. Crystals were grown at 20°C by hanging-drop vapor diffusion method by mixing equal volumes of protein samples with reservoir solutions containing 1.2 M sodium/potassium tartrate and 100 mM N-(2-acetamido)iminodiacetic acid, pH 6.5 (Forneris et al., 2007). Crystals were transferred in a solution containing 1.6 M sodium/ potassium tartrate, 100 mM N-(2-acetamido)- 2-iminodiacetic acid, pH 6.5, 10% glycerol, and 2 mM Lys4Met H3 peptide (i.e., the 21 N-terminal residues of H3 with Lys4 mutated to Met) and flash cooled in liquid nitrogen. Data collections were performed at the beamlines ID14-EH1 and ID23-EH2 of the European Synchrotron Radiation Facility. Data processing was performed using the program MOSFLM (Leslie, 1999) and CCP4 (Collaborative Computational Project, Number 4, 1994). The structures of the LSD1–CoREST complexes were solved by molecular replacement using the program AMORE of the CCP4 suite. The initial search model was generated from the Protein Data Bank entry 2IW5, deprived of cofactors, substrates, and ligands. Refinement was performed out using Refmac5 (Murshudov et al., 1997). Data collection and refinement statistics are reported in Table 2. Structure analysis, validation, and modeling were performed using the program COOT (Emsley and Cowtan, 2004). Figures were generated with PyMOL (www.pymol.org).

Biochemical assays

The enzymatic activities of LSD1 splice variants were measured at 25°C using a Cary 100 UV/Vis spectrophotometer (Varian) following published protocols (Forneris et al., 2005a) (Table 1). Methylated synthetic peptides corresponding to the 21 aa N-terminal tail of histone H3 were used as substrates for biochemical analysis (purchased from Thermo Fisher Scientific).

Cortical neurons cultures and immunostaining.

Cortical neurons cultures were obtained from embryonic day 18.5 (E18.5) rat brain (Charles River or Harlan) as described previously (Romorini et al., 2004). Neurons were plated on 12-mm diameter optical grade coverslips, and grown on 12-well plastic tissue culture plates (Iwaki; Bibby Sterilin or BD Bioscience). Calcium Phosphate transfection was performed as elective method for gene delivery, intended for subsequent reporter assays and immunostaining procedures. Briefly, transfection was carried out on E18.5 seeded cortical neurons as previously described (Xia et al. 1996) in serum free conditions with medium consisting of DNA vectors, CaCl2 125 mM and Hepes Phosphate buffer and allowed to precipitate, with a resulting 2% transfection efficiency. Low efficiency allows single cell analysis by immunostaining, preventing any possible overlap of neurites from adjacent neurons that may interfere with the morphometric analysis. Cells were fixed after 48 hrs with a PBS solution containing 4% paraformaldehyde for 10 min at room TC followed by incubation in absolute methanol for 10 min at -20C. Cells were incubated with anti-HA antibody (1:100; sc80; Santa Cruz Biotechnology) for 3 h in GDB buffer (30 mM phosphate buffer, pH 7.4,containing 0.2% gelatin, 0.5% Triton X-100, and 0.8 M NaCl), followed by 1 h incubation with cyanine 3-conjugated secondary antibody (The Jackson Laboratory) and mounted in antifading VectaDAPI medium (Vector Laboratories). For the used vectors Vectors intended for the overexpression of the LSD1 isoforms, the knock down or the expression of the mutagenized neurospecific isoforms are listed above in the mammalian and bacterial expression vectors section.

Image acquisition and morphometric analysis

Images were acquired on a Carl Zeiss LSM5 510 laser scanning confocal microscope. Acquisition was performed with a chromatic aberration-corrected objective (plan-apochromatic 1.4 oil NA DIC M27) in the provided Zeiss Immersol. Scanning lasers power from each source was set by default at minimum (Argon 488 nm and HeNe 543 nm lasers, HeNe 633 nm were used for the acquisition of neurite morphology whereas Diode 405 nm was used for DAPI counterstain). Detector gain and amplifier offset were adjusted for each image according to the palette range indicator. Multi track configuration was set on the single channel alternate acquisition (BP filters 505-530 and BP 560-625). Optical slice thickness was adjusted for each image ranging from 3 to 4 μm, with a 0.75 to 1 μm step per section, to cover the whole neurite arborisation along the Z axis. Image frame size was chosen by default at 1024 x1024 x 4 pixels, with a working conversion factor of 0.197 μm per pixel on orthogonal plane. Optical sections were merged as z-stack projections, exported LSM Meta 8-bit .lsm file and analysed on ImageJ software with correspondent scale settings. Sholl's analysis principle (Sholl, 1953) was applied to analyse neurite thickness, cumulative arborisation and branching by overlapping concentric circumferences over neuronal somas at increasing radii of 20, 25 and 30 μm. For each analysed neuron, neurites thickness was averaged on the radian arcs measurements; z-stack splitting and image analysis on isolated optical slices was always applied whenever juxtaposition of adjacent saturated neurites occurred on the joint projections. Statistical significance was assessed by bilateral Student's t-test (l Stat t I >= T ά/2) unless differently specified (l Stat t I >= T ά). The indicated n in the text for each experiment represents 3 independent experimental trials with the related SEM 0.95 confidence interval. For each trial a minimum of 15 neurons were analysed per experimental condition. For illustrative purpose, images are displayed with possible saturation of one or more channels and represented as 3D projections.

All the experimental conditions included EGFP to evaluate neuronal morphology: EGFP vector was transfected alone (mock condition) or co-transfected with either pCGN-HA-LSD1 wt and sv2a vectors or pCGN-HA-LSD1 sv2a/8a and pCGN-HA-LSD1 sv8a vectors, respectively referred to as ubiquitous condition (uLSD1) or neurospecific one (nsLSD1). The knock down of isoforms was performed by use of pSuperGFPNeo generated vectors.

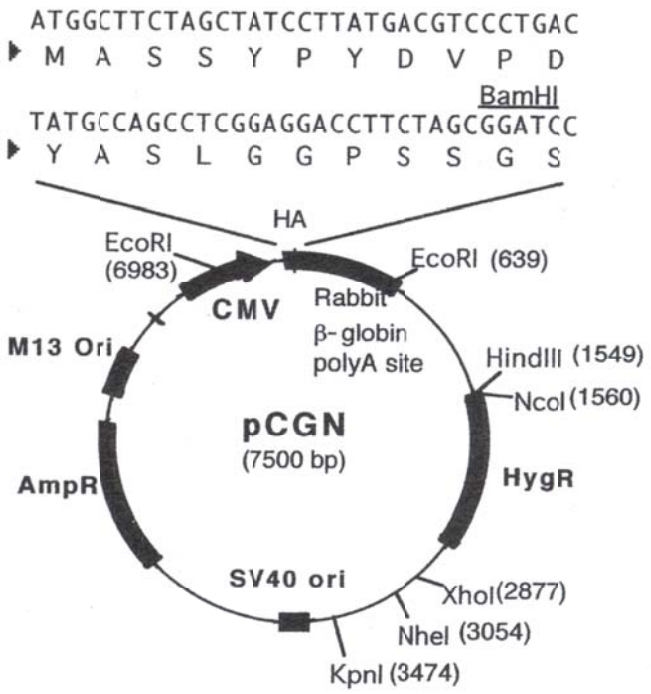
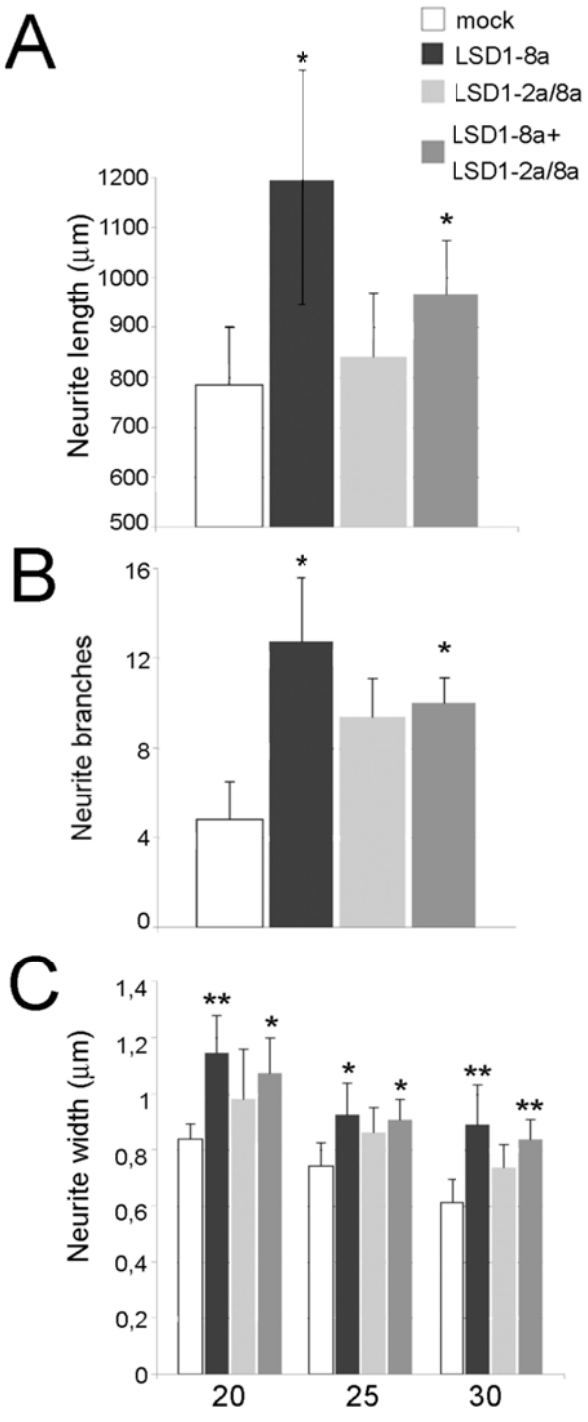


Image 4. pCGN –HA tagged backbone for delivery of LSD1 splice variants into cortical neurons


A
mock
LSD1-8a
LSD1-2a/8a
LSD1-8a+
LSD1-2a/8a
Neurite length (µm)
1200
1100
1000
900
800
700
600
500
*
*
B
Neurite branches
16
12
8
4
0
*
*
C
Neurite width (µm)
1,4
1,2
1
0,8
0,6
0,4
0,2
0
**
*
*
*
**
**
20
25
30

Supplementary figure 7: Morphometric effects of overexpression of single HA-tagged LSD1 isoforms in cortical neurons. (A) Cumulative neurite length ± SEM 0.95 C.I.: 784,8 ± 56,28 for mock series; 1194,7 ± 147,12 for neurons transfected with the sole LSD1-8a; 840,37 ± 126 for neurons transfected with the sole LSD1-2a/8a; 964,564 ± 53,43 for neurons transfected with the LSD1-2a/8a+LSD1-8a, referred to as neurospecific condition, namely nLSD1- see also Fig. 8). (B) Neurite branches ± SEM 0.95 C.I.: 4,8 ±. 1,71 for mock series; 12,7 ± 2,9 for neurons transfected with the sole LSD1-8a; 9,38 ± 1,71 for neurons transfected with the sole LSD1-2a/8a; 10 ± 1,1 for neurons transfected with the LSD1-2a/8a+LSD1-8a. (C) Neurite width measured at 20 μm distance from soma (inner radius) ± SEM 0.95 C.I.: 0,839 ±. 0,06 for mock series; 1,15 ± 0,13 for neurons transfected with the sole LSD1-8a; 0,98 ± .0,18 for neurons transfected with the sole LSD1-2a/8a; 1,07± 0,29 for neurons transfected with the LSD1-2a/8a+LSD1-8a. Neurite width measured at 25 μm (intermediate radius) ± SEM 0.95 C.I.: 0,74 ±. 0,14 for mock series; 0,92 ± 0,29 for neurons transfected with the sole LSD1-8a; 0,86 ± 0,21. for neurons transfected with the sole LSD1- 2a/8a; 0,9 ± 0,17 for neurons transfected with the LSD1-2a/8a+LSD1-8a. Neurite width measured at 30 μm (outer radius) ± SEM 0.95 C.I.: 0,61 ± 0,08 for mock series; 0,89 ± 0,36 for neurons transfected with the sole LSD1-8a; .0,74± 0,19 for neurons transfected with the sole LSD1-2a/8a; 0,84 ± 0,17 for neurons transfected with the LSD1-2a/8a+LSD1-8a. Two-tailed t-test was performed between mock condition and transfected LSD1 isoforms

Lentiviral production for silencing of LSD1 splice isoforms

A 2nd generation self-inactivating (SIN) system, characterized by the exclusion of four accessory genes (vpr, vif, vpu, nef) (Zufferey et al. 1997) and comprising the transgene vector, the pseudotyping and the packaging vectors was kindly provided by Dr. D. Trono (Lausanne, Switzerland) (Wiznerowic et al. 2003). The pLVTHM genomic vector carries the basic elements of 5'LTR, the major splice donor (SD), the packaging signal encompassing the 5' part of the gene (psi), the Rev-responsive element (RRE), the envelope splice acceptor (SA). The internal expression cassette contains the GFP encoded downstream the EF1α promoter and a 5' to 3' central polypurine tract (cPPT). The GFP is followed by the post-transcriptional regulatory element of woodchuck hepatitis virus to enhance gene expression (WPRE), which has been reported to increase the overall levels of transcripts in both producer and target cells, increasing titers and transgene expression and by the mutagenized 3'LTR containing the viral promoter in the RNA genome, hence resulting in replication-defective self-inactivating (SIN) vector (Zufferey et al 1998). In addition, the pLVTHM vector carries the H1 RNA polymerase III promoter to permit the expression of a short hairpin RNA (Abad et al. 2006) for RNA interference (shRNAi) (Wiznerowick et al. 2003).Hairpin sequences targeting LSD1 splice variants and a scrambled sequence were generated as described above. Lentiviral production was led onto Hek293T packaging cell line according to BSL2 biosafety level, FACS titrated (Becton Dickinson) and checked for replicative incompetence by measure of GFP on Hek293T cells after 36 hrs incubation with surnatant retrieved at the third replate of infected cells and different lentiviral batches were tested.

A 10^3 TU/μl titer was obtained from Hek293T cells at 48 hrs from lentiviral delivery infected in presence of charge neutralizing polybrene, under experimental conditions carrying 1% to 15% positive cells , with a minimum of 10.000 vital GFP+ve , Iodide Propidium –ve events. Lower transducing units were detected on neurons under the same experimental conditions, raising the need to further optimize lentiviral production and delivery routes. No difference in neuronal viability was detected between the experimental series, where either shRNAs or scrambled were delivered.

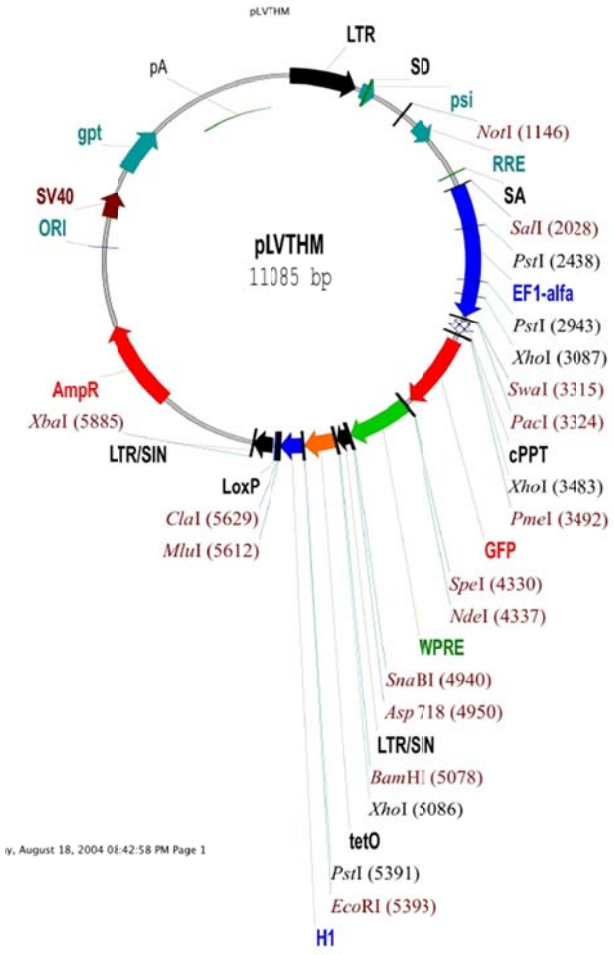


Image 5. pLVTHM transgene vector of 2nd generation for delivery of shRNAs targeted against LSD1 splice variants.

ChIP sequencing of LSD1 splice variants in cortical neurons

Two Sprague Dawley rat litters (Harlan) were sacrificed according to the IACUC animal policies. A minimum of 10 million neurons were used for high throughput ChIP-seq, at e18.5+div7 to characterize the neuro-specific LSD1 genome wide location. Chromatin immunoprecipitation of the four identified LSD1 variants was carried out by use of a rabbit polyclonal anti-LSD1 (Abcam) antibody recognizing a shared epitope.

Briefly, cells were rinsed in cold PBS in presence of protease inhibitors, cross-linked in freshly made 1% formaldehyde, quenched in 125 mM glycine and extracted nuclei were sonicated by Bioeruptor to produce 100 to 500 bp range sheared chromatin. Chromatin from a minimum of 10 million seeded cortical neurons was immunoprecipitated o.n. at 4 C as previously described (Shi et al. 2004) with a dedicated antibody or related IgG control, added for 2 hrs to pre-coated agarose beads (Invitrogen), washed and purified by organic extraction. DNA fragments coimmunoprecipitated with LSD1 were blunt-end repaired and phosphorylated using the T4 DNA polymerase, Klenow polymerase and T4 polynucleotide kinase 30 min at 20°C followed by column purification (QIAquick PCR Purification Kit, QIAGEN, part # 28104) and treated with Klenow exo$^-$ with dATP to generate 3' oligodA overhanging fragments 30 min at 37°C. Overhanging fragments were purified using the MinElute Reaction Cleanup Kit (Qiagen, Valencia, CA) and eluted in EB buffer. The resulting DNA fragments were ligated to adaptors 15 min at RT for the Illumina Cluster station–using the Illumina provided Ligase Kit (NEB), followed by gel purification and size selection: gel slices containing 175- 200bp adaptor-ligated ChIP DNA fragments were cut, shredded and further purified by gel extraction. Samples were affixed to a slide and sequenced on an Illumina Cluster Station and Genome Analyzer. The resulting 35 bp reads were formatted for alignment onto the indexed reference rat genome and visualized on UCSC genome browser. Peak calling was established by reads enrichment over a locally optimized *false discovery rate* tags threshold.

Computational analysis of high throughput Chip sequenced reads: genome indexing, reads alignment and tags density

ChIP reads were aligned onto the reference genome (Rat Nov. 2004-Baylor 3.4/rn4 assembly) comparing results from two multiple sequence alignment tools, differing for matching tolerances: bowtie (http;//bowtie-bio.sourceforge.net/index.html) (Langmead *et al.*, 2009), and GASSST (global alignment short sequence search tool, http://www.irisa.fr/symbiose/projects/gassst/html) (Rizk et al.2010).

Bowtie is a ultra-rapid short read alignment technique that applies backward search (Ferragina et al. 2000) after indexing the genome with a Burrows-Wheeler-FM Transform. Basically, it allows exact matches between query sequence and the genome to be found, with no gaps allowed, before using a backtracking procedure that allows the addition of a restricted number of errors. If one or more exact matches exist for a read, Bowtie will report one but if the best match is inexact high quality alignment may not be guaranteed. GASSST is a short read aligner for mapping reads with mismatch and *indel* (insertion-deletion) errors at a very high speed, discarding false positive positions before the refinement extension step: briefly, it applies a *seed, filter and extend* technique to globally align short sequences to local regions of complete genomes in a very short time. The *seed* step provides all potentially homologous areas in the genome with a given query sequence. An index of all possible k-mers in the genomic reference sequence is created and every query sequence from Chip reads is aligned iteratively onto the indexed genome to find matching *seeds:* only genomic sequences sharing common k-mers are considered for candidate alignments; their position and the flanking nucleotides are recorded for the subsequent filtering step. Sequences adjacent to the seeds are analyzed for discrepancies onto the indexed genome and filtered, to eliminate false positive hits that have more than a user specified number of errors. To include gaps, filtered alignments are extended with the Needleman-Wunsch dynamic programming algorithm (Needleman et al. 1970).

Peaks calling

Two different peaks calling methods, including Homer (Fig. 15 C and D) (*Hypergeometric Optimization of Motif EnRichment*, http://biowhat.ucsd.edu/homer/) and MACS (Fig.15 B)(*Model-based analysis of Chip-seq*, http://liulab.dfci.harvard.edu/MACS/index.html), were applied to identify Chip enriched regions, hereafter *peaks*, defined after clustered tags observed at a non-stochastic frequency. The tag density around a true binding site displays a bimodal enrichment pattern with an equal representation of upstream tags on Watson strand and downstream tags on Crick antisense strand, since ChIP-DNA fragments are equally likely to be sequenced at both adaptor-delimited ends. *Model-based Analysis of ChIP-Seq* (MACS) works on uniquely mapped short reads (35 bp tags on Illumina platform) taking advantage of the tags bimodal distribution and empirically models the shifting size to better locate the binding site. Given a sonication size (*bandwidth*) and a high-confidence fold-enrichment (*mfold*), MACS slides *2 bandwidth* windows across the genome to find regions with clustered non redundant tags more than *mfold* enriched relative to a random tag genome distribution. MACS randomly samples 1,000 of these high-quality peaks, separates their Watson and Crick tags, and aligns them by the midpoint between their Watson and Crick tag centers. The distance between the modes of the Watson and Crick peaks in the alignment is defined as '*d*', and MACS shifts all the tags by *d*/*2* toward the 3' ends to the most likely protein- DNA interaction site (*summit*) as peak center. After MACS shifts every tag by *d*/*2*, it slides *2d* windows across the genome to find candidate peaks with significant tag enrichment (Poisson distribution): candidate peaks with *p*-values below a user-defined threshold are called (p-value cutoff set at 10^{-5}). For a ChIP-Seq experiment with controls, MACS empirically estimates the false discovery rate (*fdr*) for each detected peak. The empirical FDR is defined as Number of control peaks / Number of ChIP peaks. MACS calculate the FDR based on the number of peaks from control over ChIP peaks that are called at the same *p*-value cutoff. This FDR estimate is more robust than calculating the FDR from randomizing tags along the genome.

HOMER determines the threshold of tags needed to call a peak significant by assuming that non-enriched ChIP-fragment concentrations are approximated by a stochastic distribution (Poisson distribution, p-value cutoff set at 0.001, tags *fdr* threshold at 11 *fdr* for bowtie and 9 *fdr* for GASSST alignments). Tags

positions are then adjusted of half the estimated ChIP-fragment length in the 3' direction relative to the original position of the tag. HOMER scans the chromosome calculating the number of clustered tags, found within the designated fixed peak size, sorting positions with the highest tag density through those with the lowest one assigning putative peaks and masking nearby positions within the minimum distance acceptable for neighboring peaks, so that local maxima densities are selected (minimum distance between peaks is set by default at 2.5 fold the estimated fragment size therefore 265 bp for bowtie aligned tags and 375 bp for GASSST aligned ones). Background filtering is applied for local and clonal tags: HOMER requires the tag density at peaks to be 4-fold greater than in the surrounding 10 kb region and discards clonal reads near repeat elements whenever the ratio between the expected tags containing unique positions at the peak and the observed tags containing unique positions is too high (fold enrichment limit of expected unique tag positions, default: 2.0).

Computation of transcription factors binding sites and *de novo* motifs by means of *Genomatix*® tools

To characterize extensively the pattern of putative TFs participating to the regulation of the identified LSD1 target genes, prediction of TFBS was performed by means of MatInspector® tool available online extending promoters regions (200 bp width) of 350 bp upstream and 50 bp downstream the located peak, since 600 bp long promoters are randomly extracted from MatInspector® library as control set for p-value calculation of TF enrichment (0.75/Opt.) (http://www.Genomatix® .de/online_help/help/scores.html). MatInspector® uses the core region of the matrix, represented by four consecutive nucleotides with the highest occurrence to preselect putative matches, the nucleotide distribution matrix, the Ci-vector describing the aligned nucleotides distribution and their conservation, the optimized threshold and the family information to scan sequences of unlimited length for matches to the consensus matrix description. The optimized threshold for the weight matrices is set by default at the minimum similarity level allowing up to three false positive matches in 10.000 bp of non-regulatory test sequences (core/matrix similarity cutoff set by default at 0.75/optimized).

De novo motifs search (CoreSearch®) was applied onto promoters annotated ChIP-seq regions (1000 bp<TSS) and a subset of intergenic annotated regions (3000-1000 bp<TSS) by means of CoreSearch ®tool, with a minimum core length of 7 bp, a minimum occurrence in 5% of input sequences on both strands, unlimited number of motif matches per sequence and equal nucleotide expected distribution. The matrix similarity threshold was set at 0.7.

Overrepresentation of TFs (RegionMiner®) was applied on the same sets without altering the extension of the input sequences, since overrepresentation calculation is based on the total number of base pairs in the considered sequence set to exclude any false positive match that may arise from a common TFs search task applied on extended sequences (MatInspector®). The following parameters were considered: *promoter association* referring to TFBS families known to occur more than twice as often in promoters as in genomic sequence, total number of input sequences displaying a match with the known TFBS matrix and the total number of matches. Further parameters are the *expected number of matches* (equally sized background sample retrieved from genome or non-related promoters), *overrepresentation* (fold increase of matches found in the input set relative to the observed number of matches found in the equally sized background set) and *the z-score* (distance from the population mean in units of the population standard deviation): the difference between the number of observed matches (input set) and expected matches (background set) is corrected per 0.5 factor and divided on the background set standard deviation (Sui et al. 2005) where a Z-score below -2 or above 2 can be considered statistically significant, corresponding to a p-value < 0.05.

RESULTS

Identification of novel LSD1 isoforms evolutionarily conserved in mammals

The annotated human *LSD1* gene (also known as *AOF2*) sequence (GenBank accession number NM_015013) encodes for a mRNA that results from the fusion of 19 exons and gives rise to a protein of 852 aa (Fig. 1*A*) whose biochemistry has already been characterized (Forneris et al.2005). I performed a computational analysis of LSD1 orthologues by means of Vista Genome Browser, by drawing genomic alignment at LSD1 coding region, with human sequence as consensus. at locus chr1: 23,218,576-23,282,768 at 1p36.12, as reported in *UCSC Genome Browser (Mar.2006 Human Assembly)* database. Sequence homology cutoff was set at 70% of conservation identity and results were graphed in a peak and valley graph, displaying percentage conservation at a given genomic coordinate within an overall window 800 bp wide. The comparative analysis revealed that LSD1 sequence is highly conserved among species and mostly among vertebrates with two regions emerging for the highest conservation degree nearby exon 2 and exon 8 (hereafter exon E2a and E8a, respectively) of BC040194 transcript originarily isolated from human adult hippocampus (Fig. 1B and C), indicating the possible existence of two additional exons subject to alternative splicing; remarkably, introns that flank such annotated exons display a high percentage of conservation, suggesting an elevate degree of conservation at regulatory splice sites. In this LSD1 isoform, the amino acids coded by E2a localize between the N-terminal disordered region and the SWIRM domain, whereas the four residues of E8a immediately precede the CoREST-binding tower domain, which is inserted within the amine oxidase domain (Fig. 1*A*). The alternatively spliced introns present canonical donor/acceptor splice sites (supplemental Table 1, available at www.jneurosci.org as supplemental material from Zibetti et al.2010) and display a very high conservation degree between human and mouse (Fig. 1*C*), a typical feature of alternatively spliced exons (Sorek and Ast, 2003).

Since inclusion of either predicted exons would neither alter the open reading frame nor introduce premature stop codons that would drive nonsense mediated decay, it was reasonable to hypothesize that full-length alternative spliced isoforms deriving from combinatorial inclusion of the predicted exons may retain functional properties. Therefore I performed an extended search for partial and complete transcripts containing one or both identified LSD1 exons was performed in the genomic sequences of different mammalian species, based on human LSD1 mRNA (GenBank accession number BC040194) (http://www.ncbi.nlm.nih.gov/BLAST/): all investigated species displayed at least one alternative exon and, most importantly, all E8a-containing expressed sequence tags (ESTs) derive from the nervous

system, as shown in supplemental Table 2 (available at www.jneurosci.org as supplemental material from Zibetti et al. 2010). I extended the analysis to other vertebrates to compare *AOF2* gene structure throughout evolution: whereas E2a was present and highly conserved in lizard (data not shown), chicken, and mammals, E8a is fully preserved in mammals only. E2a is 60 bp long and encodes for 20 aa, whereas the E8a is 12 bp long and is translated into 4 aa with sequence Asp-Thr-Val-Lys. The inclusion of the two exons does not alter the reading frame and results in a protein of 876 aa that I named LSD1-2a/8a.

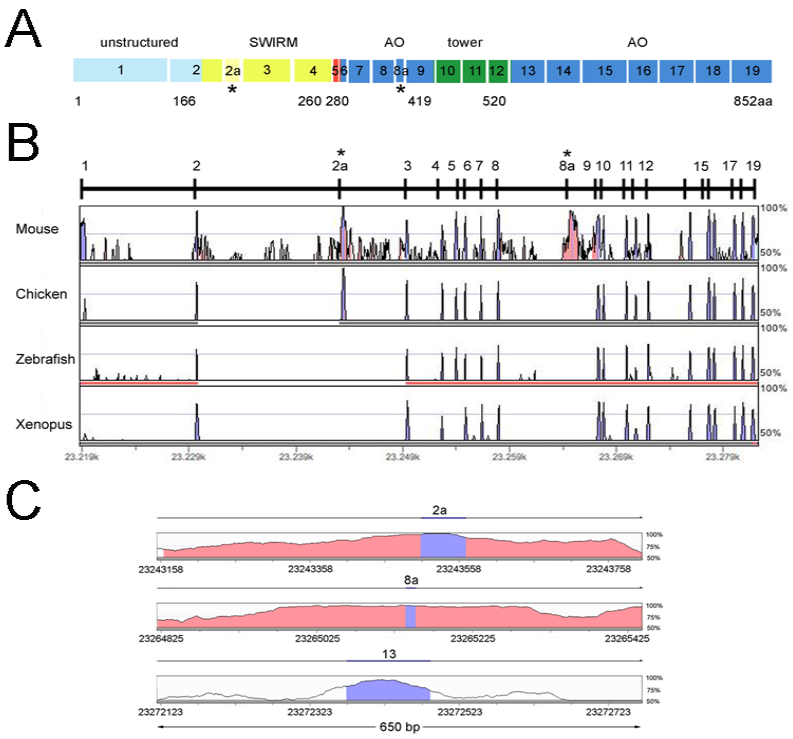


Figure 1. Genomic organization of human *LSD1* gene. ***A***, Schematic representation of the human LSD1 protein domains together with its exons ranging from 1 to 19; asterisks indicate the location of annotated alternative exons (E2a and E8a). Different colors indicate functional domains. N-terminal unstructured region coded by exons 1–2, SWIRM domain coded by exons 2– 4, the SWIRM-oxidase connector coded by exon 5, the amine oxidase domain coded by exons 6–9 and exons 13–19, and the tower domain (coded by exons 10 –12. Residue Met1 of this sequence corresponds to the first amino acid of the protein characterized (Shi et al., 2004). ***B***, Human *AOF2* alignment across vertebrates by GenomeVista browser (http://pipeline.lbl.gov/cgi-bin/GenomeVista). The "peaks and valleys" graphs represent percentage conservation at a given genomic coordinate between aligned sequences and the human sequence. Human exons are numbered. The top and bottom percentage bounds are shown to the right of every row. Regions of high conservation are colored as exons (blue) or noncoding (pink). Conserved regions are defined as regions with identity of 70% or higher that are wider than or equal to "minimal conservation width" (100 bp). ***C***, Enlarged view of 650 bp of the alignment between human and mouse intronic regions containing the two alternatively spliced exons (E2a and E8a) and one constitutively included exon (E13). Highlighted bars above the conservation area correspond to annotated E2a, E8a, and E13 of human *AOF2*. Dark gray areas within the conservation graph mark exons; light gray areas mark conserved (above 70%) non-exonic sequences.

Exon E8a-retaining LSD1 splice variant is restricted to neuronal lineage

The existence of an annotated human mRNA containing both E2a and E8a in an open reading frame prompted to the experimental validation of alternatively spliced isoforms containing the identified exons and to characterize their transcriptional profile on several samples derived from human, mouse, and rat tissues. I first investigated the presence of LSD1 splicing isoforms in a panel of total RNA samples from adult human tissues and cell lines (Fig. 2*A*,*B*) (supplemental Fig. 1, available at www.jneurosci.org as supplemental material from Zibetti et al. 2010). Through selective amplification of isoforms containing E8a (hPCR1), I found that LSD1-2a/8a was present in brain and testis, whereas the LSD1 with the sole E8a (LSD1-8a) was detected exclusively in brain tissues. Conversely, the amplification of isoforms excluding E8a (Fig. 2*B*,hPCR2) gave two products in all analyzed tissues, indicating that native LSD1 and LSD1-2a isoforms are ubiquitously expressed (see scheme in Fig. 2*C*). The identification of four mRNA species transcribed from the human *AOF2* gene indicates that the retention of E2a and E8a are two independent events and the inclusion of E8a is a tissue-specific event that occurs only in neuronal tissues and testis.

I extended the characterization of the tissue distribution of the four LSD1 splicing isoforms in adult mouse brain using the same approach performed on human tissues. As shown in Figure 2*D*, all the analyzed areas coexpress the four LSD1 isoforms and the overall amount of LSD1 measured by real-time qPCR showed comparable amounts of LSD1 transcripts in the different areas (data not shown). To further refine the analysis, I set up a method, namely rqf-PCR (see methods) in which co-amplified LSD1 isoforms are analyzed at high resolution by capillary electrophoresis.

This method allowed the detection of quantitative differences in the expression levels of LSD1 isoforms among brain areas. This analysis (Fig. 2*D*,*E*) revealed that the four isoforms are similarly expressed in all investigated brain regions, suggesting a controlled balance between the inclusion and exclusion events within adult murine and rat CNS. Furthermore, because neuronal tissues contain variable percentage of glial components, I asked whether the presence of the neurospecific LSD1 (herein referred to as nLSD1) doublet detected in all evaluated nervous tissues might be attributable to neuronal rather than glial histotype or both. The analysis performed on selected rat primary cultures, that is hippocampal neurons, cortical neurons and cortical astroglia by fluorescent rqf- PCR revealed that glia is completely devoid of nLSD1 isoforms but retained the ubiquitous LSD1 (uLSD1) ones (Fig. 2*F*), whereas both hippocampal and

cortical neurons retain the full pattern of the four identified isoforms, confirming that E8a inclusion within nLSD1 variants strictly relates to the neuronal histotype.

Because glial components (identified as GFAP-positive cells) do not exceed 2–3% in the evaluated neuronal cultures and the overall LSD1 protein levels did not appear to differ between neurons and glia (data not shown), I conclude that rqf-PCR quantification of coexisting LSD1 variants is reliably attributable to neurons.

Alternative splicing generates four functional proteins

Rat brain and heart tissues were assayed for the expression of LSD1 by SDS-PAGE and immunoblotting by a panLSD1 antibody revealed a discrete doublet of 110 kDa in both tissues (Diagenode).

I further characterized the doublet by comparing the migration of endogenous LSD1 with tagged isoforms individually transfected in HeLa cells. As shown in Figure 2*G*, the isoforms that contain exon E2a (LSD1-2a and LSD1-2a/8a) are upper shifted from the isoforms without it (LSD1 and LSD1-8a: the electrophoresis mobility displayed by the tagged LSD1 isoforms resumes the migration pattern of endogenous LSD1 proteins, suggesting that the doublet shared by and observed in neuronal and non-neuronal tissues is most probably attributable to the presence and absence of exon E2a (20 aa long) within LSD1. Furthermore, because inclusion of the sole exon E8a (4 AA only) produces no detectable variation in electrophoretic mobility of the 110 kDa protein, the LSD1 and LSD1-2a cannot be distinguished from LSD1-8a and LSD1-2a/8a, respectively. To demonstrate the existence of endogenous proteins containing the exon E8a, an antibody was specifically produced against a short peptide containing the amino acids Asp-Thr-Val-Lys (exon E8a sequence). The anti-E8a antibody specificity was tested by ELISA against the recombinant proteins (data not shown) as Ill as against the transfected cDNAs in HeLa cells (Fig. 2*H*). This antibody detected the presence of E8a-containing isoforms in rat brain tissues but not in heart, confirming that the four mRNA generated by alternative splicing are indeed translated in four proteins.

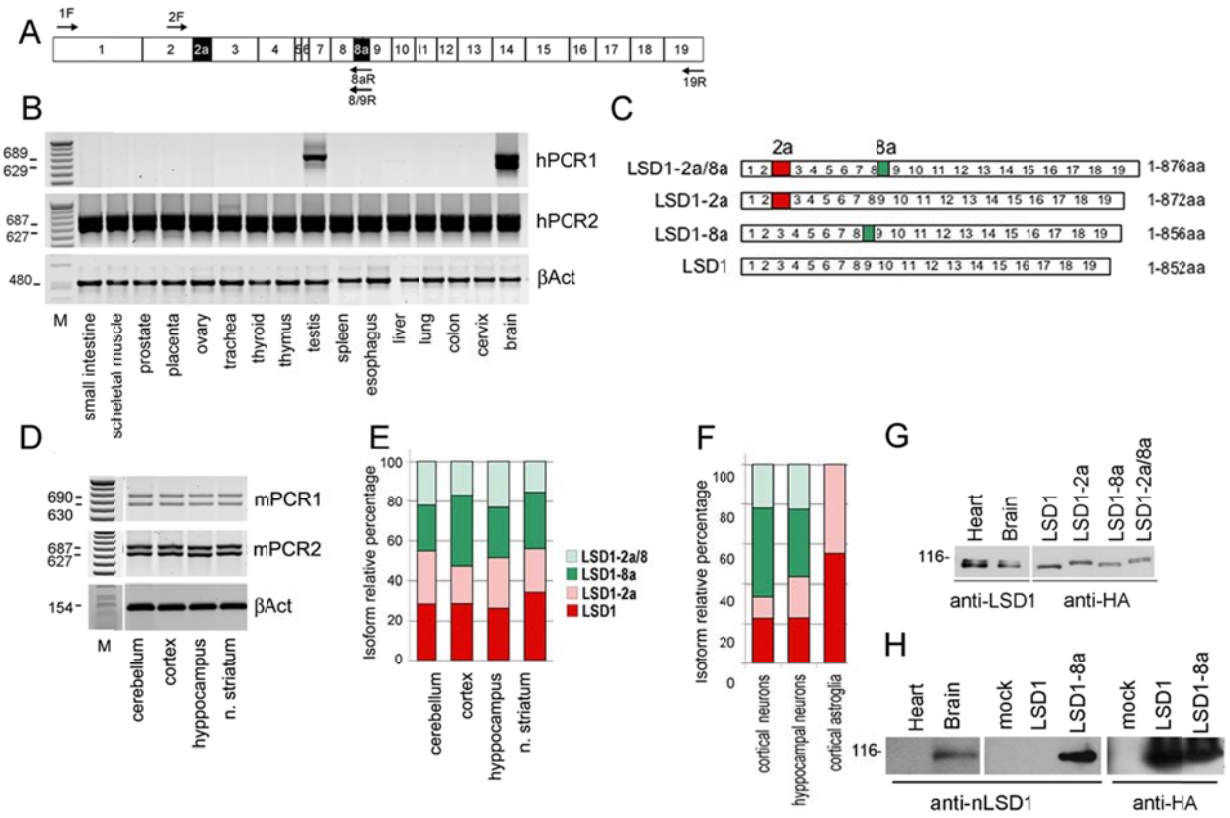


Figure 2. Mammalian LSD1 transcript undergoes alternative splicing and produces four isoforms with different tissue distribution. ***A***, Exon structure of the mammalian *LSD1* gene and position of the primers used to identify full-length, polyadenylated transcripts characterized by the presence or absence of E2a and E8a. Primers are indicated as F for forward or R for reverse, and the numbers indicate the exons where each primer anneals. Total RNA from human adult tissues (***B***) and total RNA from mouse adult brain tissues (***D***) were tested for LSD1 splicing variant expression by RT-PCR. The cDNA were amplified with primers covering the entire coding sequence (primers 1F and 19R) and re-amplified with two different nested PCR, one including E8a (PCR1) and one excluding E8a (PCR2). β-Actin was used as control. ***C***, Structure of the four LSD1 variants. ***E***, ***F***, Isoform-relative quantification of LSD1 splicing variants in mouse brain areas, primary neuronal cultures, and astroglia. rqf-PCR on cDNA obtained from total RNA of the indicated samples. Amplicons were quantified by related fluorescence units (RFU) by GeneMapper software and graphed as percentage relative to the sum of all the isoforms. Neurospecific LSD1 isoforms are shown in black and dark gray, whereas ubiquitous ones are shown in light gray and white. ***G***, Splicing generates four different LSD1 proteins. Western blots of total protein extracts from mouse brain and heart probed with a panLSD1 antibody and migration of recombinant LSD1 isoforms transfected in HeLa cells probed with anti-HA antibody. All the indicated samples were run on the same polyacrylamide gel. ***H***, Anti-LSD1-8a antibody specificity was assessed by Western blot on total protein extracts from heart and brain rat tissues and HeLa cells transfected with pCGN vector (mock), HA-tagged LSD1, or HA-tagged LSD1-8a cDNAs.

Effect of E2a and E8a inclusion on LSD1 enzymatic activity *in vitro* and three-dimensional structure

To evaluate the enzymatic activity of the four different LSD1 splice variants, the three human recombinant LSD1 isoforms (LSD1-2a/8a, LSD1-2a, LSD1-8a) were analyzed by comparative biochemical assays using histone H3 peptides. Both E2a and E8a had a strong destabilizing effect on the purified recombinant LSD1, which prevented their further biochemical investigation.

Fortunately, *in vitro* reconstitution of the complex formed by LSD1 with the corepressor protein CoREST by tandem-affinity purification strategies (Forneris et al., 2007) provided far more stable protein samples, enabling their structural and biochemical characterization. This finding indicated that the splicing variants retain the ability to form a stable complex with CoREST. Biochemical assays using histone H3 peptides revealed that all three LSD1 isoforms bound to CoREST can demethylate Lys4 of histone H3 with a catalytic efficiency virtually identical to that displayed by conventional LSD1 (Table 1). Moreover, they are totally inactive on peptides monomethylated on Lys9.

Recombinant LSD1-2a/8a, LSD1-2a, and LSD1-8a proteins were further investigated by x-ray crystallography. Crystals of their complex with the C-terminal region of CoREST (residues 305-482) and a 21 aa H3 peptide were obtained under identical conditions to those used for LSD1 protein (Forneris et al., 2007) (Fig. 3*A*). As for the native enzyme, also in the splicing variants, the N-terminal residues preceding Pro171 are disordered and not visible in the electron density map: this finding implies that, even in the presence of exon E2a (inserted between residues 170 and 171), the N-terminal region of LSD1 remains unstructured, at least in the crystalline state. Therefore, the structures containing E2a are indistinguishable from those lacking this insertion, the structural analysis should be focused to LSD1-8a in complex with CoREST and the histone peptide.

The overall conformation of LSD1-8a is very similar to that of the native protein with a root-mean-square deviation of 0.30 Å for 666 Cα atoms. The E8a residues Asp-Thr-Val-Lys inserted between Ala369 and Asp370 of the conventional isoform generate an antiparallel β-turn located in proximity of one of the two helices that define the tower domain (Fig. 3*A*). Structural superpositions show that the presence of E8a does not cause any local conformational change (Fig. 3*B*). Likewise, the conformations of CoREST and of the bound histone peptide are identical to those observed in the structure of LSD1–CoREST–peptide complex. The exon residues are located on the rim of the open cleft that forms the substrate-binding site, but they are not in direct contact with either the histone peptide or CoREST (Fig. 3*A*). This observation is

in agreement with the biochemical evidence that the enzymatic activity and substrate specificity of LSD1 splice variants are very similar to those of conventional LSD1 (Table 1). It is noteworthy that the E8a residues form a sort of protrusion that emerges from the main body of the protein. Such a feature indicates that these residues could easily form a docking site for other protein partners, which remain to be identified.

Table 1. Kinetic parameters for LSD1 and the splicing variants

	k_{cat} (min^{-1})[a]	K_m (μM)[a]
LSD1 + CoREST	7.35 ± 0.28[b]	5.12 ± 1.04[b]
LSD1-2a + CoREST	9.38 ± 0.65	6.31 ± 1.44
LSD1-8a + CoREST	5.19 ± 0.48	4.55 ± 1.65
LSD1-2a/8a + CoREST	5.47 ± 0.20	7.52 ± 1.00

[a]Apparent steady-state kinetic parameters were determined as described by using a 21 aa H3 peptide monomethylated at Lys4 (Forneris et al., 2005).

[b]Data were taken from Forneris et al. (2007).

Table 2. Data collection and refinement statistics

	LSD1-8a + CoREST complex with H3 inhibitor peptide[a]
Space group	I222
Unit cell (Å)	$a = 119.7$
	$b = 181.2$
	$c = 233.4$
Resolution (Å)	3.0
R_{sym}[b,c] (%)	11.3 (49.7)
Completeness[c] (%)	98.9 (99.8)
Unique reflections	50,357
Redundancy	4.1 (4.1)
I/σ[c]	10.2 (2.1)
R_{cryst}[d] (%)	20.7
R_{free}[d] (%)	24.9
RMS bond length (Å)	0.017
RMS bond angles (°)	1.81

[a]The final model consists of residues 171-836 of LSD1-8a (including residues Asp-Thr-Val-Lys corresponding to exon 8a, which are inserted after Ala369 and are named 369A-369B-369C-369D), a FAD molecule, residues 308-440 of CoREST, and residues 1-16 of the Lys4Met peptide.

[b]$R_{sym} = \Sigma|I_i - <I>|/\Sigma I_i$, where I_i is the intensity of ith observation and $<I>$ is the mean intensity of the reflection.

[c]Values in parentheses are for reflections in the highest resolution shell.

[d]$R_{cryst} = \Sigma|F_{obs} - F_{calc}|/\Sigma|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. The set of reflections used for R_{free} calculations and excluded from refinement was extracted from the structure factor file relative to the Protein Data Bank entry 2IW5.

Adapted from Zibetti et al. J. Neurosci., February 17, 2010 • 30(7):2521–2532

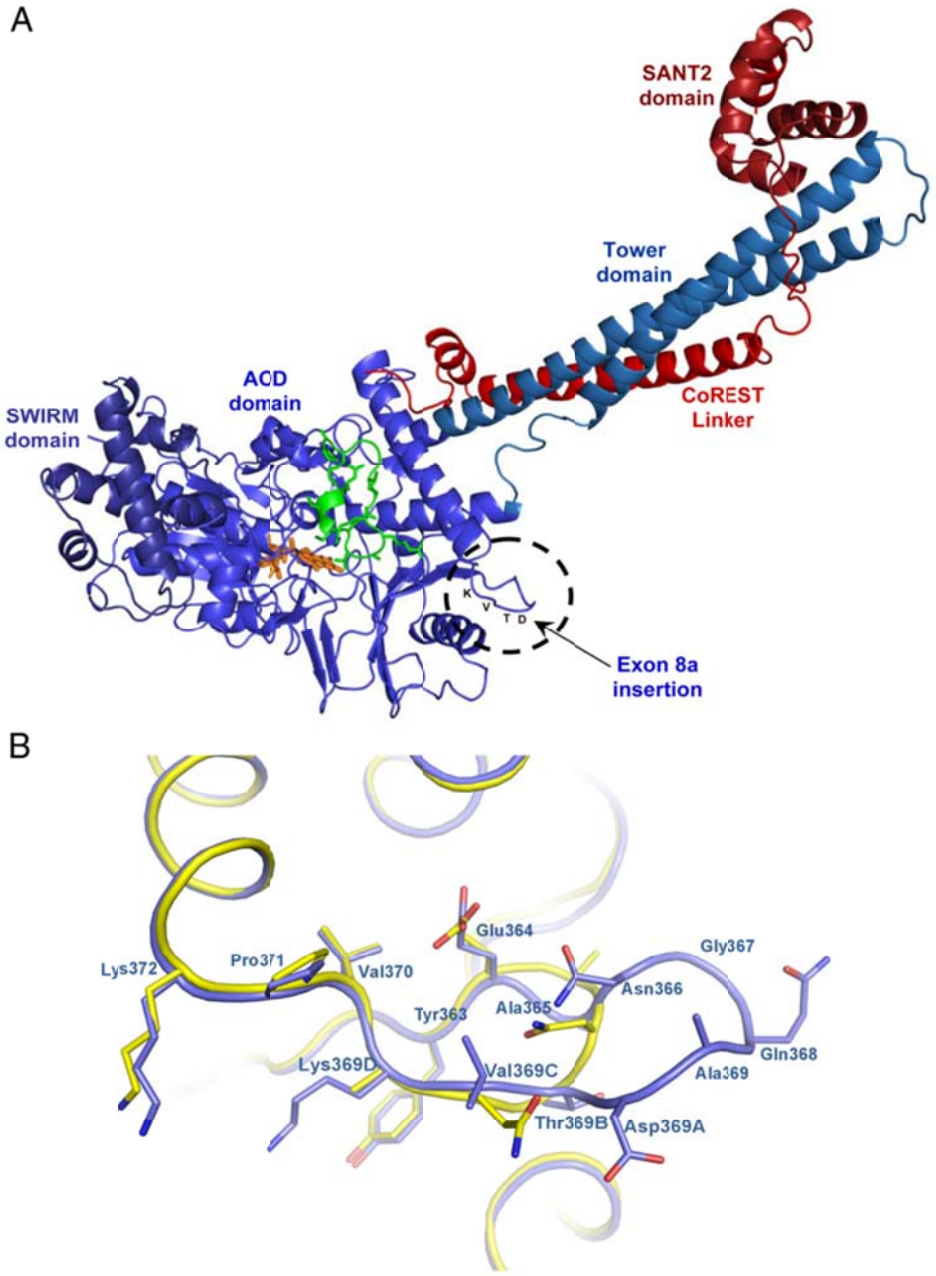


Figure 3. Comparative structural analysis of LSD1 and the LSD1-8a splice variant. ***A***, Overall crystal structure of LSD1-8a–CoREST in complex with a histone peptide. LSD1-8a (residues 171-840) is in light blue, CoREST (residues 308-440) in red, and the histone H3 peptide (residues 1-16) in green. The FAD cofactor is in the orange ball-and-stick representation. The insertion site of E8a (residues Asp369A-Thr369B-Val369C-Lys369D) is highlighted. ***B***, Close-up view of LSD1-8a structure at the site of exon E8a insertion. LSD1-8a structure is in blue, and it is superimposed onto native LSD1 (yellow; Protein Data Bank entry 2v1d) (Forneris et al., 2007). The orientation of the proteins is the same as in Figure 3*A*. The side chains of exon E8a residues are labeled in bold. Exon E8a insertion protrudes from the main body of the protein. Adapted from Zibetti et al. J. Neurosci., February 17, 2010 • 30(7):2521–2532

Inclusion of the neurospecific E8a exon modulates LSD1 repressive activity on a reporter gene

LSD1 repressive function on target genes was further investigated to determine whether the inclusion of either exon E2a or E8a in LSD1 isoforms might alter LSD1 regulatory activity. Therefore, I transfected LSD1 splice isoforms in different cellular systems and compared their effect on luciferase reporter gene. As shown in Figure 4*A*, in HeLa cells, the inclusion of exon E2a does not cause any change in luciferase expression (LSD1-2a vs LSD1, 1.31 ± 0.48 vs 1.12 ± 0.24; p = 0.34, *t* test), whereas the presence of exon E8a results in a significantly reduced repression of luciferase reporter, as evaluated in the HeLa cell line (LSD1-8a vs LSD1, 3.11 ± 1.16 vs 1.12±0.24, p =0.008; LSD1-2a/8a vs LSD1, 4.17±0.78 vs 1.12 ± 0.24, p =5.3*e*-16, *t* test). Differences in repressive activity were further evaluated on a scale of reporter: repressor molar ratios in HeLa and SH-SY5Y cell lines (supplemental Fig. 3, available at www.jneurosci.org as supplemental material). Notably, also in rat cortical neurons, in which the neuronal LSD1 isoforms are physiologically expressed, LSD1-8a and LSD1 exhibit different repressor properties (LSD1-8a vs LSD1 in Fig. 4*B* at a molar ratio 1:0.5, 89.41 ± 2.82 vs 69.99 ± 5.68, p = 0.018; molar ratio 1:1, 64.16 ± 4.01 vs 43.85 ± 3.22, p = 0.029, *t* test).

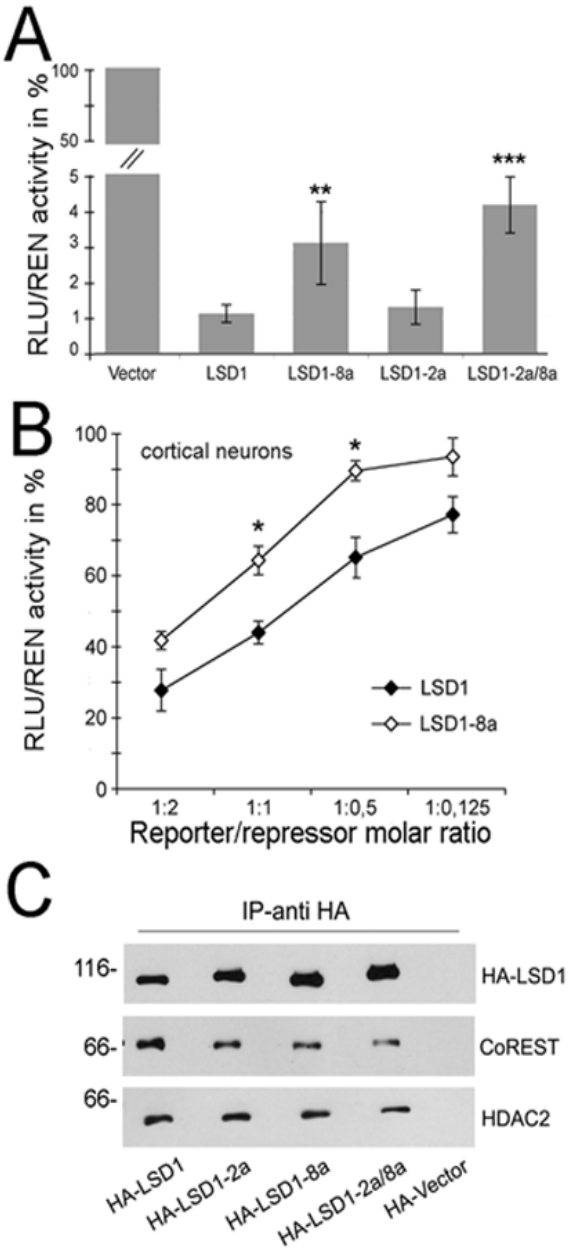


Figure 4. Effect of the included exon E2a and/or E8a on LSD1 repressor activity and affinity for corepressor partners. ***A***, The four indicated LSD1 splice variants fused to Gal4 –DBD were assayed for their ability to repress a reporter gene on HeLa cells at a constant reporter/repressor molar ratio. ***B***, In rat cortical neurons, LSD1 and LSD1-8a were compared at different reporter/ repressor molar ratios. The *luciferase* activity normalized over the activity of a co-transfected *renilla* reporter is expressed as percentage of the activity of the Gal4 –DBD vector at each molar ratio. Values are derived from at least three independent experiments. In ***A*** and ***B***, a Student's *t* test (*l* Stat $t\ I > \alpha/2$) was applied to percentage values by comparing splicing isoforms with LSD1. $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$. ***C***, Whole-cell immunocomplexes from HeLa cells overexpressing the indicated HA–LSD1 isoforms were obtained by HA antibodies and separated by SDS-PAGE. The Western blots were probed with antibodies to HA, CoREST, and HDAC2. Adapted from Zibetti et al. J. Neurosci., February 17, 2010 • 30(7):2521–2532

All recombinant LSD1 splice isoforms can be assembled into a CoREST complex

Given that LSD1 corepressor activity relies on the multiprotein complex it belongs to and the presence of E2a, E8a, or both might interfere with the formation of the complex with CoREST and associated proteins, HeLa cells were transfected and four LSD1 immunocomplexes were isolated and the presence of CoREST andHDAC2 was verified (Fig.4*C*). As well, LSD1 could be retrieved from brain derived CoREST immunocomplexes (supplemental Fig. 4*A*, available at www. jneurosci.org as supplemental material from Zibetti et al. 2010) although differences in enrichment of LSD1 splice isoforms can be detected (not shown). This finding, together with the biochemical and structural data (Fig. 3), demonstrate that inclusion of either exon does not preclude LSD1 interaction with known molecular partners and that LSD1 splice variants can function as corepressor factors similarly to the conventional isoform.

A mechanism to explain the functional implication of the LSD1 isoforms diversity might arise from the combinatorial incorporation of different isoforms. Indeed, assembly of higher-order HDAC1/2 complexes requires two hetero-trimers each formed by CoREST, LSD1, and either HDAC1 or HDAC2 (Humphrey et al., 2001). Therefore, to test whether different isoforms might be incorporated into the same higher-order HDAC1/2 complex HeLa cells were co-transfected with GFP-tagged LSD1 along with different HA tagged LSD1 isoforms. Supplemental Figure 4*B* (available at www. jneurosci.org as supplemental material) shows that it is possible to form hetero-oligomeric complexes in which different LSD1 isoforms are incorporated.

Expression of neurospecific LSD1 isoforms is regulated throughout brain development

Considering that the inclusion of E8a is a neuro-restricted event, I inferred its function within CNS by relating its expression profile to peculiar stages of neuronal differentiation, by means of rqf-PCR. I collected rat cortical tissues at several developmental stages (Fig. 5*A*,*B*) performing a time course analysis of the relative amount of each variant: at early embryonic stages, all of the four LSD1 isoforms are detectable, with preponderant expression of the LSD1 and LSD1-2a (42 and 38%, respectively; data not shown) over neurospecific isoforms (LSD1-8a and LSD1-2a/8a, 12 and 8%, respectively; data not shown). Later on, within the perinatal window between E18.5 and postnatal day 1 (P1), a rapid inversion of the proportions occurs: E8a- containing isoforms undergo a threefold increase shifting from 12 to 50% and from 8 to 25%, respectively, with a concomitant threefold decrease of LSD1 and LSD1-2a isoforms (Fig. 5*A*,*B*), and preponderance of E8a-containing isoforms is maintained until P7. Thereafter, LSD1 isoforms reach comparable levels and stabilize to the values measured in adult cortical cortex (compare with Fig. 2*E*). Time course analysis performed on cerebellar tissues revealed a similar expression pattern of LSD1 splice isoforms (Fig. 5 *G*,*H*). From these data, I calculated the inclusion frequency of either exon during development. Inclusion frequency of E2a, derived from the sum of LSD1-2a and LSD1-2a/8a (Fig. 5*C*, white squares), is rather constant, whereas the inclusion of E8a, calculated as the sum of LSD1-8a and LSD1-2a/8a (Fig. 5*C*, black squares), appears to be developmentally regulated. Overall LSD1 transcription and protein levels decrease between E18.5 and P1 and remain stable along subsequent stages of development (Fig. 5*D*,*E*). Because astroglia express only the ubiquitous isoforms (LSD1 and LSD1-2a) (Fig. 2*F*), a decrease in tissue glial composition may misleadingly indicate an increase in neurospecific isoforms (LSD1-8a and LSD1-2a/8a). Therefore, I also analyzed GFAP expression inferring glial relative contribution to isoform quantification. GFAP real-time qPCR analysis was performed from E18.5 to postnatal stages (supplemental Fig. 5*A*, available at www.jneurosci.org as supplemental material). Consistent with previous reports (Qian et al., 2000; Fox et al., 2004), GFAP is expressed at low levels in the perinatal window and constantly increases during postnatal development. This demonstrates that the detected increase of E8a is indeed caused by a neurospecific splice event, with no regard to glial composition. A parallel analysis showed that several synaptic markers arise during the perinatal window (E18.5 to P1) with a progressive increase over developmental stages (Fig. 5*F*), indicating that the inclusion frequency of E8a increases concomitantly with early stages of synaptogenesis.

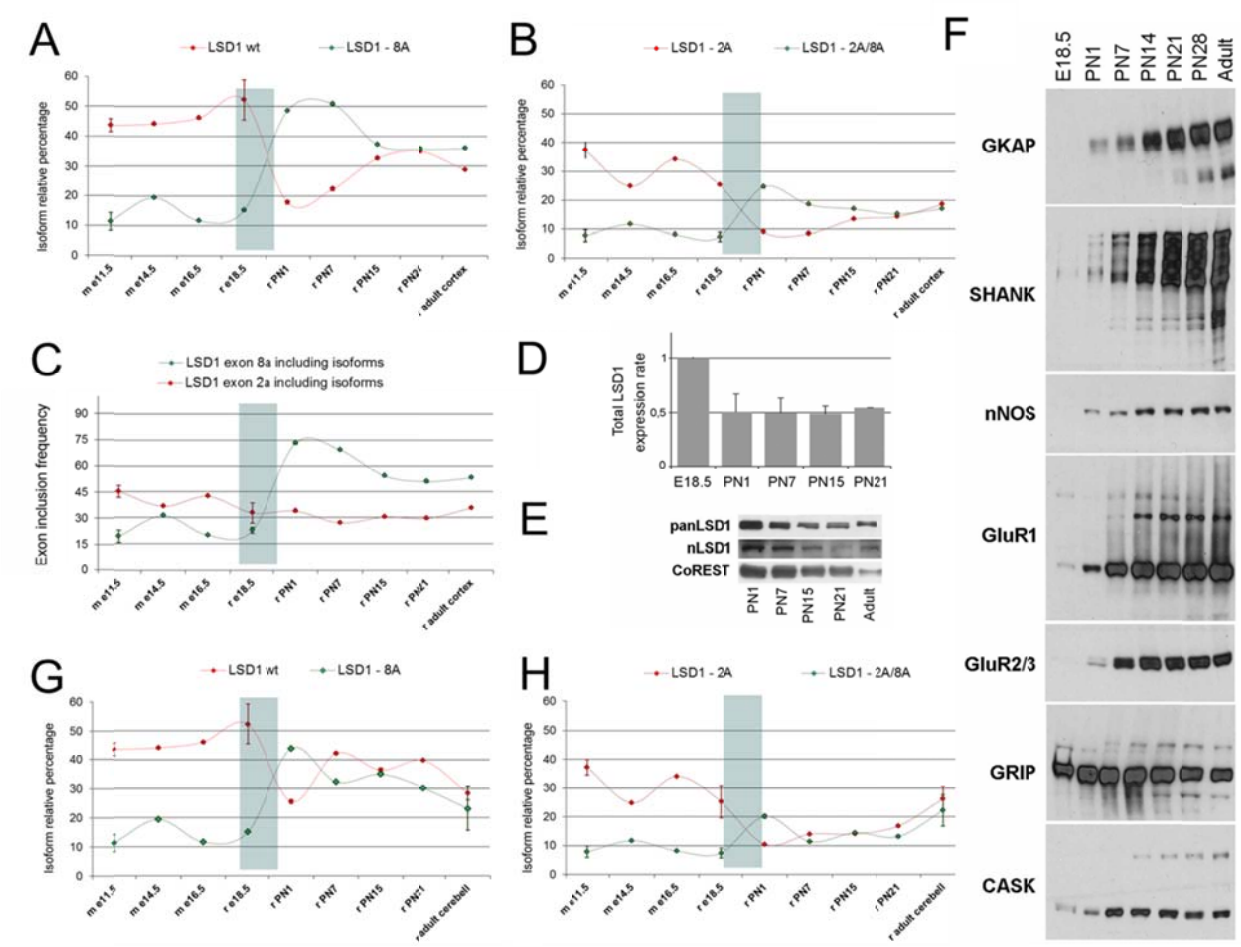


Figure 5. Inclusion of neurospecific E8a in LSD1 transcripts is dynamically modulated during neuronal development, whereas inclusion of E2a is a steady event. The relative amount of each LSD1 isoform was measured by rqf-PCR; cDNA were obtained from total RNA from rat embryonic cortex (E18.5), postnatal rat cortex (PN), and adult rat cortex. Graphs represent the relative percentage of each isoform with respect to the sum of the four. Only two isoforms are shown per graph; ***A*** compares LSD1 isoform with LSD1-8a, whereas ***B*** compares LSD1-2a isoform with LSD1-2a/8a. Values shown are mean ± SD. ***C***, Exon inclusion frequency of exons E2a and E8a. Each represented series relates to the overall inclusion of either E8a (black squares) calculated as the sum of LSD1-2a/8a and LSD1-8a relative percentage and E2a (white squares), calculated as sum of LSD1-2a and LSD1-2a/8a relative percentage, at each indicated developmental stage. ***D***, Total LSD1 transcript quantification by qRT-PCR on total RNA extract from the indicated rat cortex samples normalized on β-actin. Samples are expressed as fold increase relative to the LSD1 value at E18.5. Western blot on total protein samples from the indicated development cortical stages with a panLSD1 antibody or neurospecific LSD1 antibody (***E***) and with the indicated synaptic markers (***F***). Analysis performed on cerebellar tissue with exon E8a (***G***) and exon E2a (***H)*** . Note that nLSD1 can be detected as soon as E11.5 in mouse and E13 in rat at comparable crown-rump embryo lengths and corresponding to stage 14 in Carnegie scale of embryonic development. Adapted from Zibetti et al. J. Neurosci., February 17, 2010 • 30(7):2521–2532

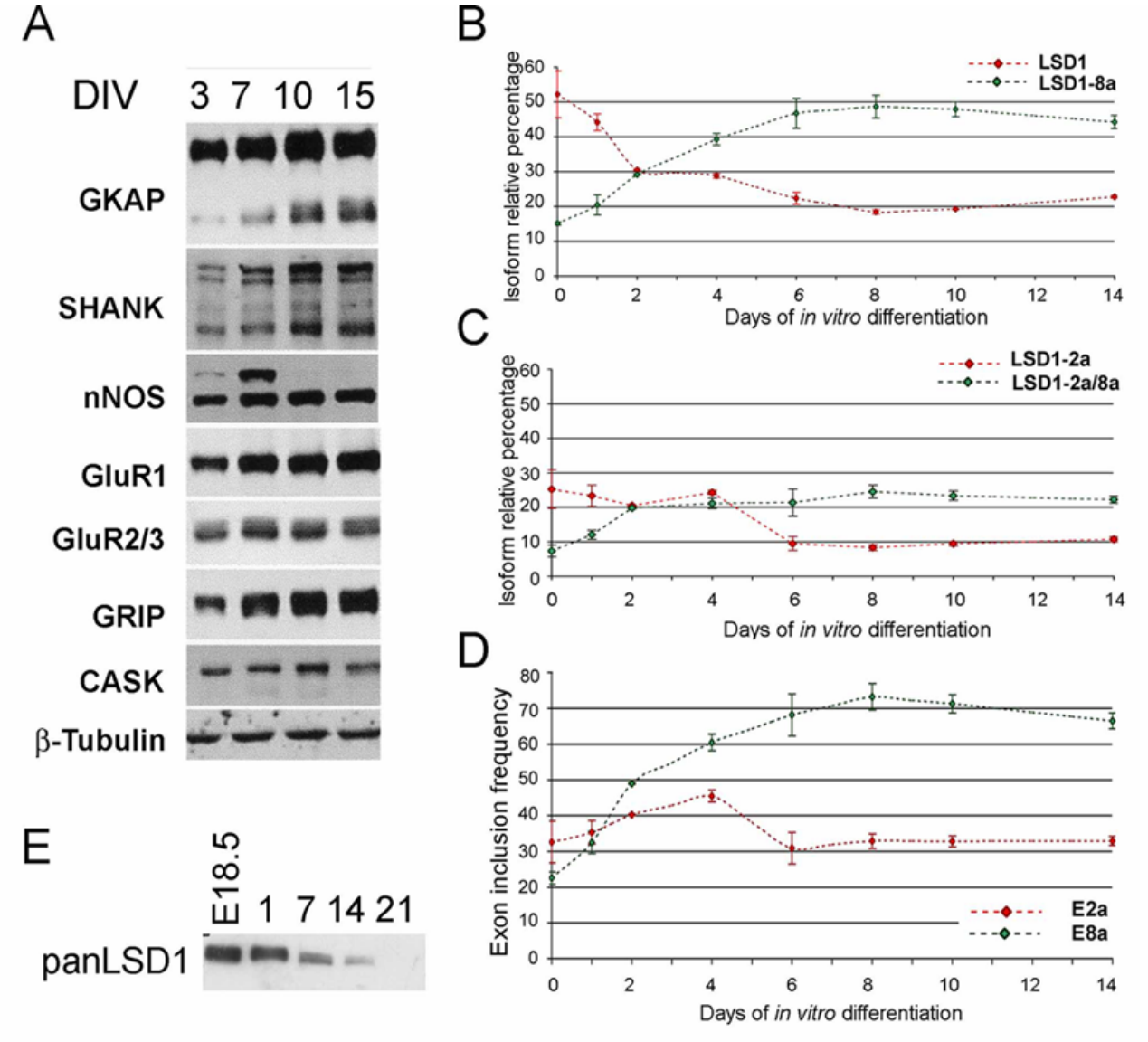


Figure 6. LSD1 splicing analysis in a rat cortical neuron maturation system. ***A***, *In vitro* maturation of cortical neurons prepared from E18.5 embryos (DIV0) was assessed by Western blot analysis of the indicated synaptic markers on total protein samples from the indicated DIV. Graphs represent the relative percentage of each isoform with respect to the sum of the four. Only two isoforms are shown per graph; ***B*** compares LSD1 isoform with LSD1-8a, whereas ***C*** compares LSD1-2a isoform with LSD1-2a/8a. Values shown are mean±SD. ***D***, Exon inclusion frequency of exons E2a and E8a. Each represented series relates to the overall inclusion of either E8a (black squares), calculated as the sum of LSD1-2a/8a and LSD1-8a relative percentage, and E2a (white squares), calculated as the sum of LSD1-2a and LSD1-2a/8a relative percentage, at each indicated developmental stage. ***E***, Western blot on total protein samples from the indicated DIV with a panLSD1 antibody. Adapted from Zibetti et al. J. Neurosci., February 17, 2010 • 30(7):2521–2532

nsLSD1 expression profile mirrors critical steps of neuronal development and contributes to neurite morphogenesis

Since the peculiar expression profile of LSD1 made reasonable to hypothesize its possible implication in neuronal development, I perturbed the expression of LSD1 isoforms in rat cortical neurons, which represent a suitable model of neuronal maturation, as assessed through the expression of synaptic markers mirroring postnatal development (Fig. 6*A*) (Lee and Sheng, 2000; Sala et al., 2000, Gaudilliere et al. 2004). This model also recapitulates the physiological expression pattern of LSD1 isoforms that I initially observed *in vivo*: the four LSD1 isoforms are all detectable at day *in vitro* 0 (DIV0) (corresponding to rat E18.5), with LSD1-2a and LSD1 being most abundant; as neuronal maturation proceeds, neurospecific E8a-containing LSD1 isoforms progressively increase (DIV2) and become preponderant (DIV6) (Fig. 6*B*,*C*). The overall inclusion frequency of either exon confirmed that E8a inclusion is developmentally regulated, whereas E2a inclusion does not change (Fig. 6*D*). Furthermore, LSD1 protein level analyzed by a panLSD1 antibody showed the same decrease during neuronal maturation that was observed *in vivo* (Fig. 6*E*). Also in this case, I verified whether any detectable variation of LSD1 isoforms might be influenced by a change in the proportion between neurons and glia during maturation *in vitro*. As indicator of astroglial contribution, I measured GFAP transcript, confirming (Fox et al., 2004) a persistently low expression from embryonic stage DIV0 (E18.5) to DIV4 and a robust increase detectable not sooner than DIV8 (supplemental Fig. 5*B*, available at www.jneurosci.org as supplemental material from Zibetti et al.2010) that is after the inversion of LSD1 isoforms proportion has occurred.

Again, I can reliably assume that the increase of E8a inclusion in the cellular model is attributable to the neuronal differentiation, with no regard to glial composition. To infer the function of LSD1 isoforms within neurons, I knocked them down differentially by generating short hairpin RNAs (shRNAs) specific for either neurospecific exon E8a or the splice junction between exon E8 and E9, which is shared among ubiquitous LSD1 isoforms. Hairpins containing vectors were generated and tested for isoform specificity and efficacy (supplemental Fig. 6, available at www.jneurosci.org as supplemental material). I transfected cortical neurons at DIV4 with hairpins containing vectors and analyzed at DIV8 (Fig. 7*A*–*C*) and evaluated phenotypic traits that describe neuronal morphogenesis during *in vitro* maturation, including cumulative neurite length, the number of branches, and neurite width by Sholl's analysis performed on

increasing radii from centered soma (Fig. 7*D–F*) (Sholl et al. 1953, Gutierrez et al.2007). The knockdown of ubiquitous isoforms ensued little or no effect when compared with control according to cumulative neurite arborisation, branch count, and neurite width. Conversely, the silencing of neurospecific isoforms altered neurite morphogenesis by eliciting a significant decrease of the cumulative neurite arborisation (Fig. 7*D*), a reduced number of secondary branches (Fig. 7*E*), and a reduced average neurite width (Fig. 7*F*). [The following series are scrambled compared with shRNAvs8a or shRNAvs8/9 conditions with related *p* values: neurite length (in μm) in Figure 7*D*, 897.21 ± 225.9, 664.42 ± 170.9, p=4.99e-4; 1063.44±229.5, p=0.06; neurite branches in Figure 7*E*, 11.08 ± 1.54, 8.88 ± 1.59, p = 0.048; 11.16 ±2.13, p = 0.95; neurite width (in μm)in Figure 7*F* at inner radius, 1.42±0.12, 0.95 ±0.078, p = 4.95e-9; 1.29 ±0.14, p = 0.16; intermediate radius, 1.27 ±0.11, 0.85 ± 0.074, p = 7.21e-9; 1.12 ±0.12, p = 0.08; outer radius, 1.13 ±0.1, 0.79 ±0.064, p = 1.01e-6; 0.98 ± 0.098, p = 0.05; two tailed *t* test.]

Furthermore, I evaluated neurite morphogenesis after overexpressing LSD1 neurospecific isoforms, since this was expected to exert opposite effects to those observed during knockdown experiments. Neurons were transfected at DIV4 with enhanced GFP (EGFP) vector alone (mock condition) (Fig. 8*A*) and co-transfected with HA-tagged LSD1 isoforms, both neuronal (pCGN–LSD1-8a plus pCGN–LSD1-2a/8a indicated as nLSD1 condition in Fig. 8*B*) and ubiquitous ones (pCGN–LSD1 plus pCGN–LSD1-2a indicated as uLSD1 condition in Fig. 8*C*). As expected, overexpression of neurospecific isoforms induced an increase in the morphometric parameters compared with controls (Fig. 8*D–F*), whereas overexpression of ubiquitous isoforms did not result in any significant effect. [The following series are mock compared with nLSD1 or uLSD1 conditions with related *p* values: neurite length (in μm) in Figure 8*D*, 783.7 ±116.9, 943.8 ±132.1, p = 0.03; 799.9 ±101.5, p = 0.41; neurite branches in Figure 8*E*, 17.24 ± 5.67, 23.92 ± 4.64, p = 0.03; 12.56 ± 2.63, p = 0.07; neurite width (in μm) in Figure 8*F* at inner radius, 0.835 ± 0.09, 1.239 ± 0.11, p = 2.12e-8; 0.855 ± 0.07, p=0.36; intermediate radius, 0.7339±0.09, 1.1313±0.11,p=2.27e-8; 0.7419±0.06, p=0.44; outer radius, 0.5995±0.08, 0.8876 ±0.07, p = 2.01e-7; 0.6225 ±0.05, p = 0.32; one-tailed *t* test]. To discern the relative contribution of each E8a-containing isoform in mediating the morphogenic effect on neurons, I performed a parallel experiment, transfecting either LSD1-8a or LSD1-2a/8a. As shown in supplemental Figure 7 (available at www.jneurosci.org as supplemental material from Zibetti et al. 2010), LSD1-8a is responsible for the morphogenic effect with LSD1-2a/8a partially

recapitulating the phenotype that was observed under the nLSD1 condition (Fig. 8) in which both isoforms are present.

In conclusion, while the knockdown of neurospecific LSD1 isoforms delays neurite morphogenesis, overexpression of the same seems to anticipate features that normally arise at later stages under physiological conditions. Conversely, perturbation of ubiquitous LSD1 variants, either by knock down or transfection, sorted no statistically significant effect when compared with controls. Because experimental conditions only differed for E8a retention in both experiments, the morphogenic effects likely rely on exon E8a presence.

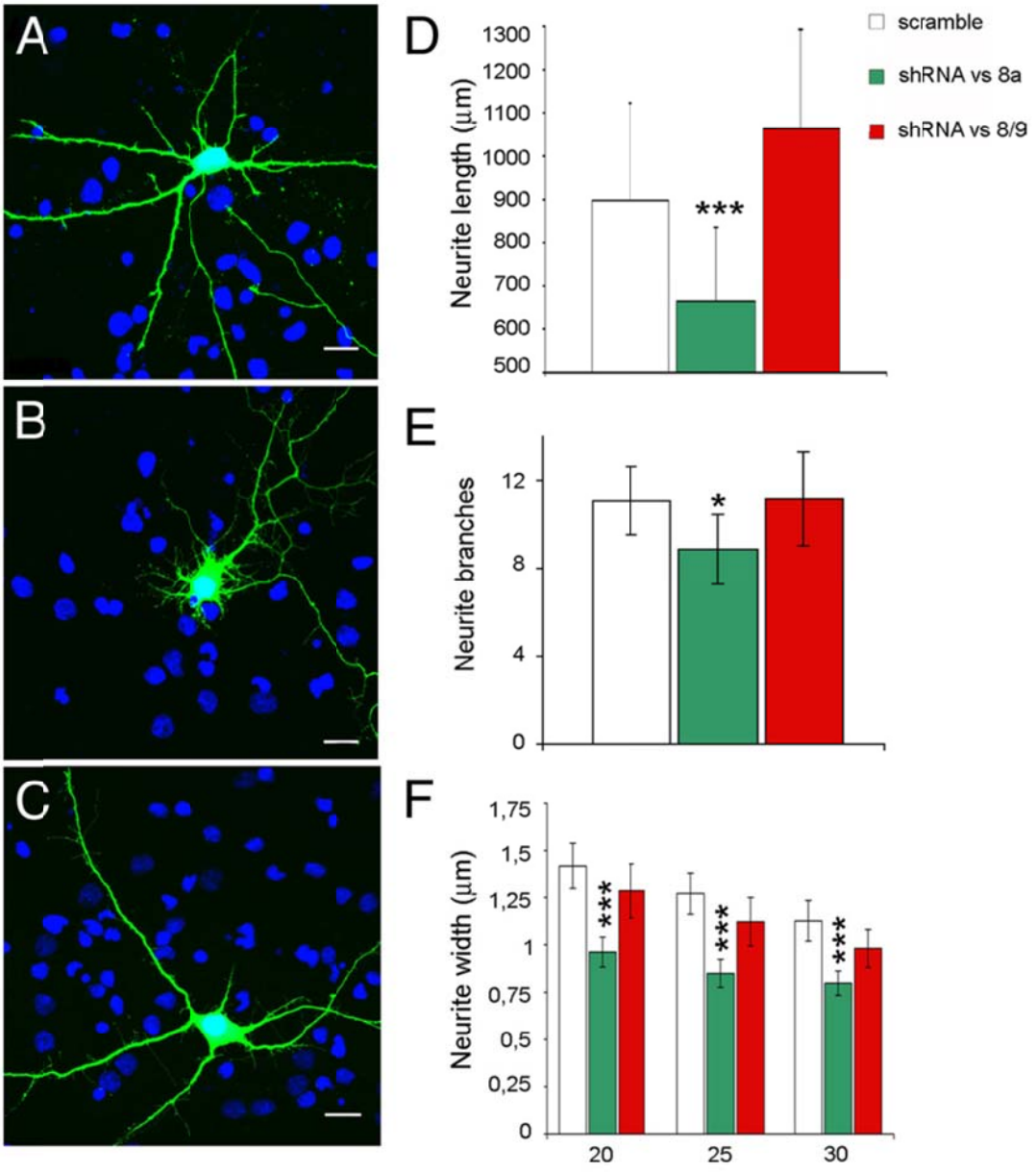


Figure 7. Effect of neurospecific or ubiquitous LSD1 knockdown by shRNAs on neurite morphology in rat cortical neurons. Cultured cortical neurons were transiently transfected with pSuper GFP Neo control (scramble) (***A***), pSuper engineered with shRNA against exon E8a (shRNA vs 8a) (***B***), and pSuper engineered with shRNA against the splice junction between exons E8 and E9 (shRNA vs 8/9) (***C***). Morphology was analyzed for EGFP-positive neurons with DAPI (4,6-diamidino-2-phenylindole) counterstain. ***D***, Cumulative neurite length in differentially LSD1 knocked down neurons is indicated as average±SEM 0.95 C.I. in micrometers. ***E***, Secondary branches count is indicated as average ± 0.95 C.I.. ***F***, Average neurite width by Sholl analysis calculated on inner, intermediate, and outer Sholl's circles corresponding to 20, 25, and 30µm radii, respectively. Values shown are mean±SEM 0.95 C.I. width in micrometers. Student's *t* test (*l* Stat *t I*> *T*α/2) was applied to values by comparing each condition with control scramble. **p*<0.05; ***p*<0.01; ****p*<0.001. Scale bars, 20µm. Adapted from Zibetti et al. J. Neurosci., February 17, 2010 • 30(7):2521–2532

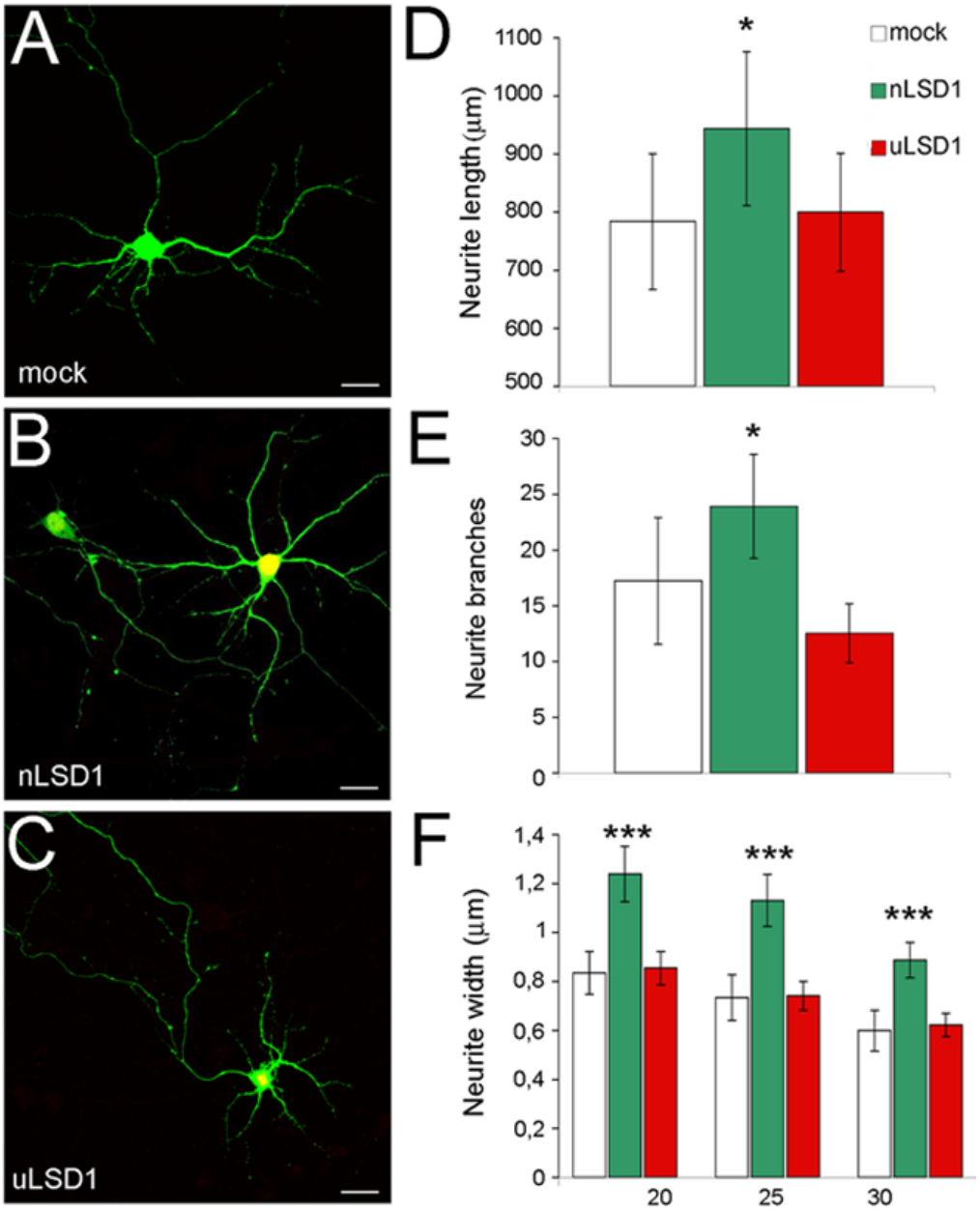


Figure 8. Effect of overexpression of neurospecific or ubiquitous LSD1 isoforms on neurite morphology in rat cortical neurons. Primary rat cortical neurons were transiently cotransfected with pCGN vector (mock) (***A***), EGFP together with HA–LSD1-8a plus LSD1- 2a/8a (nLSD1) (***B***), HA–LSD1 plus LSD1-2a (uLSD1) (***C***). Morphology was analyzed in EGFP-positive neurons (mock) or double-labeled EGFP- and HA-positive neurons. ***D***, Cumulative neurite length in differentially LSD1 transfected neurons is indicated as mean ± 0.95 C.I. in micrometers. ***E***, Secondary branches count is indicated ±SEM 0.95 C.I.. ***F***, Average neurite width by Sholl analysis calculated on inner, intermediate, and outer Sholl's circles corresponding to 20, 25, and 30 µm radii, respectively. Values shown are mean ± 0.95 C.I. width in micrometers. Student's *t* test (*l* Stat *t I>Tα*) was applied to values by comparing each condition with mock. **p*<0.05; ***p*<0.01; ****p*<0.001. Scale bars, 20µm. Adapted from Zibetti et al. J. Neurosci., February 17, 2010 • 30(7):2521–2532

Computational analysis of putative LSD1 phosphorylation sites reveals a high score threonine residue within the neuro-specific exon which *in vivo* is susceptible of post-translational modification

Since post-translational modifications occurring on histone modifying proteins have been reported to exert a powerful effect in terms of catalytic activity (Lemercier et al. 2003, Nott et al. 2008,), subcellular localization (Cai et al.2010), resulting in a change in the epigenetic functions, I scanned the 876 aa long LSD1 variant through NetPhos v 2.0 software predicting all putative phosphorylation sites (http://www.cbs.dtu.dk/services/NetPhos). As result, 14 Tyrosine, 22 Serine and 14 Threonine residues were found, among which a high score phosphorylated Thr residue predicted on 391 aminoacidic residue, falling within the neuro-specific exon DTVK (Fig. 9A and B).

NetPhos v2.0 - Threonine predictions 876 aa

Name	Pos	Context	Score	Pred
				v
Sequence	19	AAAATGTEA	0.561	*T*
Sequence	21	AATGTEAGP	0.233	.
Sequence	27	AGPGTAGGS	0.140	.
Sequence	59	VGERTPRKK	0.990	*T*
Sequence	88	QAGPTVVPG	0.188	.
Sequence	95	PGSATPMET	0.979	*T*
Sequence	99	TPMETGIAE	0.075	.
Sequence	104	GIAETPEGR	0.745	*T*
Sequence	110	EGRRTSRRK	0.989	*T*
Sequence	209	HDRMTSQEA	0.955	*T*
Sequence	226	GPQQTQKVF	0.986	*T*
Sequence	237	IRNRTLQLW	0.035	.
Sequence	250	KIQLTFEAT	0.126	.
Sequence	254	TFEATLQQL	0.063	.
Sequence	266	YNSDTVLVH	0.051	.
Sequence	298	PIKKTGKVI	0.334	.
Sequence	325	GMDVTLLEA	0.387	.
Sequence	339	GRVATFRKG	0.946	*T*
Sequence	355	AMVVTGLGG	0.027	.
Sequence	391	GQADTVKVP	0.934	*T*
Sequence	413	LLEATSYLS	0.155	.
Sequence	461	KIVKTQEEL	0.642	*T*
Sequence	499	PRDITAEFL	0.123	.
Sequence	512	HRDLTALCK	0.266	.
Sequence	524	ELAETQGKL	0.214	.
Sequence	566	FANATPLST	0.316	.
Sequence	570	TPLSTLSLK	0.040	.
Sequence	585	DFEFTGSHL	0.257	.
Sequence	590	GSHLTVRNG	0.809	*T*
Sequence	612	IKLNTAVRQ	0.055	.
Sequence	620	QVRYTASGC	0.094	.
Sequence	631	IAVNTRSTS	0.169	.
Sequence	634	NTRSTSQTF	0.580	*T*
Sequence	637	STSQTFIYK	0.022	.
Sequence	648	AVLCTLPLG	0.006	.
Sequence	672	PEWKTSAVQ	0.208	.
Sequence	708	HVGSTTASR	0.088	.
Sequence	709	VGSTTASRG	0.164	.
Sequence	770	QPKETVVSR	0.237	.
Sequence	805	AQPITPGPS	0.280	.
Sequence	827	AGEHTIRNY	0.682	*T*
Sequence	834	NYPATVHGA	0.306	.
Sequence	859	GAMYTLPRQ	0.029	.
Sequence	865	PRQATPGVP	0.980	*T*

Figure 9a. Prediction of sites susceptible of post-translational modifications was computed on NetPhos v.2.0 software. Predicted threonine sites with the related probability scores are indicated from the 876 AA long nsLSD1 isoform

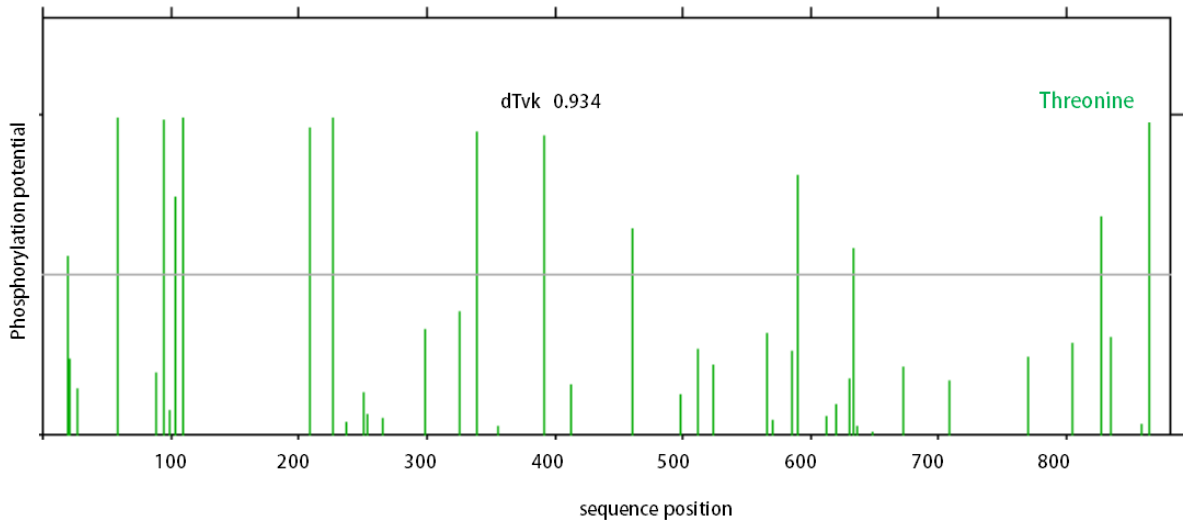


Figure 9b. Probability plot of phosphorylation threonine sites spanning through the 876 AA long nsLSD1 variant (NetPhos v2.0)

Given the high score prediction, I immunoprecipitated LSD1 complex from rat cerebral tissue at PN1, when nLSD1 protein had proven to be highly expressed (see Fig. 6E), separated fractions by SDS-PAGE and processed samples (see Fig. 10A and methods) for mass spectrometry analysis. From ionization of tryptic peptides (Arg-Lys) high intensity (m/z) molecular ions were retrieved with a wide sequence coverage attribuable to LSD1. Molecular ions obtained by peptide mass fingerprinting correspond to LSD1 peptides with and without exon E2a, as well as the neuro-specific exon E8a. In particular three C-Me-cys modified peptides deriving from tryptic miscleavage were identified: a 14 aa peptide (CPLYEANGQADTVK), a 17 aa one (CPLYEANGQADTVKPK) and a 19 aa one (CPLYEANGQADTVKPKEK) with a m/z ratio of 1725.022, 1874 e 226,460 respectively. The *base peak* related to the 19 aa long peptide exhibits an addictive *m/Z* value of 80 Da more than the predicted nominal mass, attribuable to the presence of a single phosphate group; moreover *m/Z* ratio associated to the 17 aa long peptide is attribuable to the loss of a H20 molecule from Ser(Y) or Thr(T), which relates to the presence of a phosphate modification (CPLYEANGQADTVKPK). Ionization of the 14 aa long tryptic peptide generated a high intensity *base peak* of 1725.022 *m/Z* (Fig.10B), whose further fragmentation produced a 1565,996 *m/Z* peak attributable to the unmodified exon e8a, where the difference of 160 Da can be explained with two phosphate groups (CPL**Y**EANGQAD**T**VKPKEK), confirming the hypothesis that the DTVK aminoacidic stretch coded by exon 8a is susceptible of phosphorylation *in vivo*.

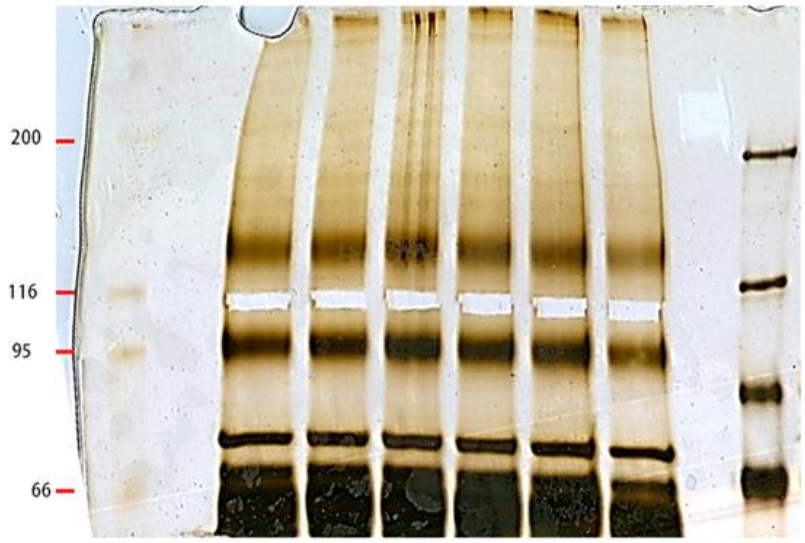


Figure 10A. High sensitivity MS-compatible acidic silver staining of LSD1 immunoprecipitated complex from PN1 rat brains

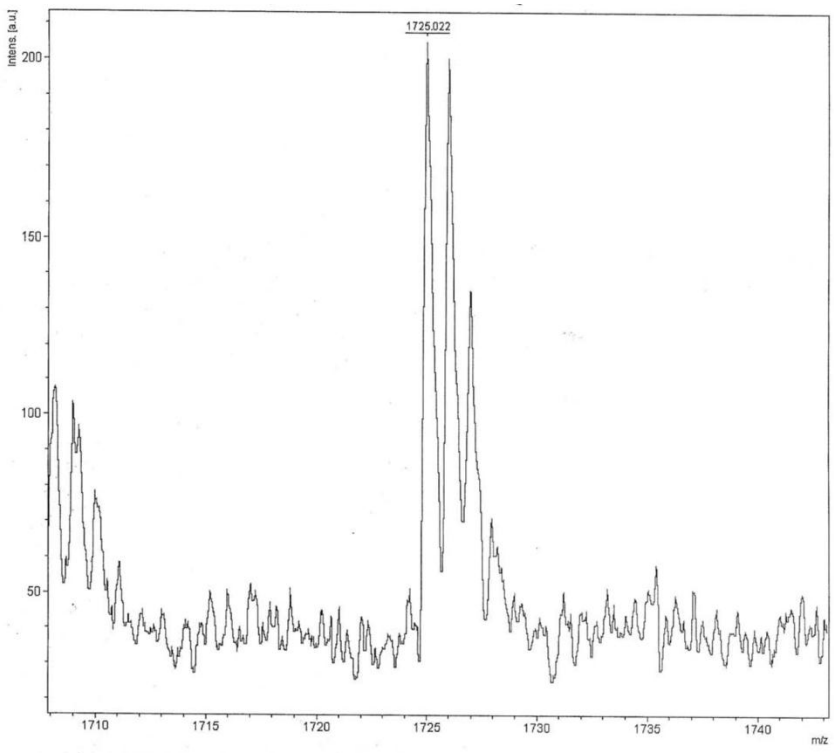


Figure 10b. Mass spectrum of volatilized tryptic (Arg-Lys) peptides from 876 aa LSD1: here displayed molecular ions from fragmentation of peptide 360-373 : *base peak* of 1725.022 *m/Z* (CPLYEANGQADTVKPKEK) related to a double phosphorylation status and lower intensity peaks related to a single phosphorylation status (DTVK) and unmodified exon E8a containing peptide of 1565,996 *m/Z*

Exon E8a phoshorylation relieves nLSD1 from regulatory activity

To investigate whether nLSD1 regulatory activity on genes could be preferentially attributed to its post-translational modified or unmodified status, I generated vectors mimicking hyper phosphorylation (T371D) or hypo phosphorylation (T371A) of the threonine residue by single aminoacidic substitution and performed luciferase reporter assay on rat cortical neurons. As previously observed, luciferase reporter gene exhibits a higher activity under the influence of the nLSD1 isoform compared to the wtLSD1 repressor (Fig.11A, 65.44 ± 9.03 vs 40.41 ± 6.66, *p-value* < 0.001, two tailed t-test). Moreover, hyper phosphorylation of the nLSD1 variant further relieves reporter gene from repression (T371D: 78.16±10.42 vs 40.41 ± 6.66, *p-value* < 0.001) whereas hypo phosphorylation of the same fully rescues wtLSD1 isoform regulatory activity (T371A: 40.48 ± 9.98 vs 40.41 ± 6.66, *p-value* > 0.05), indicating in the phosphorylation of exon E8a a mechanistic basis for functional diversification of LSD1 splice variants in the CNS.

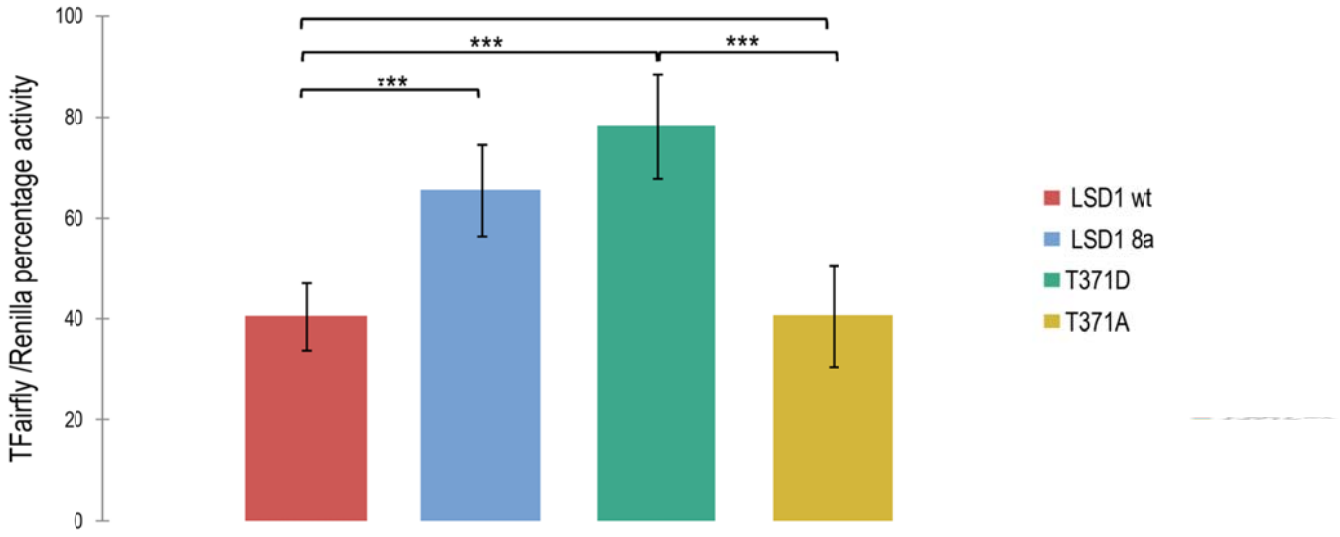


Figure 11a. Luciferase reporter expression upon induction of LSD1 neurospecific isoforms in either phosphorylated(T371D) or unphosphorylated (T371A) forms: indicated values are average ± 0.95*C.I. Averaged Luk activity was expressed as percentage relative to mock condition(not shown as the Gal4DBD vector), upon Renilla normalizarion. P-values (|t|>=Tα/2) for two tailed T-Student are indicated: LSD1 wt vs nLSD1 8a condition < 0.001, LSD1 wt vs nLSD1 T371D condition < 0.001, nLSD1 T371D vs nLSD1 T371A condition < 0.001, LSD1 wt vs nLSD1T371A condition > 0.05

Notably, no statistical significance between native nLSD1 condition and the hyper phosphorylated one can be detected (Fig.11 b, percentage values refer to differences between reporter activity by default and under treatment: nLSD1 34.5 ± 9.03 vs T371D 21.84 ± 10.42, *p-value* > 0.05), whereas its regulatory role is enhanced when no phosphorylation occurs (nLSD1 34.55 ± 9.03 vs T371A 59.52 ± 9.98, p –value < 0.001).

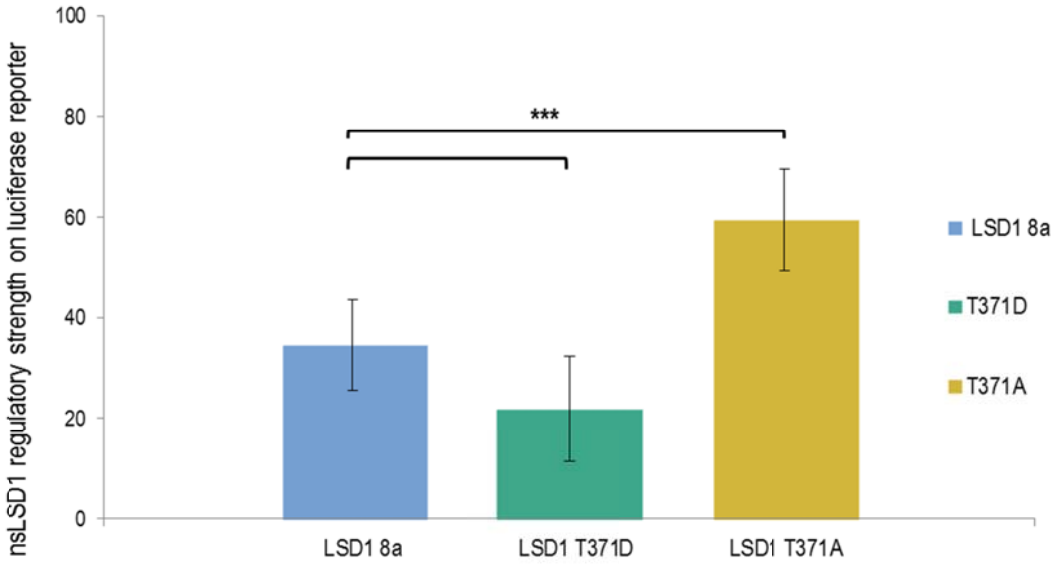


Figure 11b. nLSD1 regulatory strength over luciferase reporter. Indicated percentage values refer to the differences between reporter activity by default (100%) and under treatment (Fig.11a)

To evaluate any morphological differences in cortical neurons arising from preferential expression of the phosphorylated rather than the unphosphorylated nLSD1, I transfected mutagenized vectors T371D or T371A at DIV4 and analyzed morphology at DIV8. Once more, cumulative neurite arborisation and branches were evaluated: again, no statistical difference could be retrieved between native nLSD1 transfection and phosphorylated one on cumulative neurite arborisation, although a trend towards reduction could be observed in a few cases (Fig.12D Two-tailed eteroschedastic Student T test: nLSD1 vs nLSD1mutT371D 976,75 +/- 117,2 µm, 894,17+/- 99,4 µm, p-value > 0,05), whereas unphosphorylated counterpart enhances neurite sprouting (Fig.12D nLSD1 vs nLSD1 mut T371A: 976,75 +/- 117,2 µm, 1154,35+/- 113 µm, p-value = 0.0008), suggesting that most of the native neuro-specific isoform may undergo phosphorylation in vivo and the extent at which E8a modification occurs would determine the

final regulatory outcome. Moreover, differently from what I could observe on the reporter assay performed on a constitutively active general promoter, none of the phosphorylated counterparts mimic wtLSD1 effect that was observed previously (Fig.8) since neurite outgrowth is either statistically unaltered (T371D) or increased (T371A) relative to native nLSD1 induced condition, further suggesting that genes regulated by nLSD1 may differ indeed from those regulated by the ubiquitous isoforms. Similarly, no difference in neurite branching can be detected between native nsLSD1 and the phosphorylated isoforms, whereas a significant increase can be attributed to the hypo phosphorylated counterpart(Fig.12E nLSD1, nLSD1 mut T371D and nLSD1 mut T371A: 18,56 +/- 1,76 vs 18,97 +/- 2,06, 22,1 +/-1,92 respectively).

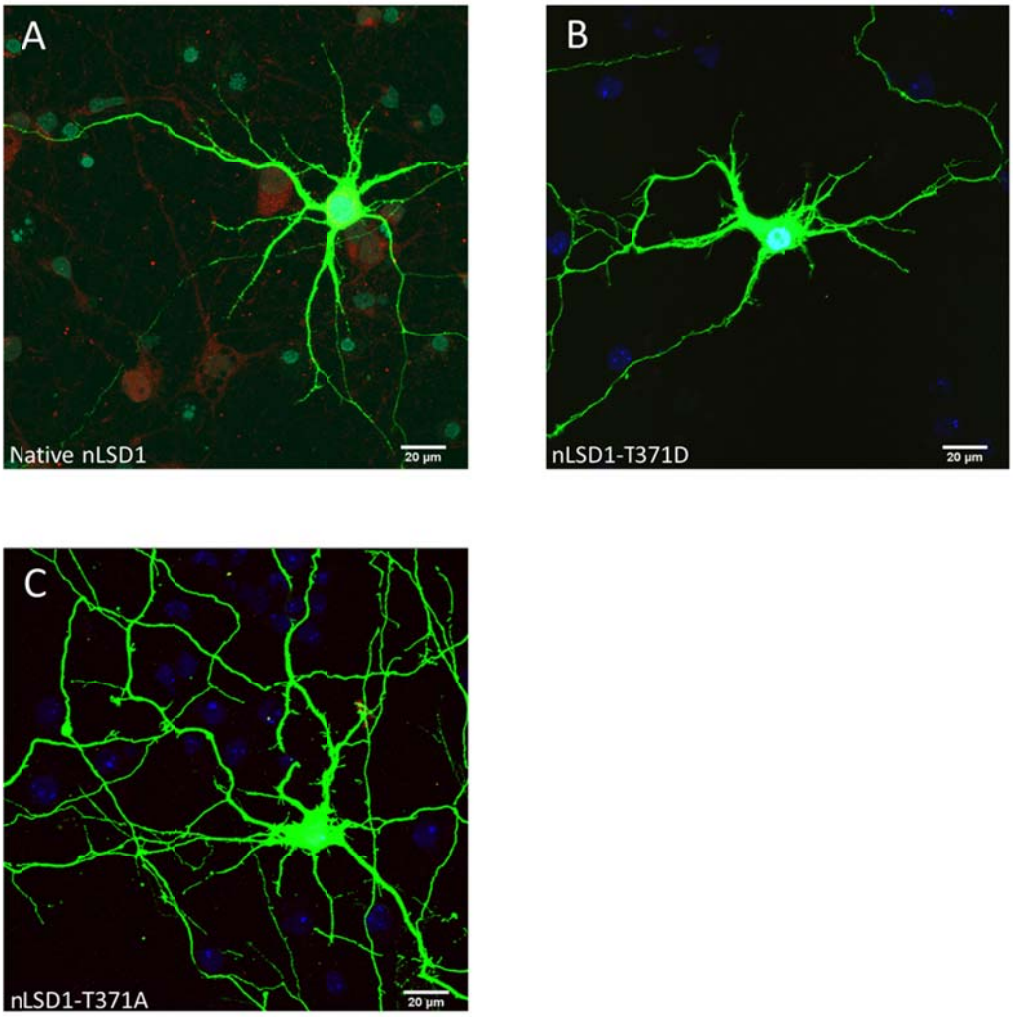


Figure 12a,b,c. Immunostaining of rat cortical neurons transfected at div4 with vectors mimicking hyper phosphorylation or hypo phosphorylation of nsLSD1 at exon E8a and morphology was analyzed at div 7

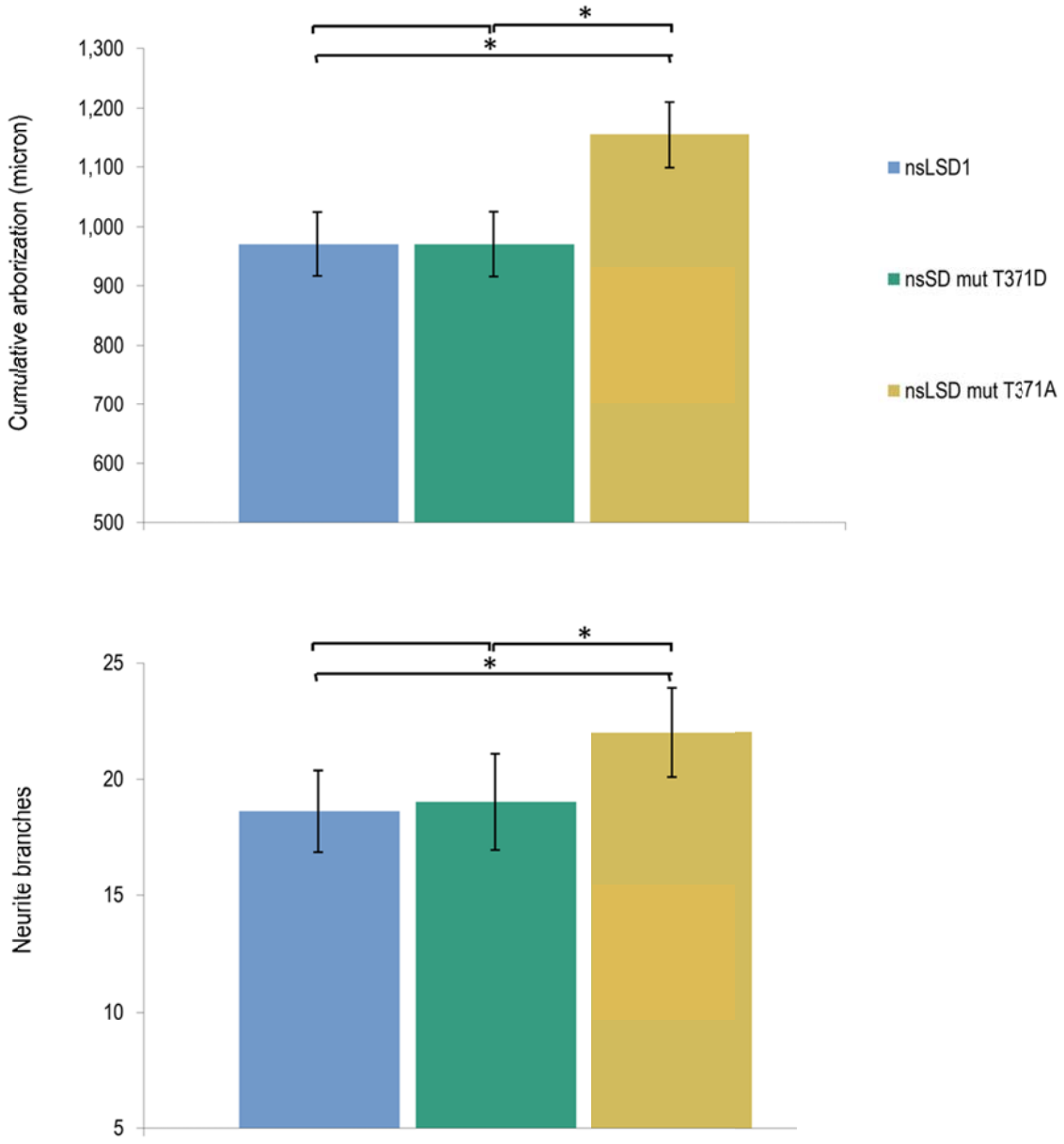


Figure 12d,e. Cumulative neurite arborisation(d), neurite branches (e) in cortical neurons under preferential expression of phosphorylated or unphosphorylated nLSD1 isoform. Indicated bars are average ± s.e.m. Here indicated as s.e.m.*0.95 confidence intervals. c) Cumulative arborisation: 970,91 µm +/- 110,7 µm for nLSD1, 970,57 +/- 111,9 for mutagenized nLSD1-T371D and 1154,35 +/- 113 for nLSD1-T371A . Two tailed eteroschedastic Student T-test p-values are indicated : 0.021 nLSD1 vs nLSD1-T371A and 0.021 nLSD1-T371D vs nLSD1- T371A d) Neurite total branches 18,56 +/- 1,76 for nsLSD1 8a 18,97 +/- 2,06 for mutagenized T371D exon E8a, 22,1 +/-1,92 for T371A exon E8a. Two tailed eteroschedastic Student T-test p-values are indicated: 0,01 nsLSD1 vs nLSD1-T371A and 0.03 nLSD1-T371D vs nLSD1- T371A

LSD1 genome wide location in cortical neurons by high throughput ChIP-sequencing

To further characterize the role of LSD1 splice variants in the regulation of gene expression in the nervous system, I performed chromatin immunoprecipitation followed by high throughput sequencing on rat cortical neurons (experiment performed at Rosenfeld laboratory, UCSD, School of Medicine, San Diego, CA). ChIP-seq was done at e18.5+div7, when the neurospecific variant proved to be highly expressed (see Fig.6) by means of an antibody recognizing all the four isoforms (Fig. 10), providing the overall LSD1 genome wide location in the CNS, with a major contribution from the neuro-specific LSD1 isoform, according to the *in vitro* developmental stage.

I pursued a subtractive approach to infer gene sets specifically related to the LSD1 neurospecific isoforms, comparing a mock sample where all the four variants are expressed with a knocked sample where either the neurospecific or the ubiquitous variants are selectively silenced by means of lentiviral- mediated shRNAs delivery. Even though all the ChIP fragments are expected to be called over threshold (non stochastic frequency of sequenced tags, *Poisson* distribution), those obtained under the two experimental conditions should differ for the score over *false discovery rate* since the enrichment of the related tags depends on the contributing LSD1 isoform being immunoprecipitated: RNA interference of a specific LSD1 isoform would determine a depletion (lower score over *fdr*) of the corresponding fragments rather than a complete loss (no significant enrichment over local background), depending on the efficacy of the silencing. Conversely, *peaks calling* for the cognate LSD1 isoforms wouldn't be affected, since their expression would be unaltered relative to the mock condition.

While further optimization is still required for lentiviral- mediated LSD1 silencing, ChIP-sequencing in control conditions led so far to an extensive characterization of genes that are complexively regulated by all LSD1 splice variants expressed in developing cortical neurons at DIV7, providing 11 million reads, 94% of which uniquely mappable to genomic positions and 10% of which annotated to promoters residing at less than 1000 bp from a referenced TSS.

Computational analysis of high throughput Chip sequenced reads: genome indexing, reads alignment and tags density

The obtained 35-bp long ChIP –seq reads were aligned onto the indexed reference genome (Rat Nov. 2004-Baylor 3.4/rn4 assembly) comparing results from two open source multiple sequence alignment tools, differing for matching tolerances: bowtie (http;//bowtie-bio.sourceforge.net/index.html) (Langmead *et al.*, 2009), and GASSST (global alignment short sequence search tool, http://www.irisa.fr/symbiose/projects/gassst/html) (Rizk et al. 2010) (see methods). UCSC genome browser (http://genome.ucsc.edu/) was configured to display custom tracks from bowtie and GASSST aligned tags onto the Rat Nov.2004 Baylor 3.4 rn4 assembly and to visualize the related ChIP-fragment densities, as the total number of overlapping tags at each position in the genome.

Quality Controls and Parameter Estimation

The minimum amount of immunoprecipitated material was guaranteed, as verified by the clonal tag counts (<1.2 tags per position) (*tag count* by HOMER at http://biowhat.ucsd.edu/homer/chipseq/index.html): the DNA from sonicated Chip sample should fragment in a random pattern, making relatively rare that fragments start in exactly the same location in the genome, which is unavoidable for regions with high ChIP enrichment, though if most fragments in the sample have several reads in the same positions, this is indicative of clonal reads since a limiting amount of starting material will likely lead to over sequencing of the same fragments, generating high clonal tag counts. Bowtie alignment of LSD1 chip reads provided 1.061107 average tags per position, comparable to GASSST aligned average tags (1.063926) (Fig.13).Moreover, since the estimation of the fragment length affects the prediction of the binding site, an autocorrelation analysis of the sequenced tags position was performed. The estimates of the fragment length and peak width were computed (*tag autocorrelation tool*, HOMER) comparing results obtained from bowtie and GASSST (Fig.14)

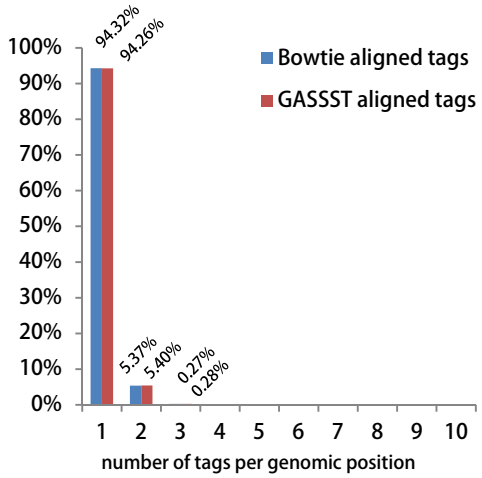


Figure 13. Tags count plot: the total number of tags per genomic position was calculated to check for clonality of sequenced reads.

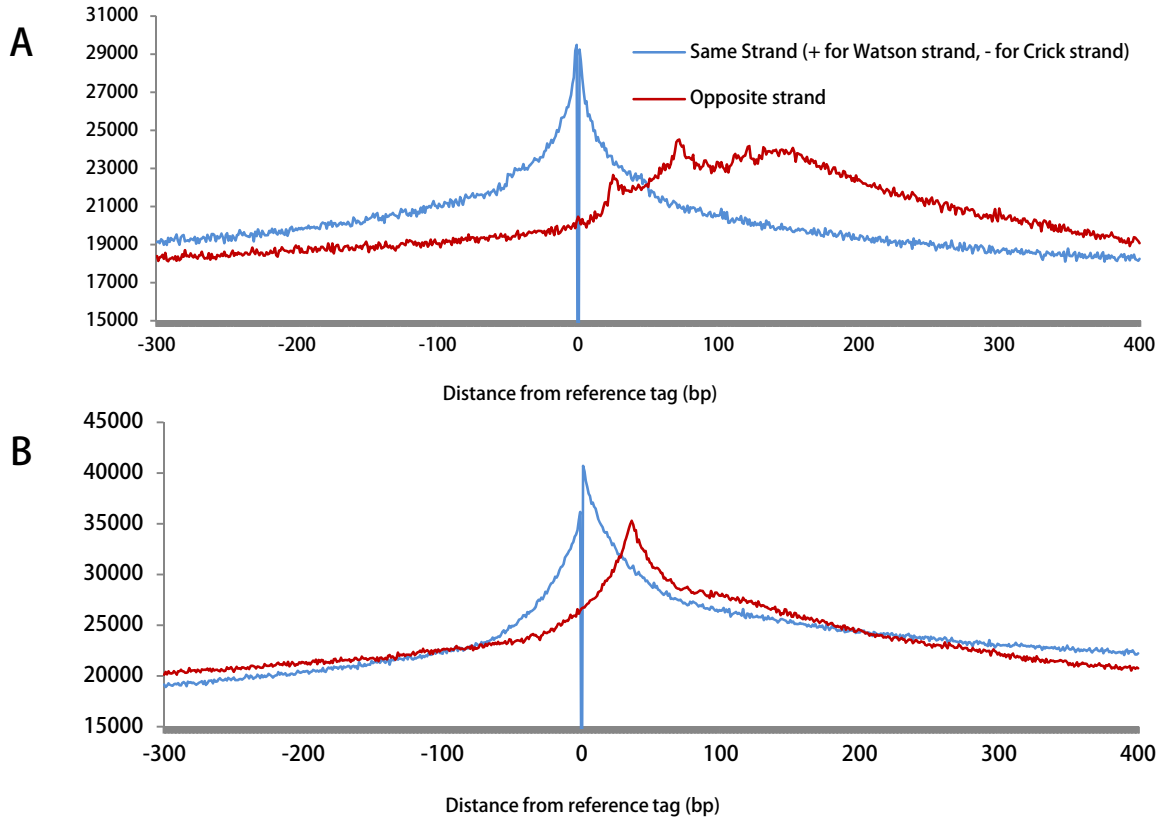


Figure 14a,b. Chip Seq Tags autocorrelation plots. Distances separating each tag from the subsequent one along the same chromosome were plotted on histogram referring to the same strand or the opposite. The estimate of the sequencing fragment length is derived from the maximum in autocorrelation signal on the opposite strand downstream of the reference tag. The level of opposite strand found at the reference tag provides an estimate of the background of the ChIP seq signal, since tags found in opposite strands facing away from each other cannot measure the same protein bound to the DNA. From the background level peak width, which is the relative range in fragment size, is also retrieved.

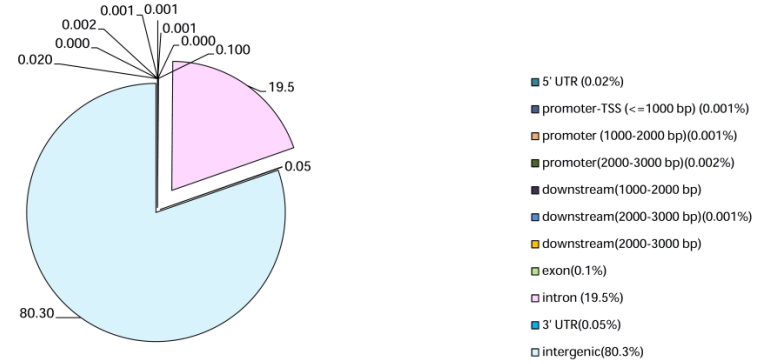
Peaks calling and annotation

High throughput sequencing of LSD1 sample through Solexa platform (Illumina) generated 12.416.494 total reads resulting in 11.700.737 bowtie identified tags (unique mappable positions) versus 11.101.216 unique position detected by GASSST. Fragment length estimate from bowtie aligned tags is 76 bp with a peak width of 268 bp, whereas GASSST provided a fragment length estimate of 150 bp and an peak width estimate of 150 bp (*tag info*, HOMER Fig.14).

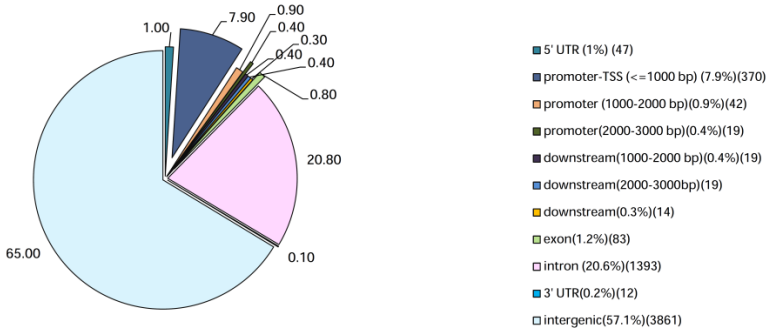
Two different peaks calling methods, including Homer (Fig. 15 c and d) (*Hypergeometric Optimization of Motif EnRichment*, http://biowhat.ucsd.edu/homer/) and MACS (Fig.15 b)(*Model-based analysis of Chip-seq*, http://liulab.dfci.harvard.edu/MACS/index.html), were applied to identify Chip enriched regions, hereafter *peaks*, defined after clustered tags observed at a non-stochastic frequency. Key features of MACS are empirical modeling of '*d*' distance between *maxima* on sense and antisense tags, 3' terminal tag shifting by *d/2* to putative protein-DNA interaction site and the use of a dynamic λ local to capture local biases in the genome. Genome is scanned for candidate peaks with significant tag enrichment and those below a user-defined threshold are called a (p-value cutoff set at 10^{-5}) (see methods).

HOMER determines the threshold of tags needed to call a peak significant by assuming that non-enriched ChIP-fragment concentrations are approximated by a stochastic distribution (Poisson distribution, p-value cutoff set at 0.001; see methods for *fdr* threshold). Briefly, tags positions are adjusted of half the estimated ChIP-fragment length in the 3' direction relative to the original position of the tag and genome is scanned applying a fixed peak size, sorting positions with the highest tag density through those with the lowest one assigning putative peaks and masking nearby positions within the minimum distance acceptable for neighboring peaks (see methods). Results from peaks annotation were plotted basing on the proximity to the nearest annotated gene; promoters were called whether located within 1000 bp upstream a referenced TSS(Homer), or up to 3000 bp (CEAS software).

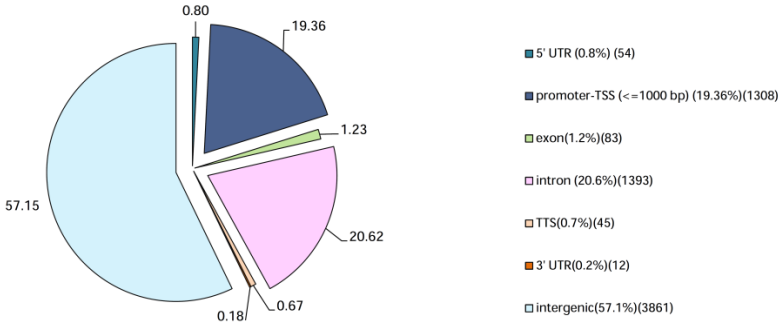
a) Genome (Rat Nov. 2004 (Baylor 3.4/rn4) Assembly



b) ChIP peak annotation on bowtie aligned tags (*n* = 4687), peaks called by MACS



c) ChIP peak annotation on bowtie aligned tags (n = 6756), peaks called by Homer



d) ChIP peak annotation on GASSST aligned tags (*n* = 6702), peaks called by Homer

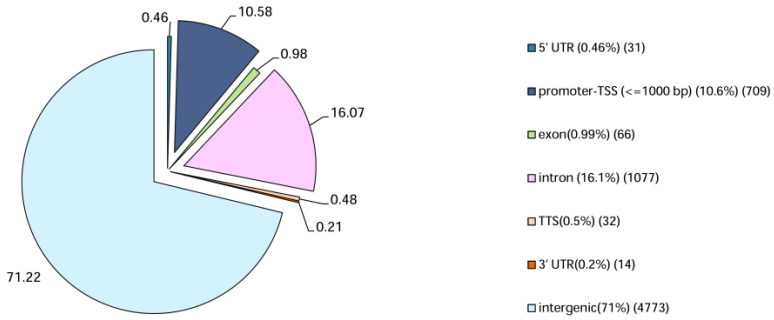
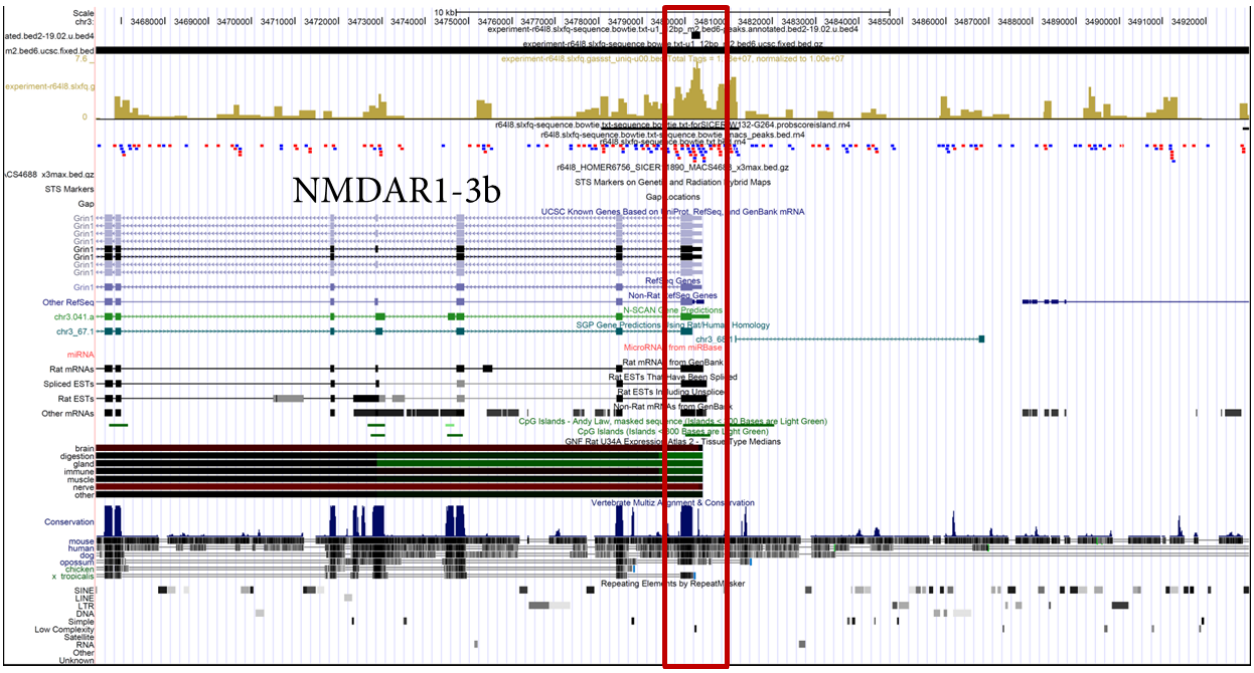


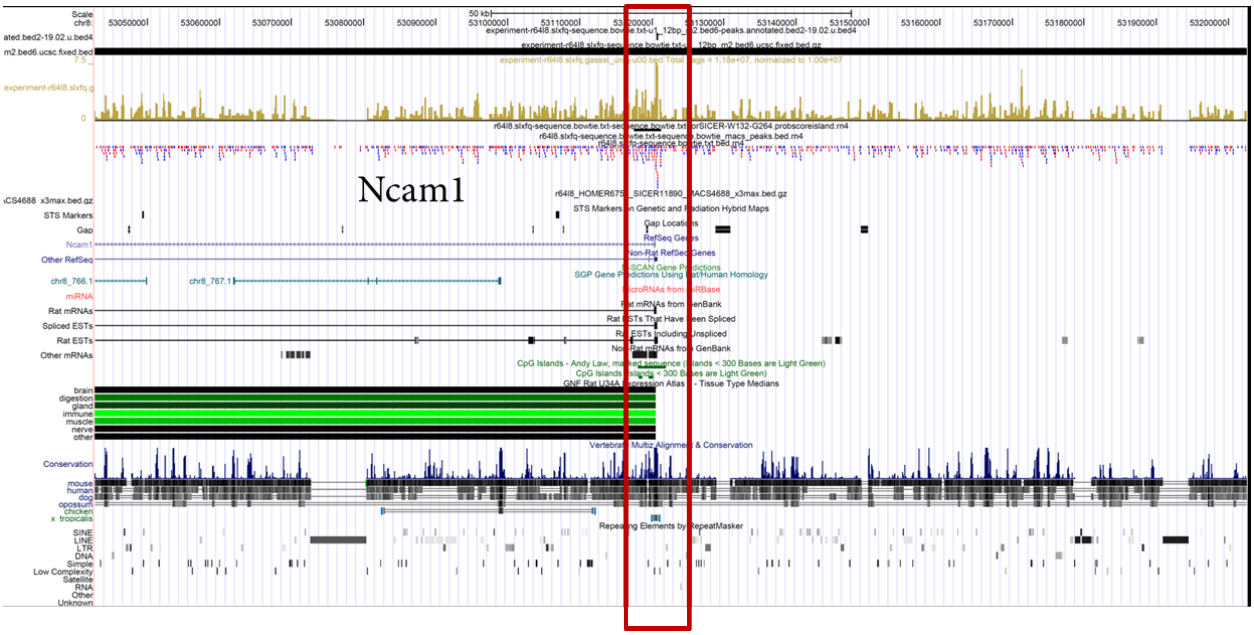
Figure 15 a,b,c,d. a)Genomic distribution of rat reference genome (Rat Nov.2004 Baylor 3.4/rn4 Assembly) over gene categories. Computation was performed by means of CEAS software b) Genomic distribution of LSD1 sv Chip-seq peaks (n=4687) in E18.5+div7 rat cortical neurons. Sequenced tags were aligned onto Ref Genome (Rat Nov.2004 Baylor 3.4 rn4 Assembly) by means of bowtie. 12.416.496 total reads generated 11.700.737 unique mappable positions. Fragment length estimate used for ChIP seq is 76 bp with a peak width of 268 bp. Peaks called by MACS were annotated by means of CEAS software. c)Genomic distribution of LSD1 sv Chip-seq peaks (n=6756) in E18.5+div7 rat cortical neurons. Sequences tags were aligned onto Ref Genome (Rat Nov.2004 Baylor 3.4 rn4 Assembly) by means of bowtie. 12.416.496 total reads generated 11.700.737 unique mappable positions. Fragment length estimate used for ChIP seq is 76 bp with a peak width of 268 bp. Peaks were called and annotated by means of HOMER software. d)Genomic distribution of LSD1 sv Chip-seq peaks (n=6702) in E18.5+div7 rat cortical neurons. Sequences tags were aligned onto Ref Genome (Rat Nov.2004 Baylor 3.4 rn4 Assembly) by means of GASSST. 11.810.875 total reads generated 11.101.216 unique mappable positions. Fragment length estimate used for ChIP seq is 150 bp with a peak width of 150 bp. Peaks were called and annotated by means of HOMER software.

Peaks annotated to promoters were displayed on UCSC genome browser (Fig.16a and b) checking for their proximity to the TSS of a known rat RefSeq gene (-1000-200 bp), for an equal distribution of sense and antisense tags and for the 5'UTR genomic conservation (Multiz Alignment tool). GNF atlas expression and rat ESTs (expressed sequence tags) were also displayed to detect any specific or differential expression of the target gene in the central or peripheral nervous system. Genomic sequence gaps, neighboring CpG islands and repeats were also considered for possible tags alignments bias.

Whenever the nearest reference rat sequence is provisional or not available in UCSC integrated database, and on the basis of evidence from ESTs and N-Scan, it is reasonable to suspect that a miscategorization may have occurred: in this regard, the proximity of the intergenic peak to an orthologous validated gene could possibly allow further refinement of the promoters annotation.

Furthermore, in some cases (Calm3, Ndufb4, Syt6, Tdrd3), multiple LSD1 locations could be retrieved short and long distance from the same TSS, with the distant peaks matching regions genomically conserved and devoid of ESTs (Fig.16b), making improbable that such binding locations may relate to other un-annotated promoters. For such cases, it is feasible to hypothesize a wide-range regulatory role rather than a promoter –restricted function. Therefore I considered at first all the LSD1 annotated gene categories for downstream analysis and opted subsequently for a more conservative approach on the subset of promoters regions.


NMDAR1-3b


Ncam1

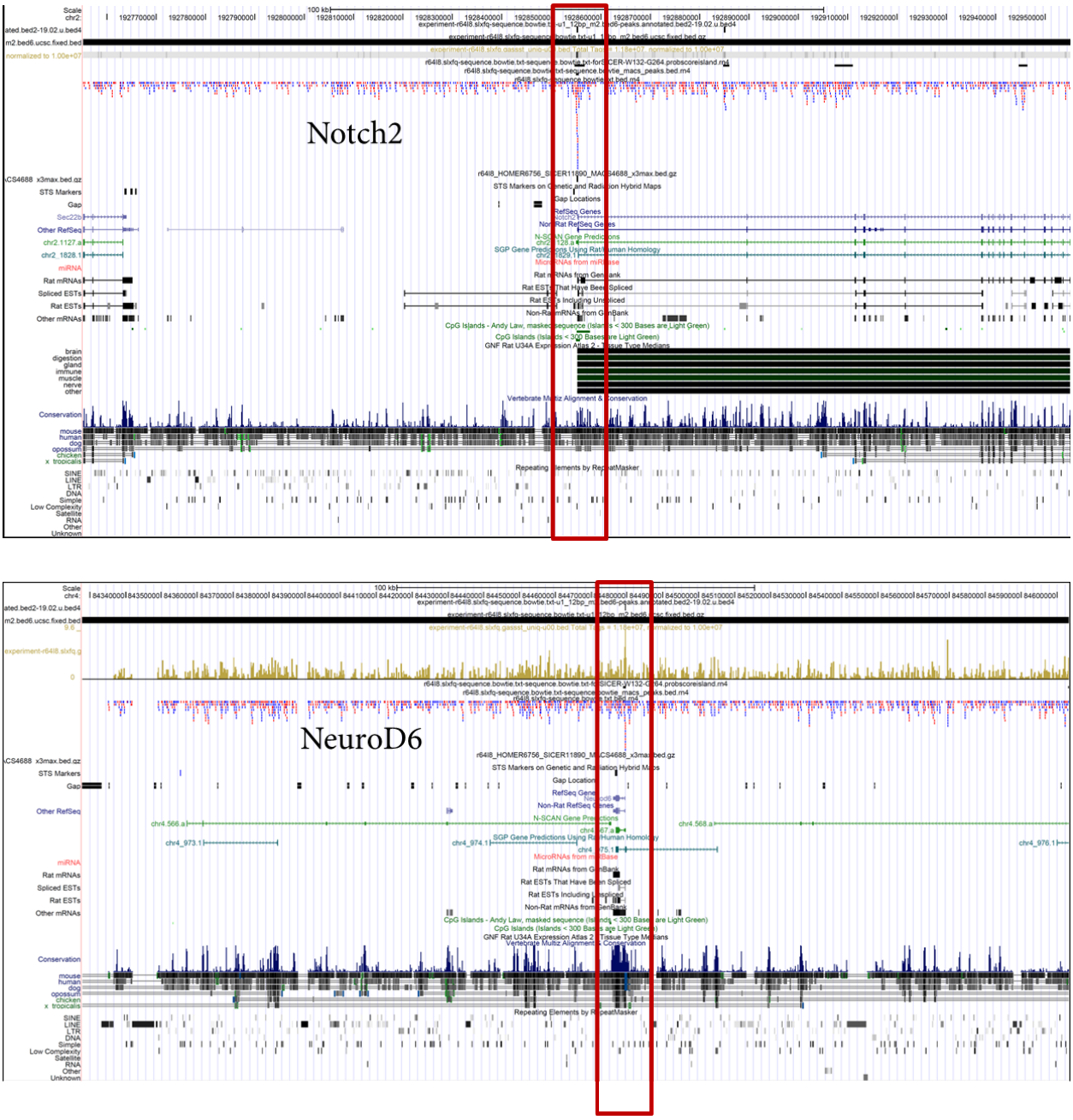
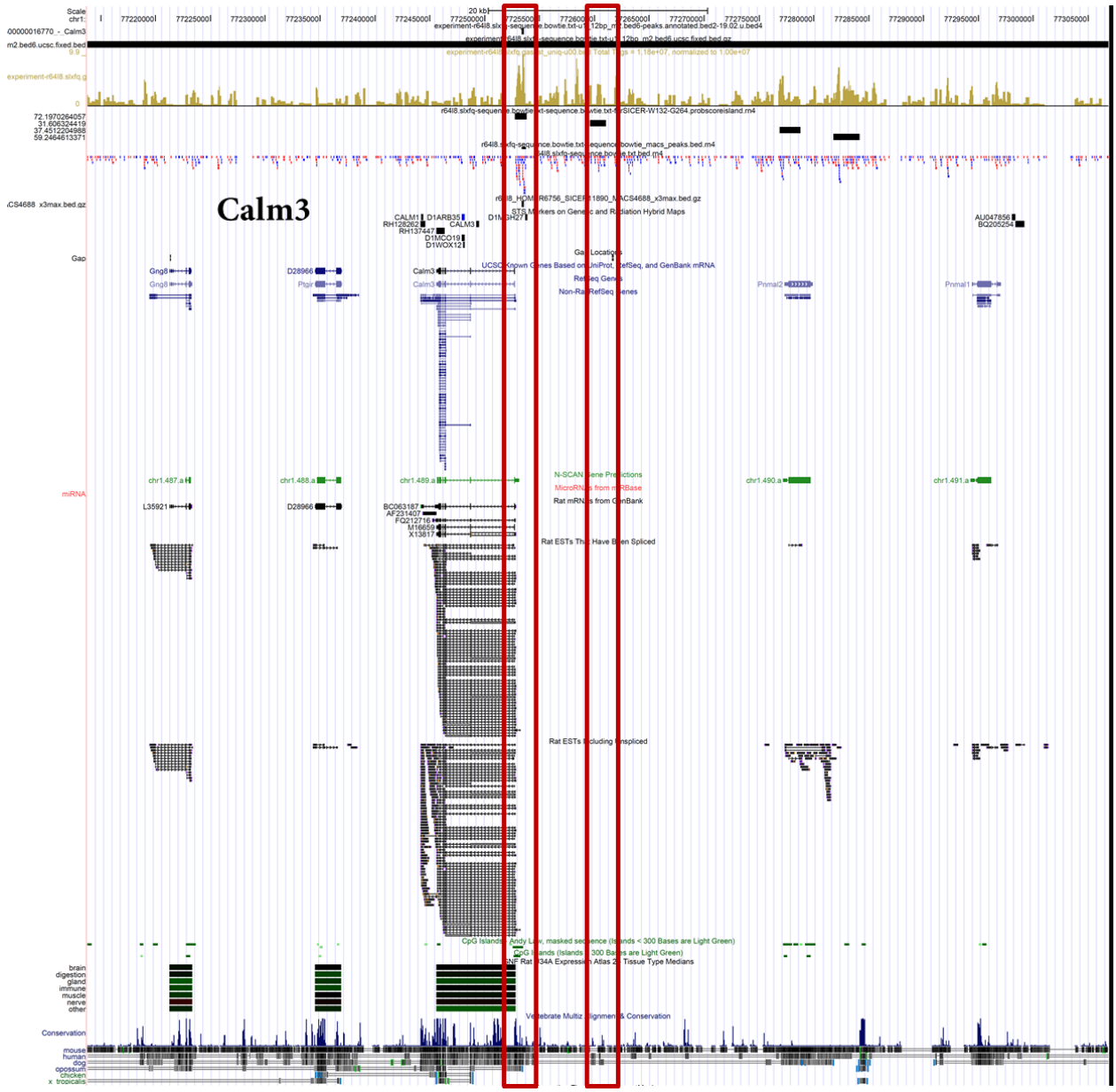


Figure 16a. Examples of custom tracks from LSD1-ChIP seq experiment displayed by means of UCSC genome browser. Windows are focused on clustered tags for which a peak was called by means of Sicer(11890 total peaks), MACS(4687 total peaks) and Homer (6756 total peaks) on either bowtie or GASSST aligned tags. Owing to the difference in peak calling sensitivity images can be displayed at different scales for illustrative purpose. Here indicated four peaks that were called (FDR 0.001, FDR tags threshold 9.0) and annotated to promoters residing at less than 1000 bp from the nearest TSS: NMDAR1-3b(focus ratio 0.6, peak score 10, predicted location 103 bp >5 UTR), Ncam1 (focus ratio 0.8, peak score 10, predicted location 138 bp < TSS), Notch2 (focus ratio 0.89, peak score 21, predicted location 91 bp < TSS) and NeuroD6 (focus ratio 0.7, peak score 10, predicted location 216 bp < TSS).


Calm3
Gng8
D28966
Ptgir
Calm3
Pnmal2
Pnmal1
chr1.487.a
chr1.488.a
chr1.489.a
chr1.490.a
chr1.491.a
L35921
D28966
BC063187
AF231407
FQ212716
M16650
X13817
brain
digestion
gland
immune
muscle
nerve
other
Conservation
mouse
human
dog
opossum
chicken
x. tropicalis

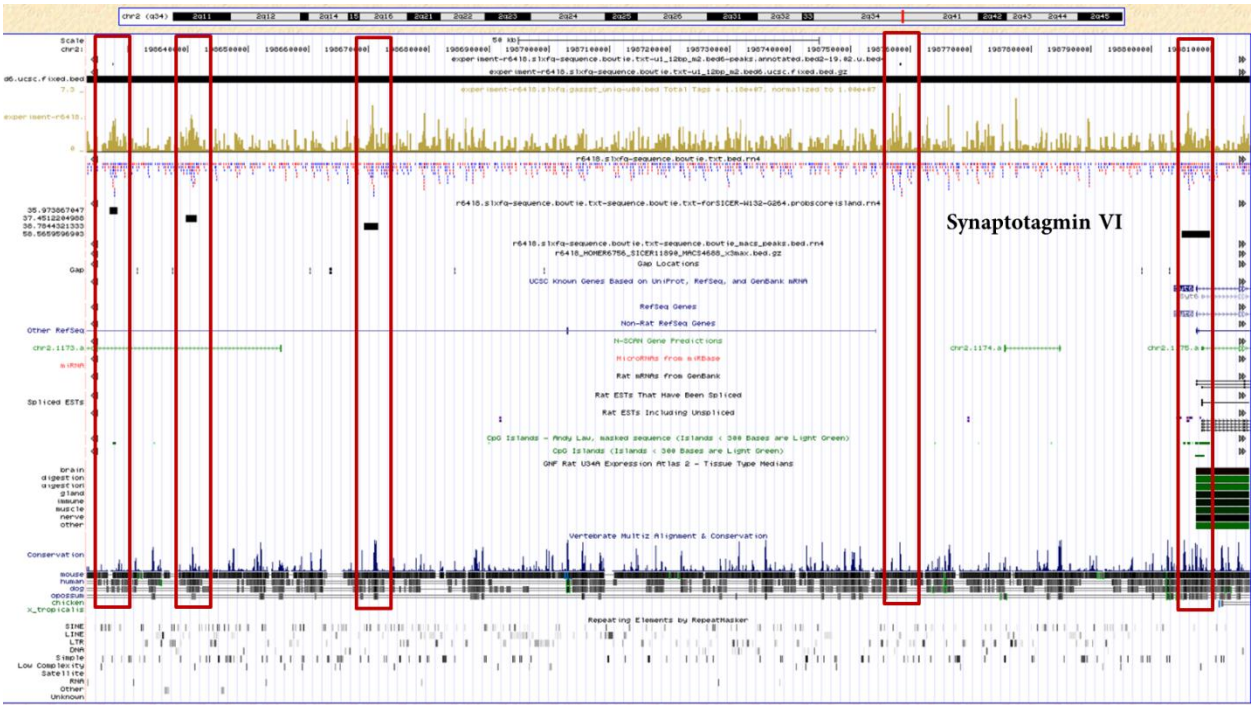


Figure16b. Here shown the LSD1-ChIP seq custom track focused on gene IDs for which multiple LSD1 locations have been found, with at least one occurring within 1000 bp from TSS: calm3 (focus ratio of 0.7, peak score of 12, predicted location 801 bp < TSS) and (focus ratio of 0.65, peak score of 9, predicted location at 411 bp < TSS), synVI (focus ratio of 0.7, peak score of 9). For the distally located peaks no ESTs can be found.

Gene ontology of LSD1 regulated genes

Functional enrichment analysis was performed at first on all the annotated regions obtained from LSD1-ChIP seq that were clustered by means of Gene Ontology categories (herein GO) related to gene functions and the involved biological pathways. The relative GO enrichment was calculated on a background set of gene IDs, assuming the cumulative hypergeometric distribution cutoff at 0.05 by means of HOMER. Nervous tissue enrichment (Fig.17a) and a selection of significantly enriched categories (Fig.17b) are displayed hereafter, related to CNS organogenesis, cell commitment, neurophysiology, synaptogenesis and axonogenesis, indicating LSD1 role in neuronal chromatin remodeling. Gene-enrichment in annotation terms was confirmed on DAVID (http://david.abcc.ncifcrf.gov), by means of a modified Fisher Exact Test test where an *EASE score* is applied penalizing the count of positive agreements by 1, to reduce false positive outputs. The categories associated to ectoderm development and cell cycle checkpoints were retrieved more conservatively from the selected subset of promoters IDs (Fig.17c), for which LSD1 may play a direct regulatory role in the transcription.

Tisse specificity terms	Count	%	PValue	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
Brain	564	16.08671	5.10E-09	2179	2034	9530	1.212729	9.94E-07	9.94E-07	6.36E-06
Hippocampus	106	3.023388	0.013028	2179	378	9530	1.22645	0.922479	0.721575	15.09927
Ovary	126	3.593839	0.036541	2179	474	9530	1.162593	0.999296	0.911047	37.16421
Brain cortex	19	0.541928	0.051039	2179	54	9530	1.538847	0.999963	0.922223	47.99892

Figure 17a. Tissue enrichment analysis was applied to all the LSD1-ChIP seq annotated peaks by means of DAVID (http://david.abcc.ncifcrf.gov).

		Embryogenesis,patternspecification and organogenesis	GO enriched categories		
P-value	LogP	Term	GO ID	Num of Genes in Term	num of Target Genes in Term
1.63E-05	-1.10E+01	forebrain development	GO:0030900	133	51
2.04E-05	-1.08E+01	anatomical structure development	GO:0048856	1785	463
3.23E-05	-1.03E+01	VSPIT1_Q6	VSPIT1_Q6	127	51
4.51E-05	-1.00E+01	pallium development	GO:0021543	44	22
2.55E-04	-8.28E+00	regulation of cell proliferation	GO:0042127	613	172
4.33E-04	-7.75E+00	PATTERN_SPECIFICATION_PROCESS	PATTERN_SPECIFICATION_PROCESS	22	13
5.77E-04	-7.46E+00	telencephalon development	GO:0021537	70	28
6.92E-04	-7.28E+00	pituitary gland development	GO:0021983	27	14
6.92E-04	-7.28E+00	hippocampus development	GO:0021766	27	14
7.60E-04	-7.18E+00	cerebral cortex development	GO:0021987	33	16
7.67E-04	-7.17E+00	limbic system development	GO:0021761	36	17
6.57E-04	-7.33E+00	CENTRAL_NERVOUS_SYSTEM_DEVELOPMENT	CENTRAL_NERVOUS_SYSTEM_DEVELOPMENT	88	35
1.01E-03	-6.90E+00	organ development	GO:0048513	1280	328
1.85E-03	-6.29E+00	NERVOUS_SYSTEM_DEVELOPMENT	NERVOUS_SYSTEM_DEVELOPMENT	293	92
2.18E-03	-6.13E+00	forebrain neuron differentiation	GO:0021879	8	6
2.41E-03	-6.03E+00	diencephalon development	GO:0021536	36	16
2.65E-03	-5.94E+00	BRAIN_DEVELOPMENT	BRAIN_DEVELOPMENT	34	16
2.67E-03	-5.93E+00	EMBRYONIC_DEVELOPMENT	EMBRYONIC_DEVELOPMENT	46	20
2.76E-03	-5.89E+00	gastrulation	GO:0007369	46	19
3.08E-03	-5.78E+00	tissue development	GO:0009888	497	136
3.09E-03	-5.78E+00	developmental growth	GO:0048589	109	37
3.43E-03	-5.67E+00	regulation of cell development	GO:0060284	250	74
8.19E-03	-4.81E+00	embryonic digit morphogenesis	GO:0042733	15	8
1.04E-02	-4.57E+00	female gonad development	GO:0008585	58	21
1.09E-02	-4.52E+00	anatomical structure morphogenesis	GO:0009653	774	198
1.23E-02	-4.40E+00	formation of primary germ layer	GO:0001704	25	11
1.28E-02	-4.36E+00	chordate embryonic development	GO:0043009	239	68
1.31E-02	-4.34E+00	EMBRYONIC_MORPHOGENESIS	EMBRYONIC_MORPHOGENESIS	15	8
1.31E-02	-4.33E+00	tissue morphogenesis	GO:0048729	176	52
1.32E-02	-4.33E+00	gastrulation with mouth forming second	GO:0001702	16	8
1.33E-02	-4.32E+00	forebrain cell migration	GO:0021885	19	9
1.33E-02	-4.32E+00	lung morphogenesis	GO:0060425	19	9
1.49E-02	-4.21E+00	embryo development	GO:0009790	395	106
1.48E-02	-4.21E+00	MORF_GSPT1	MORF_GSPT1	30	13
1.63E-02	-4.12E+00	regulation of odontogenesis	GO:0042481	8	5
1.95E-02	-3.94E+00	specification of symmetry	GO:0009799	20	9
1.95E-02	-3.94E+00	mesoderm formation	GO:0001707	20	9
2.04E-02	-3.90E+00	regulation of epithelial to mesenchymal transition	GO:0010717	14	7
2.11E-02	-3.86E+00	positive regulation of ossification	GO:0045778	30	12
2.11E-02	-3.86E+00	dorsal/ventral pattern formation	GO:0009953	30	12
2.11E-02	-3.86E+00	MORF_EIF4A2	MORF_EIF4A2	92	31
2.26E-02	-3.79E+00	embryonic morphogenesis	GO:0048598	193	55
2.28E-02	-3.78E+00	regulation of ossification	GO:0030278	73	24
2.32E-02	-3.76E+00	camera-type eye morphogenesis	GO:0048593	27	11
2.69E-02	-3.62E+00	gland development	GO:0048732	155	45
2.75E-02	-3.59E+00	mesoderm morphogenesis	GO:0048332	21	9
2.75E-02	-3.59E+00	labyrinthine layer development	GO:0060711	21	9
2.95E-02	-3.52E+00	telencephalon cell migration	GO:0022029	18	8
2.95E-02	-3.52E+00	determination of left/right symmetry	GO:0007368	18	8
2.95E-02	-3.52E+00	positive regulation of multicellular organism growth	GO:0040018	18	8
3.00E-02	-3.51E+00	NEURON_DEVELOPMENT	NEURON_DEVELOPMENT	49	18
3.01E-02	-3.50E+00	negative regulation of osteoclast differentiation	GO:0045671	9	5
3.01E-02	-3.50E+00	retina morphogenesis in camera-type eye	GO:0060042	9	5
3.01E-02	-3.50E+00	placenta blood vessel development	GO:0060674	9	5
3.55E-02	-3.34E+00	eye morphogenesis	GO:0048592	39	14
3.55E-02	-3.34E+00	ORGAN_MORPHOGENESIS	ORGAN_MORPHOGENESIS	110	35
3.63E-02	-3.32E+00	cochlea development	GO:0090102	4	3
3.63E-02	-3.32E+00	regulation of odontogenesis of dentine-containing tooth	GO:0042487	4	3
3.68E-02	-3.30E+00	sensory organ development	GO:0007423	170	48
3.83E-02	-3.26E+00	tube morphogenesis	GO:0035239	111	33
4.13E-02	-3.19E+00	determination of bilateral symmetry	GO:0009855	19	8
4.43E-02	-3.12E+00	ectoderm development	GO:0007398	85	26
4.50E-02	-3.10E+00	retina development in camera-type eye	GO:0060041	33	12
4.52E-02	-3.10E+00	lung development	GO:0030324	89	27
4.59E-02	-3.08E+00	respiratory system development	GO:0060541	93	28
4.74E-02	-3.05E+00	surfactant homeostasis	GO:0043129	7	4
4.74E-02	-3.05E+00	embryonic camera-type eye morphogenesis	GO:0048596	7	4
4.82E-02	-3.03E+00	vasculature development	GO:0001944	181	50
4.92E-02	-3.01E+00	endothelial tube morphogenesis	GO:0061154	2	2
4.92E-02	-3.01E+00	glomerular epithelial cell differentiation	GO:0072311	2	2
4.92E-02	-3.01E+00	globus pallidus development	GO:0021759	2	2
4.92E-02	-3.01E+00	olfactory nerve development	GO:0021553	2	2
4.92E-02	-3.01E+00	glomerular epithelium development	GO:0072010	2	2
4.93E-02	-3.01E+00	negative regulation of muscle organ development	GO:0048635	10	5
4.93E-02	-3.01E+00	epithelial tube branching involved in lung morphogenesis	GO:0060441	10	5
4.95E-02	-3.01E+00	pancreas development	GO:0031016	37	13
1.01E-03	-6.90E+00	cell morphogenesis	GO:0000902	236	73

		Regulation of cell cycle	Go enriched categories		
P-value	LogP	Term	GO ID	Num of Genes in Term	num of Target Genes in Term
8.57E-05	-9.37E+00	regulation of cell cycle	GO:0051726	286	91
2.55E-04	-8.28E+00	regulation of cell proliferation	GO:0042127	613	172
5.95E-04	-7.43E+00	negative regulation of cell proliferation	GO:0008285	251	78
1.56E-04	-8.76E+00	cell morphogenesis involved in differentiation	GO:0000904	179	61
1.10E-03	-6.81E+00	positive regulation of anti-apoptosis	GO:0045768	28	14
5.72E-03	-5.16E+00	cell cycle arrest	GO:0007050	52	20
6.22E-03	-5.08E+00	positive regulation of cell differentiation	GO:0045597	228	67
7.14E-03	-4.94E+00	regulation of S phase of mitotic cell cycle	GO:0007090	12	7
8.22E-03	-4.80E+00	NEGATIVE_REGULATION_OF_CELL_CYCLE	NEGATIVE_REGULATION_OF_CELL_CYCLE	63	24
1.35E-02	-4.31E+00	negative regulation of cell death	GO:0060548	312	86
1.81E-02	-4.01E+00	homeostasis of number of cells	GO:0048872	79	26
1.86E-02	-3.99E+00	positive regulation of cell proliferation	GO:0008284	361	97
1.93E-02	-3.95E+00	muscle cell homeostasis	GO:0046716	11	6
2.13E-02	-3.85E+00	regulation of leukocyte proliferation	GO:0070663	80	26
2.95E-02	-3.52E+00	regulation of mesenchymal cell proliferation	GO:0010464	18	8
4.92E-02	-3.01E+00	mitotic metaphase/anaphase transition	GO:0007091	2	2

		Cell committment	GO enriched categories		
P-value	LogP	Term	GO ID	Num of Genes in Term	num of Target Genes in Term
1.65E-05	-1.10E+01	STEMCELL_EMBRYONIC_UP	STEMCELL_EMBRYONIC_UP	977	288
3.80E-04	-7.88E+00	STEMCELL_NEURAL_UP	STEMCELL_NEURAL_UP	1337	370
6.03E-04	-7.41E+00	STEMCELL_COMMON_UP	STEMCELL_COMMON_UP	137	50
2.41E-03	-6.03E+00	mesodermal cell differentiation	GO:0048333	4	4
9.93E-03	-4.61E+00	glial cell fate commitment	GO:0021781	5	4
1.04E-02	-4.56E+00	regulation of cell differentiation	GO:0045595	460	123
1.08E-02	-4.53E+00	regulation of chondrocyte differentiation	GO:0032330	10	6
1.09E-02	-4.52E+00	positive regulation of odontogenesis of dentine-containing tooth	GO:0042488	3	3
1.09E-02	-4.52E+00	positive regulation of odontogenesis	GO:0042482	3	3
1.09E-02	-4.52E+00	negative regulation of chondrocyte differentiation	GO:0032331	3	3
1.58E-02	-4.15E+00	neuron fate commitment	GO:0048663	29	12
1.58E-02	-4.15E+00	negative regulation of epithelial cell proliferation	GO:0050680	29	12
1.63E-02	-4.12E+00	positive regulation of erythrocyte differentiation	GO:0045648	8	5
1.93E-02	-3.95E+00	regulation of cartilage development	GO:0061035	11	6
2.74E-02	-3.60E+00	regulation of neuron differentiation	GO:0045664	175	50
2.74E-02	-3.60E+00	cell fate commitment	GO:0045165	93	29
4.92E-02	-3.01E+00	glial cell fate determination	GO:0007403	2	2
4.92E-02	-3.01E+00	glomerular visceral epithelial cell differentiation	GO:0072112	2	2
4.92E-02	-3.01E+00	positive regulation of melanocyte differentiation	GO:0045636	2	2
4.92E-02	-3.01E+00	glial cell fate specification	GO:0021780	2	2
4.92E-02	-3.01E+00	mesodermal cell fate commitment	GO:0001710	2	2
4.92E-02	-3.01E+00	oligodendrocyte cell fate commitment	GO:0021779	2	2
4.92E-02	-3.01E+00	oligodendrocyte cell fate specification	GO:0021778	2	2
4.92E-02	-3.01E+00	Schwann cell proliferation	GO:0014010	2	2
1.24E-02	-4.39E+00	STEMCELL_HEMATOPOIETIC_UP	STEMCELL_HEMATOPOIETIC_UP	1041	279

		neurophysiology	GO enriched categories		
P-value	LogP	Term	GO ID	Num of Genes in Term	num of Target Genes in Term
1.91E-03	-6.26E+00	glutamate receptor activity	GO:0008066	26	13
3.80E-03	-5.57E+00	GLUTAMATE_RECEPTOR_ACTIVITY	GLUTAMATE_RECEPTOR_ACTIVITY	18	10
5.28E-03	-5.24E+00	NMDARECEPTOR	PR00177	16	8
5.31E-03	-5.24E+00	regulation of receptor recycling	GO:0001919	9	6
1.45E-02	-4.24E+00	extracellular-glutamate-gated ion channel activity	GO:0005234	16	8
1.45E-02	-4.24E+00	ionotropic glutamate receptor activity	GO:0004970	16	8
1.67E-02	-4.09E+00	NMDA_rcpt	IPR001508	16	8
1.67E-02	-4.09E+00	Iontro_glu_rcpt	IPR001320	16	8
1.67E-02	-4.09E+00	Glutamate_receptor-rel	IPR015683	16	8
1.85E-02	-3.99E+00	transmission of nerve impulse	GO:0019226	255	71
1.96E-02	-3.93E+00	regulation of transmission of nerve impulse	GO:0051969	160	47
2.08E-02	-3.87E+00	GLUTAMATE_SIGNALING_PATHWAY	GLUTAMATE_SIGNALING_PATHWAY	16	8
2.20E-02	-3.82E+00	ionotropic glutamate receptor complex	GO:0008328	14	7
2.20E-02	-3.82E+00	calcium channel regulator activity	GO:0005246	17	8
3.66E-02	-3.31E+00	regulation of synaptic transmission	GO:0050804	150	43
3.99E-02	-3.22E+00	negative regulation of ion transport	GO:0043271	29	11
4.08E-02	-3.20E+00	METABOTROPIC_GLUTAMATE__GABA_B_LIKE_RECEPTOR_ACTIVITY	METABOTROPIC_GLUTAMATE__GABA_B_LIKE_RECEPTOR_ACTIVITY	9	5
4.92E-02	-3.01E+00	clustering of voltage-gated sodium channels	GO:0045162	2	2
4.93E-02	-3.01E+00	ionotropic glutamate receptor signaling pathway	GO:0035235	10	5

		neurogenesis	GO enriched categories		
P-value	LogP	Term	GO ID	Num of Genes in Term	num of Target Genes in Term
3.20E-12	-2.65E+01	nervous system development	GO:0007399	831	267
2.80E-08	-1.74E+01	central nervous system development	GO:0007417	358	124
1.01E-03	-6.90E+00	neuron migration	GO:0001764	46	20
1.06E-07	-1.61E+01	brain development	GO:0007420	276	99
5.31E-03	-5.24E+00	generation of neurons in the forebrain	GO:0021872	9	6
3.11E-07	-1.50E+01	neuron differentiation	GO:0030182	339	115
6.38E-07	-1.43E+01	developmental process	GO:0032502	2121	555
6.54E-07	-1.42E+01	neurogenesis	GO:0022008	523	163
2.03E-06	-1.31E+01	generation of neurons	GO:0048699	482	150
3.81E-06	-1.25E+01	V$SOX5_01	V$SOX5_01	144	59
8.62E-06	-1.17E+01	cellular developmental process	GO:0048869	1175	320
8.80E-06	-1.16E+01	positive regulation of apoptosis	GO:0043065	334	108
2.01E-04	-8.51E+00	neuron recognition	GO:0008038	17	11
3.05E-04	-8.10E+00	neuron projection development	GO:0031175	205	67
3.26E-04	-8.03E+00	FN3	SM00060	102	43
4.19E-04	-7.78E+00	postsynaptic density	GO:0014069	91	35
3.59E-04	-7.93E+00	cell morphogenesis involved in neuron differentiation	GO:0048667	144	50
7.14E-03	-4.94E+00	cerebral cortex cell migration	GO:0021795	12	7
9.93E-03	-4.61E+00	cerebral cortex radially oriented cell migration	GO:0021799	5	4
1.06E-02	-4.55E+00	regulation of neurological system process	GO:0031644	178	53
1.08E-02	-4.53E+00	cerebral cortex neuron differentiation	GO:0021895	10	6
2.46E-02	-3.71E+00	cerebral cortex GABAergic interneuron differentiation	GO:0021892	6	4
2.49E-02	-3.70E+00	regulation of neurogenesis	GO:0050767	214	60
3.99E-02	-3.22E+00	positive regulation of neuron differentiation	GO:0045666	29	11
4.31E-02	-3.14E+00	Ngfr (nerve growth factor receptor (TNFR superfamily, member 16))	24596	5	4
4.42E-02	-3.12E+00	NEUROGENESIS	NEUROGENESIS	72	24
4.92E-02	-3.01E+00	forebrain ventricular zone progenitor cell division	GO:0021869	2	2
4.92E-02	-3.01E+00	neuron cell-cell adhesion	GO:0007158	2	2
4.92E-02	-3.01E+00	forebrain neuron fate commitment	GO:0021877	2	2
4.92E-02	-3.01E+00	negative regulation of CREB transcription factor activity	GO:0032792	2	2
4.93E-02	-3.01E+00	neuron fate specification	GO:0048665	10	5
4.93E-02	-3.01E+00	neuron fate specification	GO:0048665	10	5
4.93E-02	-3.01E+00	cell proliferation in forebrain	GO:0021846	10	5
5.00E-02	-3.00E+00	neural tube formation	GO:0001841	23	9
5.00E-02	-3.00E+00	cell fate determination	GO:0001709	23	9

		Synaptogenesis and axonogenesis	GO enriched categories		
P-value	LogP	Term	GO ID	Num of Genes in Term	num of Target Genes in Term
1.43E-05	-1.12E+01	Fibronectin_III_dom	IPR008957	115	47
1.76E-05	-1.10E+01	fn3	PF00041	81	36
2.21E-05	-1.07E+01	synapse part	GO:0044456	273	91
2.34E-05	-1.07E+01	FN_III	IPR003961	104	43
3.05E-05	-1.04E+01	FN3	PS50853	100	41
5.73E-05	-9.77E+00	cell projection morphogenesis	GO:0048858	170	60
7.00E-05	-9.57E+00	axonogenesis	GO:0007409	129	48
7.11E-05	-9.55E+00	cell projection organization	GO:0030030	266	86
7.81E-05	-9.46E+00	neuron projection	GO:0043005	436	132
9.74E-05	-9.24E+00	neuron projection morphogenesis	GO:0048812	155	55
2.09E-04	-8.47E+00	dendrite	GO:0030425	240	78
1.32E-03	-6.63E+00	terminal button	GO:0043195	40	18
1.39E-03	-6.58E+00	axon	GO:0030424	205	65
1.39E-03	-6.58E+00	synaptic vesicle	GO:0008021	86	32
2.41E-03	-6.03E+00	regulation of dendritic spine development	GO:0060998	4	4
3.53E-03	-5.65E+00	synaptic vesicle membrane	GO:0030672	43	18
3.83E-03	-5.57E+00	axon guidance	GO:0007411	64	24
4.78E-03	-5.34E+00	HSA04360_AXON_GUIDANCE	HSA04360_AXON_GUIDANCE	104	37
5.31E-03	-5.24E+00	axonal fasciculation	GO:0007413	9	6
6.46E-03	-5.04E+00	synaptic transmission	GO:0007268	201	60
6.81E-03	-4.99E+00	V$CREBP1_Q2	V$CREBP1_Q2	141	47
8.57E-03	-4.76E+00	regulation of synaptic transmission, glutamatergic	GO:0051966	24	11
1.09E-02	-4.52E+00	regulation of dendritic spine morphogenesis	GO:0061001	3	3
1.21E-02	-4.42E+00	synapse organization	GO:0050808	38	15
1.44E-02	-4.24E+00	AXONOGENESIS	AXONOGENESIS	33	14
1.74E-02	-4.05E+00	neurotrophin binding	GO:0043121	8	5
2.20E-02	-3.82E+00	beta-tubulin binding	GO:0048487	14	7
2.49E-02	-3.69E+00	synaptic vesicle transport	GO:0048489	34	13
2.54E-02	-3.67E+00	regulation of axon extension	GO:0030516	24	10
2.54E-02	-3.67E+00	synaptosome	GO:0019717	110	34
3.12E-02	-3.47E+00	AXON_GUIDANCE	AXON_GUIDANCE	17	8
3.16E-02	-3.45E+00	dendritic spine	GO:0043197	74	24
3.16E-02	-3.45E+00	neuron spine	GO:0044309	74	24
3.57E-02	-3.33E+00	axon terminus	GO:0043679	60	20
3.57E-02	-3.33E+00	neuron projection terminus	GO:0044306	60	20
4.89E-02	-3.02E+00	presynaptic membrane	GO:0042734	40	14
4.92E-02	-3.01E+00	axon midline choice point recognition	GO:0016199	2	2

Figure 17b. Gene Ontology categorization was performed on all the LSD1-ChIP seq annotated peaks by means of HOMER, assuming the cumulative hypergeometric distribution cutoff at 0.05. Among the significantly enriched categories listed above: embryonic pattern specification, CNS organogenesis, cell commitment and neurogenesis, cell cycle regulation, synaptogenesis and axonogenesis.

Term	Count	%	PValue	Genes	List Total	Pop Hits	Pop Total	Fold Enrichment	Bonferroni	Benjamini	FDR
GO:0030216~keratinocyte differentiation	5	0.749	0.015	NM_013001, NM_012922, NM_031643, NM_134368, NM_001008767	419	28	12092	5.153	1.000	0.561	22.948
GO:0007398~ectoderm development	9	1.347	0.040	NM_001012105, NM_001031644, NM_013001, NM_012922, NM_031643, NM_013193, NM_134368, NM_001008767, NM_001108444	419	112	12092	2.319	1.000	0.773	50.835
GO:0000075~cell cycle checkpoint	6	0.898	0.041	NM_001044259, NM_001031644, NM_001034939, NM_053677, NM_001106138, NM_031058	419	55	12092	3.148	1.000	0.773	51.412
GO:0001701~in utero embryonic development	14	2.096	0.061	NM_053698, NM_134368, NM_001107669, NM_031058, NM_012655, NM_001108444, NM_173101, NM_001108393, NM_017022, NM_001005533, NM_031643, NM_031781, NM_001106447, NM_024360	419	233	12092	1.734	1.000	0.827	66.156
rno04360:Axon guidance	9	1.347	0.068	NM_001134958, NM_017022, NM_017309, NM_001168670, NM_031564, NM_001105717, NM_134366, NM_001107692, NM_134331	192	127	5590	2.063	1.000	0.741	57.112
PIRSF002350:calmodulin	3	0.449	0.072	NM_017309, NM_001135017, NM_031969, NM_012518	180	11	4368	6.618	1.000	1.000	61.141
GO:0010941~regulation of cell death	32	4.790	0.079	NM_019381, NM_001007686, NM_178021, NM_012699, NM_053867, NM_031058, NM_139258, NM_053610, NM_001031644, NM_001079887, NM_001034012, NM_024358, NM_001170327, NM_053698, NM_017258, NM_012922, NM_001008767, NM_022380, NM_053677, NM_032612, NM_001106138, NM_001013150, NM_134366, NM_001034107, NM_053973, NM_134331, NM_001024800, NM_001135008, NM_001100986, NM_017309, NM_001108079, NM_053703	419	689	12092	1.340	1.000	0.852	76.062
GO:0001709~cell fate determination	4	0.599	0.084	NM_024358, NM_013001, NM_134368, NM_024360	419	30	12092	3.848	1.000	0.863	78.014
GO:0051726~regulation of cell cycle	14	2.096	0.088	NM_053698, NM_001044259, NM_012922, NM_001034939, NM_022380, NM_053677, NM_001105725, NM_001106138, NM_001107669, NM_031058, NM_017022, NM_001168559, NM_001031644, NM_001108324	419	247	12092	1.636	1.000	0.865	79.661

Figure 17c. Functional gene enrichment applied on the subset of the genes IDs retrieved from promoters annotated LSD1-ChIP seq regions. Analysis was performed by means of DAVID software (http://david.abcc.ncifcrf.gov/). Here displayed a selection of the significantly enriched categories from the functional annotation chart, for which a 0.1 EASE threshold (maximum EASE score/p-value ratio) and 2 threshold count (minimum number of genes for the corresponding term) were set by default.

To characterize extensively the pattern of TFs that could possibly participate in the regulation of LSD1 target genes, I scanned the identified promoters regions for enriched TF binding sites, by means of MatInspector (http://www.Genomatix® .de/en/index.html). LSD1 bound regions (200 bp width) were extended of 350 bp upstream and 50 bp downstream the located peak, since 600 bp long promoters are randomly extracted from MatInspector library as control set for p-value calculation of TF enrichment (see table below, 0.75/Opt.), (Cartharius et al.,2005) (http://www.Genomatix® .de/online_help/help/scores.html).Two position weight family matrices, grouping similar or functionally related TF binding sites were applied, including the vertebrates (V$) and the general core promoter elements one (O$). The optimized threshold for the matrices is set by default at the minimum similarity level allowing up to three false positive matches in 10.000 bp of non-regulatory test sequences (core/matrix similarity cutoff set by default at 0.75/optimized)(Fig.18, output series obtained with default settings, see methods).

By applying a more stringent core/matrix similarity cutoff (1/opt+0.05) three TF binding sites families keep displaying a significant enrichment: O$TF2B (*p-value* 4.11908E-18),O$TF3C (0.041) core promoter family matrices and V$SIX3(0.048) vertebrate matrix (diencephalon development , GO:0021536 and forebrain anterior/posterior pattern formation GO:0021797). Among the TFBS displaying the highest number of occurrences (number of matches) and representation (number of sequences) V$TALE, V$NEUR and V$CREB (regulation of neuron differentiation, GO:0045665) and V$NKXH (neuron fate specification GO:0048665), $VSP1F and V$TCFF (anatomical structure morphogenesis,GO:0009653) were found.

Search Results (88692 matches)

[Statistics]

MatInspector Release professional 8.0.4, August 2010 — Sat Oct 16 08:50:34 2010

Solution parameters:

Sequence files:	extended 350bp <(prom 200 bp wide annotated peaks)>50bpGASST aligned Homer Called.seq (700 sequences, 420000 bp)
Family matches:	yes
MatInspector library:	Matrix Family Library Version 8.2 (January 2010)
Selected groups:	Vertebrates (0.75/Optimized)
(core/matrix sim)	General Core Promoter Elements (0.75/Optimized)

Statistics:

Matrix Family	p-value Help	No. of matches	No. of sequences
O$TF3A	0.00E+00	1	1
V$NRF1	3.01E-70	1305	338
V$E2FF	1.10E-58	2065	619
O$MTEN	3.75E-57	786	409
V$CDEF	2.34E-54	445	290
V$AHRR	1.97E-38	788	445
V$ZF5F	3.85E-38	760	304
O$XCPE	6.53E-38	616	368
V$SP1F	9.94E-33	1616	562
V$EGRF	1.21E-31	1870	542
V$WHNF	9.93E-30	462	307
V$HIFF	4.85E-25	683	345
V$CTCF	9.15E-22	1190	475
V$DEAF	8.92E-21	361	262
V$HESF	9.85E-21	951	425
O$TF2B	4.12E-18	203	158
V$HNFP	2.81E-16	181	152
V$HDBP	1.33E-14	356	205
V$CHRE	1.67E-12	233	184
V$ZBPF	1.99E-12	1653	494
O$TF2D	3.28E-11	212	175
V$PAX5	2.14E-10	670	393
O$TF3C	6.38E-10	73	66
V$MAZF	3.35E-09	817	404
V$EBOX	1.92E-08	1073	449
V$ETSF	5.93E-08	2329	674
V$PAX9	1.06E-07	192	154
V$MYBL	2.68E-07	1063	538
V$DMTF	3.85E-07	225	191
V$NRSF	1.20E-06	676	403
V$PAX3	1.91E-06	321	261
V$HICF	2.93E-05	220	179
V$XBBF	4.35E-04	719	407
V$YBXF	5.77E-04	251	206
V$ZF35	9.65E-04	232	193
V$E4FF	1.10E-03	323	209
V$AP2F	1.91E-03	327	224
V$GRHL	8.37E-03	407	268
V$KLFS	1.21E-02	1745	595
V$GLIF	1.27E-02	618	394
V$PAX1	1.30E-02	96	91
V$NOLF	1.40E-02	377	277
V$MTF1	1.97E-02	142	114

Figure 18. Prediction of putative TFs co-regulating the identified LSD1bound genes was performed extending the target promoters up to 600 bp, matching the default length of the background promoters randomly extracted from MatInspector library to perform enrichment calculation. Output matching a minimum core/matrix similarity of 0.75/optimized are displayed.

Known motifs enrichment and prediction of *de novo* motifs

Known motifs enrichment and prediction of *de novo* motifs were computed globally on all the annotated regions by means of HOMER software (*Hypergeometric Optimization of Motif EnRichment*, http://biowhat.ucsd.edu/homer/). First, input sequences corresponding to the identified peaks are selected from the genome, calculating the CpG/CG content. Genomic control fragments of a specified size and matching the CpG characteristics of the input sequences are randomly extracted. Motifs are scored for differential enrichment between two sets of genes which are represented by co-regulated gene promoters (input sequences) and either non-regulated extracted genes or background sequences, the latter assumed as a random collection of A,C,G and T. Once target and background sequences are chosen within 100 Kb and checked for CpG islands bias, HOMER searches for motifs of a specific length that are overrepresented in the target set relative to the background set. This enrichment is measured using the Fisher exact test with no requirements in terms of degeneracy or number of occurrences. Due to the finite amount of data and many degrees of freedom in a motif probability matrix, the cutoff is set at 1e-10 to exclude false positives (http://biowhat.ucsd.edu/homer/introduction/motifDetails.html). Motifs are scored and displayed as probability matrices that are representative of the consensus oligo with small arbitrary probabilities assigned to the non-consensus nucleotides. Sequence logos were generated using WebLOGO (http://Iblogo.berkeley.edu) (Fig.19b). Statistics was applied to 50.000 total sequences and 6700 target sequences (Fig.19a).

Motif Name	Consensus	P-value	Log P-value	Number of Genes	Number of regulated Genes	Number of Genes with Motif	number of regulated genes with Motif
NF1-halfsite/LNCaP-NF1-ChIP-Seq/Homer	TTGCCAAG	5.97E-30	-6.73E+01	50000	6699	7740	1359
SeqBias: GCW-triplet	GCAGCAGCAGCA	2.12E-26	-5.91E+01	50000	6699	25465	3814
CTCF/CD4+-CTCF-ChIP-Seq/Homer	ANAGTGCCACCTGGTGGCCA	2.63E-23	-5.20E+01	50000	6699	937	238
X-box/NPC-H3K4me1-ChIP-Seq/Homer	GGTTGCCATGGCAA	2.21E-22	-4.99E+01	50000	6699	495	149
RFX/K562-RFX3-ChIP-Seq/Homer	CGGTTGCCATGGCAAC	4.15E-18	-4.00E+01	50000	6699	377	115
Foxo1/RAW-Foxo1-ChIP-Seq/Homer	CTGTTTAC	2.25E-17	-3.83E+01	50000	6699	9944	1593
Stat3+il23/CD4-Stat3-ChIP-Seq/Homer	NNCTTCCNGGAAGN	8.12E-16	-3.48E+01	50000	6699	3329	604
Foxa2/Liver-Foxa2-ChIP-Seq/Homer	CNTGTTTACATA	3.70E-15	-3.32E+01	50000	6699	4304	749
NFY/Promoter/Homer	AGCCAATCGG	7.32E-13	-2.79E+01	50000	6699	3984	685
Nanog/mES-Nanog-ChIP-Seq/Homer	GGCCATTAAC	9.33E-12	-2.54E+01	50000	6699	21197	3094
NF1/LNCAP-NF1-ChIP-Seq/Homer	CTTGGCANNNTGCCAA	7.55E-11	-2.33E+01	50000	6699	1321	260
Tcf12/GM12878-Tcf12-ChIP-Seq/Homer	NCAGCTGCTG	2.48E-09	-1.98E+01	50000	6699	3895	645
Lhx3/Forebrain-p300-ChIP-Seq/Homer	CTAATTAGCN	1.90E-08	-1.78E+01	50000	6699	5215	830
MyoD/Myotube-MyoD-ChIP-Seq/Homer	AGCAGCTGCTNN	4.76E-08	-1.69E+01	50000	6699	3096	516
NFAT/Jurkat-NFATC1-ChIP-Seq/Homer	ATTTTCCATT	1.14E-06	-1.37E+01	50000	6699	4344	686
Oct4/mES-Oct4-ChIP-Seq/Homer	ATTTGCATAA	1.26E-06	-1.36E+01	50000	6699	2280	383
GFX/Promoter/Homer	ATTCTCGCGAGA	3.53E-06	-1.26E+01	50000	6699	59	22
STAT4/CD4-Stat4-ChIP-Seq/Homer	NTTTCCNGGAAA	3.68E-06	-1.25E+01	50000	6699	4626	721
Tlx/NPC-H3K4me1-ChIP-Seq/Homer	CTGGCAGNCTGCCA	7.69E-06	-1.18E+01	50000	6699	1595	274
Oct2/Bcell-Oct2-ChIP-Seq/Homer	ATATGCAAAT	9.54E-06	-1.16E+01	50000	6699	1613	276
CTCF-SatelliteElement/CD4+-CTCF-ChIP-Seq/Homer	TGCAGTTCCNNNNNTGGCCA	1.27E-05	-1.13E+01	50000	6699	86	27
JunD/K562-JunD-ChIP-Seq/Homer	ATGACGTCATCN	2.08E-05	-1.08E+01	50000	6699	361	77
YY1/Promoter/Homer	CAAGATGGCGGC	4.86E-05	-9.93E+00	50000	6699	334	71
STAT5/mCD4+-Stat5a|b-ChIP-Seq/Homer	NTTTCTNAGAAA	8.82E-05	-9.34E+00	50000	6699	1536	257
CEBP:AP1/ThioMac-CEBPb-ChIP-Seq/Homer	NATGTTGCAA	2.57E-04	-8.27E+00	50000	6699	3304	510
AR-halfsite/LNCaP-AR-ChIP-Seq/Homer	CCAGGAACAG	4.09E-04	-7.80E+00	50000	6699	19421	2727
ZBTB33/GM12878-ZBTB33-ChIP-Seq/Homer	GGNTCTCGCGAGAAC	6.13E-04	-7.40E+00	50000	6699	133	32
Stat3/mES-Stat3-ChIP-Seq/Homer	CTTCCNGGAA	6.31E-04	-7.37E+00	50000	6699	2424	379
CEBP/CEBPb-ChIP-Seq/Homer	ATTGCGCAAC	6.32E-04	-7.37E+00	50000	6699	2813	435
EWS:ERG-fusion/CADO_ES1-EWS:ERG-ChIP-Seq/Homer	ATTTCCTGTN	1.34E-03	-6.61E+00	50000	6699	3722	560
Sox2/mES-Sox2-ChIP-Seq/Homer	NCCATTGTTC	1.82E-03	-6.31E+00	50000	6699	4546	674
E2F/Cell-Cycle-Exp/Homer	TTCGCGCGAAAA	1.88E-03	-6.28E+00	50000	6699	182	39
TEAD(TEA Domain)/Fibroblast-PU.1-ChIP-Seq/Homer	NCTGGAATGC	2.96E-03	-5.82E+00	50000	6699	2930	443
c-Jun-CRE/K562-cJun-ChIP-Seq/Homer	ATGACGTCATCN	3.06E-03	-5.79E+00	50000	6699	1160	188
FOXA1/LNCAP-FOXA1-ChIP-Seq/Homer	AAAGTAAACA	3.58E-03	-5.63E+00	50000	6699	6936	1001
GFY/Promoter/Homer	ACTACAATTCCC	3.88E-03	-5.55E+00	50000	6699	250	49
ETS/Promoter/Homer	AACCGGAAGT	8.47E-03	-4.77E+00	50000	6699	1639	253
Pbx3/GM12878-PBX3-ChIP-Seq/Homer	NCTGTCAATCAN	9.28E-03	-4.68E+00	50000	6699	931	150
ELF1/Jurkat-ELF1-ChIP-Seq/Homer	ANCCGGAAGT	1.03E-02	-4.58E+00	50000	6699	2590	387
NF1:FOXA1/LNCAP-FOXA1-ChIP-Seq/Homer	NNTGTTTATTTTGGCA	1.27E-02	-4.37E+00	50000	6699	203	39
TATA-Box/Promoter/Homer	CCTTTTATAGNC	1.48E-02	-4.21E+00	50000	6699	6694	954
AARE/mES-cMyc-ChIP-Seq/Homer	GATTGCATCA	1.83E-02	-4.00E+00	50000	6699	285	51
ATF3/K562-ATF3-ChIP-Seq/Homer	NGGTCACGTGAC	1.95E-02	-3.94E+00	50000	6699	665	108
NFkB-p65/GM12787-p65-ChIP-Seq/Homer	NGGGGATTTCCC	2.04E-02	-3.89E+00	50000	6699	1963	294
Unknown/Homeobox/Limb-p300-ChIP-Seq/Homer	NGCAATTAAA	2.34E-02	-3.76E+00	50000	6699	3339	486
PAX3:FKHR-fusion/Rh4-PAX3:FKHR-ChIP-Seq/Homer	ACCGTGACTAATTNN	2.41E-02	-3.73E+00	50000	6699	946	148
SPDEF/VCaP-SPDEF-ChIP-Seq/Homer	ACATCCTGNT	3.65E-02	-3.31E+00	50000	6699	5143	731
FOXA1/MCF7-FOXA1-ChIP-Seq/Homer	AAAGTAAACA	3.71E-02	-3.30E+00	50000	6699	5739	813
CRE/Promoter/Homer	CGGTGACGTCAC	4.41E-02	-3.12E+00	50000	6699	826	128

Figure19a. Known motifs enrichment. Analysis performed by means of HOMER software on GASSST aligned tags.

1. NF1-halfsite/LnCAP-Chip seq-homer, p-value 5.97 e-30

2. Seq-bias GCw triplet, p-value 2.12 e-26

3. CTCF/CD4+-CTCF Chip seq-homer, p-value 2.63 e-23

4. X-box NPC/H3K4 1me —Chip seq-homer, p-value 2.21 e-22

5. RFX,K562 , RFX3, Chip-seq-homer, p-value 4.15 e-18

6. FOXO1-RAW/FOXO1 Chip-seq-homer, p-value 2.25 e-17

7. Stat3 + IL23/CD4-Stat3 Chip-seq-homer, p-value 8.12 e-16

Figure 19b. LSD1-sv ChIP sequenced enriched known motifs displayed as probability matrices, with the consensus sequence and the related co-occurring lower probability sequences. Motifs are displayed by means of WebLogo graphics. Known enriched motifs are listed from the top with the highest p-value significance.

De novo motifs discovery was done similarly, by computing motifs assignments on a set of total unrelated genes extracted from the rat genome and comparing the relative nucleotide occurrence with a selected number of target genes. Alignments were manually checked for any putative motif to occur in the core of the *best match* matrix. Low complexity motifs containing degenerate nucleotide sequences were discarded, as well as simple repeat motifs and motifs occurring in less than 5% of the target set (< 330 occurrences). *De novo* motifs meeting these criteria include Sox18, HNF1b, NR2e3 and TCF7. More permissively, GFY/STAF, Spt2, AR, RfxDc2, Stb3, Spdef2, were also retrieved (Fig.20a and b).

De novo motifs

P-value	log P-value	Best Match/Details	Alignment de novo motif with the best match	Total seq	target seqs	Total instances of Motifs	Total instances of motifs in targets
0.00E+00	-1.63E+03	PB0171.1_Sox18_2	ACTGAATTCCACC nnnnTGAATTCAnnnC	50	6700	335	330
0.00E+00	-1.42E+03	GFY-Staf/Promoter/Homer	AGTTCCCAATAG AACTACAATTCCCAGAATGC			298.1	293
0.00E+00	-8.94E+02	MA0387.1_SPT2	CCGTTCGTTAAA TTCNTTAANT			154	153
0.00E+00	-8.73E+02	MA0007.1_Ar	ACGTTGCTCTAT GGNNGGTACANGNTGTTCTTAN			151	150
0.00E+00	-8.23E+02	PH0168.1_Hnf1b	AACTAACTGA AGCTGTTAACTAGCCGT			411	340
7.119e-310	-7.12E+02	PB0160.1_Rfxdc2_2	CGAATCTAAGTA NTNNCGTATCCAAGTNN			176	170
2.33E-212	-4.87E+02	MA0390.1_STB3	AAATTTTGAT NNNNAGTGAAAAATTTTNGAC			230	193
1.08E-184	4.24E+02	PB0181.1_Spdef_2	GCGAATCCTA GATAACATCCTAGTAG			80	80
2.12E-170	-3.91E+02	MA0164.1_Nr2e3	AATGCTTG AAGCTTG			2911	934
2.37E-125	-2.87E+02	PB0187.1_Tcf7_2	CTTATTAC TTNCCAGCTG			362	214
5.36E-116	-2.65E+02	PF0160.1_CAGNYGKNAA	CCCATCTG TTTNCCAGCTG			4253	1102
1.08E-64	-1.47E+02	PB0056.1_Rfxdc2_1	CTAGGCGA NCCGTTGCTANGNGN			174	105

Figure 20a. *De novo* motifs enriched. Analysis performed by means of HOMER software on GASSST aligned tags

Alignment: --ACTGAATTCACC—

NNNNTGAATTCANNNC

Best assignment: Sox18

GFY/STAF/Promoters

SPT2

Ar

HNF1b

Rfxdc2,p-value 7.119 e-310

STB3, p-value 2.327 e-212

.Spdef2,p-value 1.079e-184

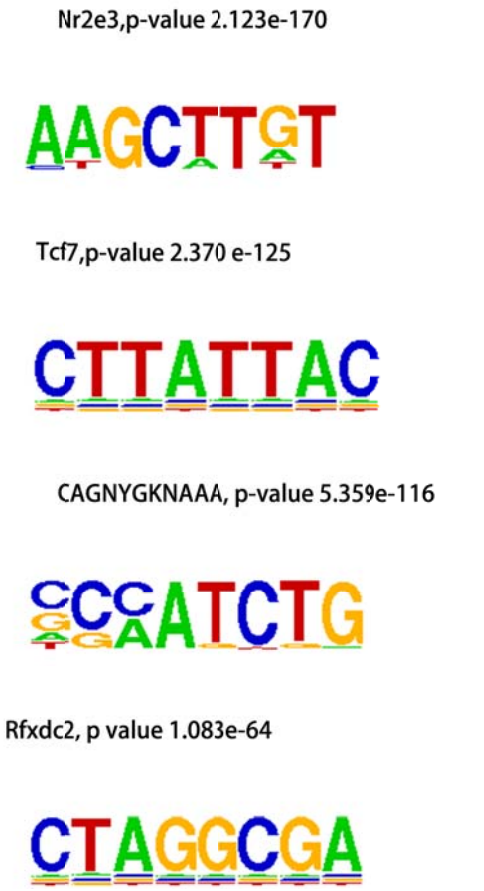


Figure 20b. LSD1-sv ChIP sequenced enriched *de novo* motifs displayed as probability matrices, with the consensus sequence and the related co-occurring lower probability sequences. b) Motifs are displayed by means of WebLogo graphics. De novo motifs are listed from the top with the highest p-value enrichment significance. The indicated name refers to the best assignment provided by HOMER. *De novo* assignments were manually checked for the alignment to occur in the core of the *best match* motif.

A parameter which is considered for the prediction of the binding site of the immunoprecipitated TF as well as for the motif enrichment analysis is the tags distribution around TSS (Fig.21) and the *focus ratio*, defined as the ratio of tags located 5' of the peak center on either strand relative to the total number of tags in the peak. Binding sites are generally focused (focus ratio > 0.8) for transcription factors and unfocused for histone modifications, as result of several closely spaced binding sites or large complexes that crosslink to multiple positions along the DNA. From GASSST aligned tags, 59.8% of identified LSD1 peaks display a focus ratio below 0.8, going along with the known LSD1 K4H3 histone modification activity. Although no DNA binding domain has been described for LSD1, according to its structural characterization (Forneris et al.2005), 40.2% of peaks display a focus ratio of at least 0.8, accounting for a direct LSD1 regulatory role at promoters regions, possibly recruiting molecular scaffolds (Tsai et al. 2010) and other transcription factors.

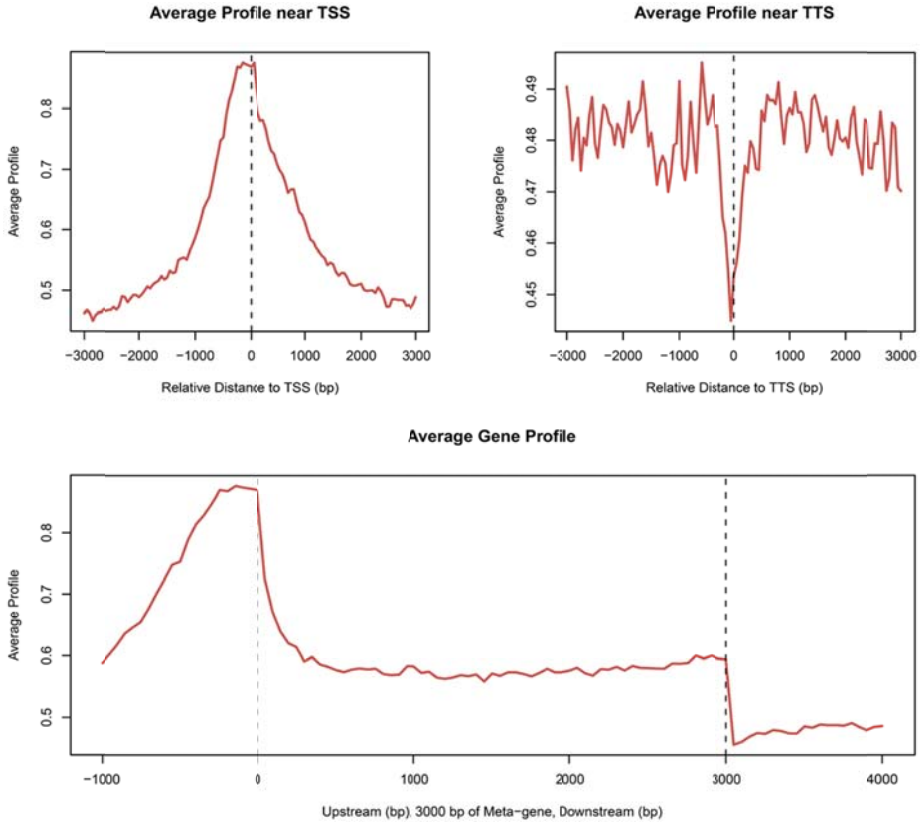


Figure 21. Average distribution of LSD1 sv-Chip tags near the transcriptional start sites, transcriptional termination site and within the gene body. LSD1 sv profile is centered on the TSS with a nucleosomal depletion at the TTS, indicating LSD1-promoter regulatory role. Average profiles were built on peaks called by MACS by means of CEAS software.

Therefore, I searched for TF binding sites overrepresented in the 200 bp wide annotated promoters RegionMiner (http://www.Genomatix® .de/en/index.html) in order to predict any structural subclasses of transcription factors that may directly interact with the LSD1 chip regions. The length of the input sequences was left unaltered (Fig.22) excluding false positive matches that may arise from a common TFs search task applied on extended sequences (Fig.18). The analysis was also performed on intergenic annotated regions localized within 3000 bp from TSS obtained from LSD1 ChIP sequencing. All occurrences of the matches are calculated by MatInspector over promoter background in Rattus Norvegicus NCBI build 4 reference genome. For each predicted TFBS the following parameters were computed (Fig. 22): *promoter association* referring to TFBS families known to occur more than twice as often in promoters as in genomic sequence, *total number of input sequences* displaying a match with the known TFBS matrix and the *total number of matches*. Further parameters are the *expected number of matches* (within an equally sized background sample retrieved from genome or non-related promoters), *overrepresentation* based on a binomial distribution (fold increase of matches found in the input set relative to the number of matches observed in the equally sized background set) and *the z-score* (distance from the population mean in units of the population standard deviation, see methods).

Results: Overrepresented TF Families

Matrix Lil Matrix Library 8.2

Extension
350 bp upstream, 50 bp downstream

Input: 700 sequence(s) with ¿ 1000bp<TSS located promoters

Z-score: the distance from the population mean in units of the population standard deviation.
Promoter association: TF Families known to occur more than twice as often in promoters as in genomic se

Listing of all TF Families

TF Famili	Prom. ass	Z-Score (i	Z-Score (promoters)
V$E2FF	no	64.01	29.71
V$NRF1	yes	196.37	25.35
V$HIFF	no	34.93	16.28
V$CDEF	yes	74.47	16.27
O$MTEN	yes	82.64	16.01
V$HESF	yes	38.69	15.02
V$ZF5F	yes	93.68	14.77
V$WHNF	yes	52.23	13.91
V$AHRR	yes	39.15	13.48
V$ETSF	no	22.66	12.11
V$MYBL	no	15.31	10.65
V$DEAF	yes	37.54	10.37
V$HNFP	yes	21.14	9.52
V$EBOX	no	28.02	9.51
V$E4FF	no	11.92	9.12
O$XCPE	yes	62.04	8.42
V$CREB	no	7.89	8.18
O$TF2D	yes	22.18	7.85
V$XBBF	no	16.62	7.56
V$DMTF	no	14.9	6.35
V$PAX3	no	12.83	6.32
V$PAX5	yes	28.76	6.29
V$GRHL	no	7.08	6.14
V$CHRE	yes	21.82	6.05
O$TF3C	no	6.1	5.79
V$NRSF	yes	26.72	5.32
V$YBXF	no	8.78	5.22
V$GABF	no	12.26	5.15
V$CAAT	no	1.32	5.03
V$PAX9	yes	22.22	4.96
V$HEAT	no	2.51	4.82
V$ZF35	no	12.78	4.36
V$SP1F	yes	60.18	3.88
V$EGRF	yes	66.89	3.81
V$SMAD	no	7.67	3.51
O$TF2B	yes	40.3	3.48
V$PAX1	no	4.51	3.37
V$GZF1	no	-1.15	2.77
V$MOKF	no	7.97	2.58
V$RARB	no	1.36	2.5
V$RBP2	no	9	2.39
V$HICF	yes	20.32	2.24
V$ZFIA	no	6.68	2.15
V$MEF3	no	4.38	1.96
V$GCNR	no	-3.57	-1.9
V$GCMF	no	5.33	-2.03
V$AP4R	no	6.34	-2.12
V$MAZF	yes	36.5	-2.16
V$P53F	no	0.2	-2.17
V$THAP	no	-4.01	-2.21
V$MYT1	no	-13	-2.26
V$RNCF	no	0.92	-2.28
V$LEFF	no	-7.44	-2.38
V$FXRE	no	-2.28	-2.39
V$PPAR	no	-4.46	-2.51
V$BRAC	no	-7.18	-2.55
V$IRFF	no	-10.59	-2.65
V$AP1R	no	-1.84	-2.67
V$SREB	no	-0.95	-2.77
V$TCFF	no	-5.38	-2.87
V$NR2F	no	-3.61	-3.03
V$ZFXY	yes	9.51	-3.04
V$PTF1	no	-1.55	-3.08
V$CHRF	no	-9.23	-3.09
V$RORA	no	-7.86	-3.18
V$YY1F	no	-9.33	-3.18
V$PAX2	no	-9.73	-3.18
V$SIXF	no	-7.74	-3.31
V$HMTB	no	-9.66	-3.39
V$SNAP	no	-10.44	-3.39
V$AIRE	no	-7.6	-3.44
V$AP1F	no	-7.35	-3.47
V$NFAT	no	-9.66	-3.48
V$RPS8	no	-4.69	-3.56
V$BCL6	no	-9.09	-3.64
V$PERO	no	-4.29	-3.71
V$HOXH	no	-10.38	-3.78
V$NF1F	no	-4.11	-3.93
V$PRDF	no	-8.21	-3.96
V$TEAF	no	-5.65	-3.98
V$MITF	no	-7.22	-4.17
V$INSM	no	5.78	-4.19
V$KLF5	no	20.65	-4.2
V$NEUR	no	-2.1	-4.2
V$MZF1	no	9.43	-4.24
V$SIX3	no	-6.69	-4.31
V$PURA	yes	3.25	-4.4
V$RREB	no	6.48	-4.57
V$CLOX	no	-16.51	-4.67
V$GFI1	no	-11.14	-4.73
V$CEBP	no	-12.4	-4.8
V$ZFHX	no	-5.25	-4.98
V$SAL1	no	-10.44	-5.22
V$RXRF	no	-2.11	-5.44
V$RUSH	no	-15.95	-5.47
V$CHOP	no	-8.49	-5.73
V$BCDF	no	-16	-6.01
V$SRFF	no	-12.18	-6.27
V$ZBPF	yes	46.65	-6.29
V$AARF	no	-9.15	-6.41
V$MYOD	no	3.56	-6.61
V$HOXC	no	-18.37	-6.95
V$PLAG	yes	16.28	-7.1
V$FAST	no	-15.89	-7.34
V$HOMF	no	-28.89	-8.23
V$NKX1	no	-16.51	-8.27
V$PAXH	no	-18.05	-8.28
V$GATA	no	-20.75	-8.51
V$NKXH	no	-22.59	-8.54
V$HNF6	no	-19.4	-8.95
V$HAND	no	-10.13	-8.98
V$ARID	no	-20.23	-9.07
V$HNF1	no	-23.73	-9.32
V$SORY	no	-25.54	-9.61
V$PARF	no	-23.33	-9.86
V$ATBF	no	-18.72	-9.87
V$MEF2	no	-21.77	-10.17
V$SATB	no	-19.22	-10.19
V$DMRT	no	-22.58	-10.29
V$CDXF	no	-22.76	-10.33
V$IRXF	no	-18.85	-10.49
O$PTBP	no	-23.3	-10.78
O$YTBP	no	-23.91	-10.92
V$EVI1	no	-22.37	-11.22
V$PDX1	no	-21.95	-11.5
V$DLXF	no	-24.14	-11.93
V$FKHD	no	-30.21	-12.14
V$ABDB	no	-29.21	-12.63
V$PIT1	no	-24.51	-12.74
O$VTBP	no	-30.03	-12.94
V$NKX6	no	-27.09	-13.82
V$CART	no	-34.7	-15.5
V$HBOX	no	-34.98	-16.89
V$BRNG	no	-33.91	-17.1
V$HOXF	no	-39.07	-19.19
V$LHXF	no	-41.19	-20.06
V$OCT1	no	-42.25	-21.63
V$BRNF	no	-45.28	-23.24

Results: Overrepresented TF Families

Extension
350 bp upstream, 50 bp downstream

Input: 114 sequence(s) with a 3000-1000bp<TSS intergenic regions

Z-score: the distance from the population mean in units of the population standard deviation.
Promoter association: TF Families known to occur more than twice as often in promoters as in genomic se

Listing of all TF Families

TF Famili	Prom. ass	Z-Score ((	Z-Score (promoters)
V$E2FF	no	13.78	4.88
O$TF3A	no	2.86	3.23
V$ZFTR	no	1.1	3.07
V$TEAF	no	1.45	2.58
V$HOMF	no	-5.1	2.18
O$MTEN	yes	17.19	2.07
V$AHRR	yes	10.75	2.06
V$NEUR	no	2.09	2.03
V$HBOX	no	-5.64	1.96
V$MYOD	no	2.17	-1.99
V$NRF1	yes	16.11	-2.11
V$ZBPF	yes	9.11	-2.11
V$EGRF	yes	9.07	-2.12
V$OCT1	no	-10.73	-2.27
V$AP2F	yes	2.08	-2.38
V$PAX5	no	2.4	-2.47
V$GABF	no	-5.5	-2.8
O$TF2B	yes	3.42	-2.88

Figure 22. RegionMiner analysis of overrepresented TFBSs was applied to the 200 bp wide promoters associated (1000<TSS)(upper table) and a subset of intergenic associated (3000-1000bp<TSS) (lower table) LSD1-ChiP-seq regions.The following parameters are indicated: promoter association,total number of input sequences, total number of matches. Further parameters are the expected number of matches overrepresentation over each considered background set the z-score . Matches displaying a significant $-2 < Z < 2$ score are indicated in bold.

The analysis of *de novo* motifs was also repeated on the annotated promoters (regions located <1000 bp from a RefSeq associated TSS) by means of CoreSearch tool(http://www.Genomatix® .de/en/index.html) (Wolfertstetter et al, 1996). Quality thresholds include the number of binding sites in a target sequence (at least 4 nt) and the number of matches to the matrix expected in a random sequence of 1000 bp (<5): the random expectation value (RE-value) of the defined motifs is given in the output(Fig.23) in order to estimate the quality of the motifs. Large-scale analysis (>250 input sequences) allowing the computation of the most represented motif was first performed scanning the overall set of promoters for a minimum 7 to 10 bp core, a minimum match occurrence in 5% of input sequences and matrix similarity cutoff at 0.8 on both strands and with unlimited number of matches per sequence. The best match identified corresponds to V$ETSF family matrix (151 matches /609 aligned seqs, embryonic development GO:0009790, organ morphogenesis GO:0009887, skeletal system development GO:0001501, multicellular organismal development GO:0007275, negative regulation of cell proliferation GO:0008285). Notably V$ETSF family matrix comprises Spdef2 previously matched with a *de novo* motif by HOMER analysis performed on the overall pool of ChIP sequences. Analysis was reiterated on subsets of 250 promoters regions and on intergenic regions located within 3000 bp from the TSS in order to compute the 10 most frequent *de novo* motifs. Among the best matches the V$FKHD family matrix was found including the FOXO1 TF , previously identified as best match by *de novo* HOMER analysis on all the LSD1 bound regions (embryonic development GO:0009790, skeletal system development GO:0001501, anatomical structure morphogenesis GO:0009653), V$RXRf (embryonic hindlimb morphogenesis GO:0035116, induction of apoptosis by hormones GO:0008628, negative regulation of cell proliferation GO:0008285, steroid metabolic process GO:0008202), V$SF1F (embryonic development GO:0009790, anatomical structure morphogenesis GO:0009653, cell differentiation GO:0030154, male gonad development GO:0008584,regulation of steroid biosynthetic process GO:0050810) and V$NRF1(generation of precursor metabolites and energy, GO:0006091).

GEMS Launcher Task: CoreSearch: Definition of unknown motifs
working on

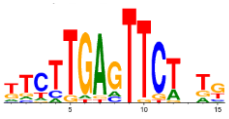
CoreSearch Release 6.2 Mon Oct 18 05:34:07 2010

Solution parameters multiple motifs

Sequence file: intergenic annotated regions (3000-1000<TSS),(114 sequences, 22800 bp)
Length of core: 10 bp
Maximum number of motifs: 10
Min. number of sequences in p 6 sequences (5 % of 114)
Number of motif matches per s any number of repetitions
A priori frequency of nucleotid determined from input sequences (A: 0.25, C: 0.25, G: 0.25, T: 0.26)
Strand(s) searched: both strands

Matrix similarity threshold: 0.8

Motifs defined	Core	Unknown Motif (IUPAC consensus)	Num of seq	Number of Matches	Re-value	best match (family matrix)
Motif 1	TCTTGAGTTC	K TCTTGAGTTC TNTK	7	21	0.42	-
Motif 2	CAGCTCTGTA	N CAGCTCTGTN	8	33	0.78	V$FKHD(4/33)
Motif 3	CAGGCTCCTG	N CAGNCTCCTG NNAG	7	31	1.07	V$GATA(4/31)
Motif 4	TCCCAGGCTC	N TCCNAGGCTS NNN	7	31	1.49	OKF(4/31),V$RXRf(4/31),V$SF1F(4/31)
Motif 5	TTGAGGAAAA	N TTNARGAAAA N	7	44	5.44	-
Motif 6	CCCCGCCCCA	N CCYCGCCCCM	7	42	3.45	V$ZBPF(18/42)
Motif 7	CAGGGACTTG	N CAGGGACTTG	7	32	1.48	V$MOKF(7/32)
Motif 8	TTCCTTTTTC	NWYWNMM YTCCTTTTTC N	7	14	0.03	V$ETSF(7/14)
Motif 9	AGAGCAACAA	N AGANCAACAA N	7	32	3.48	V$HOMF(5/32),V$LEFF(5/22)
Motif 10	TGAGACATCT	N TGAGMCATCT NBMBNKYY	6	8	0	V$HAND(4/8)



Solution parameters large scale analysis: one motif

Sequence file: promoters annotated regions (1000-TSS),(700 sequences, 140000 bp)
Length of core: 7 bp
Maximum number of motifs: 1
Min. number of sequences in p 35 sequences (5 % of 700)
Number of motif matches per s any number of repetitions
A priori frequency of nucleotid determined from input sequences (A: 0.25, C: 0.25, G: 0.25, T: 0.26)
Strand(s) searched: both strands
Matrix similarity threshold: 0.8

matrix (IUPAC consensus)	Num of Aligned Seqge similarity among r	Re-value	best match (family matrix)
NTTCANTTN.		14.16	-
NTTTAANRN.		15.52	
NTTNNTTTN.		7.63	
NAAARYTTSN		12.88	
NTTAANTN..		8.81	

Length of core: 10 bp
Matrix similarity threshold: 0.8

SANCCACTTCCGGN.	0.49	-
..NMAATYACCTTTN	1.23	
YGGGAANTGTAGTYC	0.01	
.NTNTGCTTYGTTTN	0.19	
.NTTTTTNTTCTTN.	1.72	

Length of core: 10 bp
Matrix similarity threshold: 0.7 V$ETSF(151/609)

NCCACTTCNGGN	10.91
NAATYACCTTTN	11.50
NACTWCANTTCCC	3.03
NAAACAAAGCAN	10.01
NTTTTTNTTCTN	11.70

Solution parameters multiple motifs

Sequence file: promoters annotated regions (1000-TSS),(700 sequences, 140000 bp)
Length of core: 10 bp
Maximum number of motifs: 10
Min. number of sequences in p 35 sequences (5 % of 700)
Number of motif matches per s any number of repetitions
A priori frequency of nucleotid determined from input sequences (A: 0.25, C: 0.25, G: 0.25, T: 0.26)
Strand(s) searched: both strands
Matrix similarity threshold: 0.8

Motifs defined	Core	Unknown Motif (IUPAC consensus)	Num of seq	Number of Matches	Re-value	best match (family matrix)
Motif 1	CTCAGCCTCC	NCTSCWRC CTCAGCCTCC Y	17	57	0.09	V$ETSF(12/57)
Motif 2	GGAAGCTGAG	N GGAARCTGAG N	13	105	5.29	V$ETSF(17/105)
Motif 3	CACCAGCCTG	RYTGNA CWCCAGCCTG G	14	22	0.01	-
Motif 4	GGAGAGCAGC	N GNAGAGCAGC N	15	111	4.59	V$MYOD(11/111)
Motif 5	TCCCAGCTAC	RCCTGTAR TCCCAGCTAC	16	16	0	V$ZFHX(6/15)
Motif 6	TGCGCATGCG	S TGCGCNNGCG NN	13	83	2.29	V$NRF1(48/830
Motif 7	CCCGGAAGTG	NCCGGAAGTG N	12	83	2.36	V$ETSF(63/83)
Motif 8	CACTCCAGCC	NACTCCAGCC TGN	14	49	1.69	

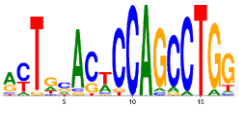


Figure 23. The definition of unknown motifs was applied selectively on the promoters set (1000 bp>TSS) or in an intergenic subset (3000-100 bp>TSS) by means of CoreSearch tool,with a minimum core length of 7 bp, a minimum occurrence of 5% input sequences on both strands, unlimited number of motif matches per sequence and equal nucleotide expected distribution. The matrix similarity threshold was set at 0.7.

DISCUSSION

A variety of chromatin remodeling complexes are thought to orchestrate transcriptional programs that lead neuronal precursors from earliest commitment to terminal differentiation, dynamically enduring in response to environmental cues (Greer et al.2008), though most epigenetic complexes display an ubiquitous pattern of expression which enables them to sort widespread biological functions, setting the need to further investigate possible neuro-specific mechanisms underlying neuronal specification and development.

Alternative splicing is considered one of the most powerful biological devices that convey diversification of gene function without a corresponding increase in gene number (Xing et al. 2006, Kopelman et al. 2005, Parmley at al. 2007) giving rise to evolutionary complexity, especially for mammalian species. Notably, most of splicing events occur on CNS related transcripts, expanding molecular variability of the resultant proteins, which participate to fundamental nervous processes from axon guidance and synapse formation (Lipscombe 2005, Ule and Darnell 2006; Licatalosi and Darnell 2006) to higher cognitive functions, including learning and memory (Grabowski et al. 2001).

Here I report that mammalian neurons express a functionally specialized splice variant of the histone demethylase LSD1, which is restricted to the neuronal histotype: the neuro-specific LSD1 variant is dynamically regulated during perinatal brain development and early synaptic establishment when it contributes to the acquisition of neurite morphology and its functional diversification partly relies on exon specific phosphorylation. Although LSD1 epigenetic function in the vertebrate nervous system has remained elusive, its implication in neuronal processes has been inferred previously through its interactor, the transcriptional silencer REST which prevents the ectopic expression of neuron-specific genes outside CNS and mediates chromatin plasticity throughout neurogenesis (Ballas et al., 2005), whereas in the adult nervous system (Garriga-Kanut et al. 2006; Palm et al., 1998) REST function can be modulated by alternative splicing (Palm et al., 1998; Shimojo et al. 1999; Zuccato et al., 2003).
A comparative analysis of LSD1 genomic sequence across vertebrates and exon retrieval from ESTs databases led me to identify three mammal LSD1 variants, generated by either single or double inclusion of two alternatively spliced exons, namely E2a and E8a. Thus, there are four LSD1 isoforms: LSD1, the conventional one, LSD1-2a, LSD1-8a, and LSD1-2a/8a. The former two are ubiquitous, while the latter ones are restricted to a terminally differentiated neuronal histotype. E2a is conserved from the lizard and

chicken genomes till mammals whereas the neuro-specific exon E8a is exclusively present in mammals. Although traces of E8a can be found also in the lizard genome suggesting that it appeared together with E2a in the common ancestor of mammals, birds and reptiles, it is absent in all non mammal organisms, indicating that selective pressure may have determined the functional preservation of E8a in mammals, possibly related to brain ontogenesis. Most relevant, the processing of LSD1 transcript undergoes a dynamic modulation within the perinatal window, when fundamental processes take place, as confirmed by the time course analysis of several synaptogenic markers. In this phase, exon E8a inclusion frequency is dramatically prompted, resulting in the preponderance of neurospecific isoforms over the ubiquitous ones. Indeed, transcriptional profiling of the developing rat brain revealed that the most dramatic changes in gene expression occur postpartum (Stead et al., 2006), underscoring the relevance of early postnatal life as a critical phase of neural organization and differentiation (Shalizi et al. 2007, Kim et al.2008). The splicing dynamics of LSD1 isoforms can be fairly reproduced in cultured cortical neurons, with an inversion of proportions occurring between the second and the fourth day *in vitro* (DIV). In this model the overexpression of neurospecific LSD1 isoforms at early stages promotes neurite morphogenesis, as revealed by the analysis of neurite arborisation, branches complexity and neurite thickness, whereas the imbalance of the ubiquitous isoforms devoid of E8a elicits no effect, suggesting that the four amino acids peptide coded by E8a is responsible for the observed phenotype, reflecting the timing of neuronal development; consistently, the reversal of this effect can be observed upon silencing of the neuro-specific nsLSD1 splice variants. Complexively, these results relate the "wave-like" perinatal expression of exon E8a to the acquisition of neuronal morphology, setting the basis for LSD1 role in the regulation of neurodevelopment-related genes.

Different *modus operandi* can be prospected for the nsLSD1 splice isoform: the inclusion of E8a may provide a tool to tune LSD1 epigenetic activity on targets that are shared by all the splice variants, as they are contemporarily expressed in the nervous system. In such case, a distinction between neurospecific and ubiquitous isoforms may arise from posttranslational modifications on exon E8a generating a loop protruding from the surface of the protein, as revealed by the structural characterization. Notably, computational analysis of putative LSD1 phosphorylation sites predicted a high score threonine residue within the neuro-specific exon which *in vivo* proved susceptible of post-translational modification at early

postnatal stages, according to mass spectrometry analysis of the immunoprecipitated LSD1 neuropeptides. Results from luciferase reporter assay and morphologic analysis on cortical neurons upon preferential expression of the modified rather than unmodified isoform clearly indicate in the phosphorylation of the neuro-specific exon a mechanistic basis for functional diversification of LSD1 variants in the nervous system: when compared to the wtLSD1 isoform, phosphorylated nsLSD1 relieves the reporter gene from repression, which conversely can be rescued by the unphosphorylated counterpart suggesting that, while ubiquitous isoforms repress neuronal genes by default (Shi et al.2004), the modified nsLSD1 would relieve them from repression at specific developmental stages and the extent at which phosphorylation occurs would determine the final regulatory outcome. Noteworthy is the fact that no statistic difference emerges from the transfection of the native nsLSD1 compared to the phosphorylated variant, suggesting that most of native nsLSD1 is likely to undergo phosphorylation *in vivo,* at least at the considered developmental stages (PN1 in rat cortex and div4 in cortical neurons), possibly turning LSD1 into a sensor of environmental cues and a platform of converging signaling pathways. A similar scenario has already been described for MeCP2 in adult cortical neurons, where a depolarizing treatment induces a massive phosphorylation of the CpG binding protein causing its dissociation from target genes and their repression relief by chromatin remodeling (Ballas et al.2005, Zhou et al.2006). Moreover, since the four LSD1 isoforms can be expressed within the same histotype where they variably interact, a combinatorial assembly into a common CoREST complex may occur, widening the *repertoire* of the co-repressor complexes: the transcriptional regulation would rely on the frequency at which LSD1 splicing and phosphorylation occur, providing a double layered control to modulate its function in the brain. Notably, the assembly of neuro-specific factors in different combination has already been described (Olave et al., 2002) increasing the complexity of epigenetic regulation within CNS (Lessard et al.2007, Wu et al., 2007).

Another *modus operandi* can be prospected for LSD1 splice variants, by assuming that promoters occupancy may rely on the interactivity with different molecular partners, which is likely considering that no DNA binding domain has been structurally defined for LSD1 so far (Forneris et al. 2005). Experimental data demonstrated that the inclusion of neurospecific E8a within the amine oxidase domain does not preclude LSD1 binding to histone peptides and recombinant LSD1 variants retain comparable demethylase activity if tested in the presence of the C-terminus of CoREST which acts as a stabilizing

factor in vitro and is essential to prevent protein degradation in vivo (Shi et al. 2004), suggesting that a reduced availability of or affinity for CoREST might affect LSD1 isoforms activity. Insightful data also come from lincRNA *Hotair* which is mainly expressed in primary foreskin fibroblasts where it acts as bimodular scaffold recruiting polycomb repressive complex 2 at 5'terminal and LSD1 at 3' terminal. Upon *Hotair* interference LSD1 occupancy at *HOXD* proximal promoter is lost (Tsai et al. 2010), suggesting that specific lincRNAs could potentially direct complex patterns of chromatin states at specific genes in a spatially and temporally organized manner during development and disease states (Guttman et al. 2009). Consistently, in neurons the ubiquitous LSD1 variants cannot recapitulate the neurite traits elicited by the neuro-specific counterparts regardless of their phosphorylation state, indicating that the genes regulated by nsLSD1 may differ indeed from those regulated by the ubiquitous isoforms: the early embryonic stage at which the neurospecific variants can be detected, their association with a terminally differentiated neuro-restricted phenotype and their morphogenic role on neurites may possibly underlie the regulation of specific genes related to the neuronal commitment and the exit from the cell cycle as well as neuronal maturation. Notably, telencephalic tissues from E13.5 embryos, stage at which nLSD1 can be detected, have been considered a primary source of early neuronal and glial restricted progenitors (Maric et al. 2003). To test this hypothesis I performed LSD1 chromatin immunoprecipitation followed by high throughput sequencing on rat cortical neurons. ChIP-seq was done at e18.5+div7, when the neurospecific variant proved to be highly expressed by means of an antibody recognizing all the four isoforms providing the overall LSD1 genome wide location in the CNS, with a major contribution from the neuro-specific LSD1 isoform, according to the *in vitro* developmental stage. Even though a comparison of the obtained targets upon overexpression of the different tagged LSD1 variants may promptly provide verification of this hypothesis, limiting quantities of endogenous interactors permissive to LSD1 promoters binding (Tsai et al.2010) may lead to false negative results, whereas a subtractive approach would be resilient to such bias. Therefore I pursued a subtractive approach to infer gene sets specifically related to the LSD1 neurospecific isoforms, comparing a mock sample where all the four variants are expressed with a knocked sample where either the neurospecific or the ubiquitous variants are selectively silenced by means of lentiviral- mediated shRNAs delivery.

Despite all the ChIP fragments are expected to be called over threshold, the two experimental conditions should differ for the enrichment of the related sequenced tags depending on the contributing LSD1

isoform being immunoprecipitated: RNA interference of a specific LSD1 variant would determine a depletion of the corresponding fragments rather than a complete loss, depending on the efficacy of the silencing. Conversely, *peaks calling* for the cognate LSD1 isoforms wouldn't be affected, as their expression would be unaltered relative to the mock condition. While further optimization is still required to achieve lentiviral- mediated silencing and distinction of targets associated to the different splice variants, ChIP-sequencing performed in control conditions led so far to an extensive characterization of genes that are complexively regulated by all LSD1 splice variants expressed in developing cortical neurons, providing 11 million reads, 94% of which uniquely mappable to genomic positions resulting in 6700 called peaks, 10% of which annotated to promoters, residing at less than 1000 bp from a referenced TSS, where the sharp profile characterizing the average tags distribution clearly indicates a direct role for LSD1 in promoters regulation. Nonetheless, given that the peaks assignment to gene categories relies on the proximity to the nearest reference gene and a high proportion of coding sequences in the rat genome are provisional or incomplete, most of peaks are necessarily annotated as intergenic, possibly leading to an underestimation of LSD1 regulated promoters. Moreover, a wide-range regulatory role for LSD1 rather than a promoter –restricted function can be hypothesized for those genes where multiple binding locations can be found at short and long distance from the same TSS, with the distant peaks matching regions genomically conserved and devoid of ESTs, making improbable that such binding locations may relate to other un-annotated promoters. Therefore I considered at first all the annotated peaks for downstream functional analysis and opted subsequently for a more conservative approach on the selected subset of promoters. From gene ontology categorization several classes were found involved in developmental processes and control of cell fate decisions, regulation of cell cycle, calcium induced neuronal signalling and inherent categories were found for TFs whose binding sites are predicted to recur in the LSD1 identified promoters. Interestingly, for some of the genes for which LSD1 displays multiple binding locations short and long distance to the TSS, an altered expression has been reported upon long-lasting synaptic enhancement (Kawaai et al. 2010, Greer et al. 2008, Aizawa et al. 2004, Pfenning et al. 2010), suggesting that LSD1 cooperative binding at promoters regions and putative enhancer sites may occur. Notably, widespread transcription at neuronal enhancers and coding regions upon depolarization has already been described, relating eRNAs synthesis to K4H3-1me levels (Kim et al. 2010). Considering that enhancers recruit the transcriptional co-activator p300/CBP, bind histone H3 monomethylated at

lysine 4 and that they are located distally from known transcription start sites (Heintzman et al. 2007, Visel et al. 2009, Xi et al. 2007), the definition of of LSD1/p300 co-occurrent binding patterns in cortical neurons might provide insightful hints on a possible K4H3 histone demethylase role in the regulation of enhancer activity.

To conclude, the present study provides evidence that the time-lapsed expression of the neurospecific LSD1 splice isoforms contributes to the early neurite morphogenesis of cortical neurons where all the splice variants are compresent and that phosphorylation at exon E8a constitutes a mechanistic basis for their functional specialization in the CNS, where the frequency at which LSD1 splicing and phosphorylation occur would provide a double layered control to modulate its activity, determining the final regulatory outcome on target genes. Future directions may focus on the identification of the kinase responsible for nsLSD1 phosphorylation in vivo and its molecular interactors by mass –spectrometry while further efforts will allow the selective identification of target genes that are uniquely associated to the nsLSD1 variant through its exclusive genome-wide location.

BIBLIOGRAPHY

Abad MA., Enguita M., DeGregorio-Rocasolano N., Ferrer I., Trullas R. (2006), Neuronal Pentraxin 1 Contributes to the Neuronal Damage Evoked by Amyloid-ß and Is Overexpressed in Dystrophic Neurites in Alzheimer's Brain. J. Neurosci., 26: 12735 – 12747

Aizawa, H., Hu, S.C., Bobb, K., Balakrishnan, K., Ince, G., Gurevich, I., Cowan,M., and Ghosh, A. (2004). Dendrite development regulated by CREST,a calcium-regulated transcriptional activator. Science 303, 197–202.

Allis CD, Berger SL, Cote J, Dent S, Jenuwien T, Kouzarides T, Pillus L,Reinberg D, Shi Y, Shiekhattar R, Shilatifard A, Workman J, Zhang Y (2007) New nomenclature for chromatin-modifying enzymes. Cell 131:633– 636.

Andre´s ME, Burger C, Peral-Rubio MJ, Battaglioli E, Anderson ME, Grimes J,Dallman J, Ballas N, Mandel G (1999) CoREST: a functional corepressor required for regulation of neural-specific gene expression. Proc Natl Acad Sci U S A 96:9873–9878.

Bai L, Morozov AV (2010) Gene regulation by nucleosome positioning. Trends Genet. 26,476-83

Ballas N, Battaglioli E, Atouf F, Andres ME, Chenoweth J, Anderson ME, Burger C, Moniwa M, Davie JR, Bowers WJ, Federoff HJ, Rose DW, Rosenfeld MG, Brehm P, MandelG (2001) Regulation of neuronal traits by a novel transcriptional complex. Neuron 31:353–365.

Ballas N, Grunseich C, Lu DD, Speh JC, Mandel G (2005) REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. Cell 121:645– 657.

Barak O, Lazzaro MA, Cooch NS, Picketts DJ, Shiekhattar R (2004) A tissue specific, naturally occurring human SNF2L variant inactivates chromatin remodeling. J Biol Chem 279:45130–45138.

Battaglioli E, Andre´s ME, RoseDW,Chenoweth JG, Rosenfeld MG, Anderson ME, Mandel G. (2002) REST repression of neuronal genes requires components of the hSWI.SNF complex. J Biol Chem 277:41038–41045.

Binda C, Valente S, Romanenghi M, Pilotto S, Cirilli R, Karytinos A, Ciossani G, Botrugno OA, Forneris F, Tardugno M, Edmondson DE, Minucci S, Mattevi A, Mai A (2010) Biochemical, structural and biological evaluation of tranylcypromine derivatives as inhibitors of LSD1 and LSD2. J Am Chem Soc., 132, 6827-33.

Borrelli E, Nestler EJ, Allis CD, Sassone-Corsi P (2008) Decoding the epigenetic language of neuronal plasticity. Neuron 60:961–974.

Cai Q., Robertson ES. (2010) Ubiquitin/SUMO modification regulates VHL protein stability and nucleocytoplasmic localization. PLoS One. 5, e12636.1-9

Catharius K, Frech K, Groet K, Klocke B, Haltmeier M, Klingenhoff. A, Frisch M, Bayerlein M, Werner T (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics. 21,2933-42

Chang B. , Chen Y., Zhao Y.,Bruick R. (2007) JMJD6 is a histone arginine demethylase.Science 318, 444–447

Chen G, Nguyen PH, Courey AJ (1998) A role for Groucho tetramerization in transcriptional repression. Mol Cell Biol 18:7259 – 7268.

Chen Y, Yang Y, Wang F, Wan K, Yamane K, Zhang Y, LeiM (2006) Crystal structure of human histone lysine-specific demethylase 1 (LSD1). Proc Natl Acad Sci U S A 103:13956 –13961.

Choi J., Jang H,, Kim H., Kim S.T., Cho E.J., Youn H.D. (2010) Histone demethylase LSD1 is required to induce skeletal muscle differentiationby regulating myogenic factors.Biochem Biophys. Res. Comm. 401, 327-332

Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: programs for protein crystallography. Acta Crystallogr D 50:760 –763.

Dallman JE, Allopenna J, Bassett A, Travers A, Mandel G (2004) A conserved role but different partners for the transcriptional corepressor CoREST in fly and mammalian nervous system formation. J Neurosci
24:7186 –7193.

Dunn MJ., Crisp SJ.(1994) Detection of proteins in polyacrylamide gels using an ultrasensitive silver staining technique. Methods Mol Biol., 32:113-8.

Dulac C. (2010). Brain function and chromatin plasticity. Nature 465, 728-735

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D 60:2126 –2132.

Fang R, Barbera AJ, Xu Y, Rutenberg M, Leonor T, Bi Q, Lan F, Mei P, Yuan GC, Lian C, Peng J, Cheng D, Sui G, Kaiser UB, Shi Y, Shi YG.(2010). Human LSD2/KDM1b/AOF1 regulates gene transcription by modulating intragenic H3K4me2 methylation. Mol Cell.,39,222-33

Ferragina G.,Manzini P. (2000) Opportunistic Data Structures with Applications. Proceedings of FOCS, 41st IEEE Symposium on Foundations of Computer Science,390 – 398.

Flavell, S.W., and Greenberg, M.E. (2008). Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. Annu. Rev. Neurosci. 31, 563–590

Forneris F, Binda C, Vanoni MA, Mattevi A, Battaglioli E (2005a) Histone demethylation catalysed by LSD1 is a flavin-dependent oxidative process. FEBS Lett 579:2203–2207.

Forneris F, Binda C, Vanoni MA, Battaglioli E, Mattevi A (2005b) Human histone demethylase LSD1 reads the histone code. J Biol Chem 280:41360–41365.

Forneris F, Binda C, Dall'Aglio A, Fraaije MW, Battaglioli E, Mattevi A (2006) A highly specific mechanism of histone H3–K4 recognition by histone demethylase LSD1. J Biol Chem 281:35289 –35295.

Forneris F, Binda C, Adamo A, Battaglioli E, Mattevi A (2007) Structural basis of LSD1-CoREST selectivity in histone H3 recognition. J Biol Chem 282:20070 –20074.

Forneris F, Binda C, Battaglioli E, Mattevi A (2008) LSD1: oxidative chemistry for multifaceted functions in chromatin regulation. Trends Biochem Sci 33:181–189.

Foster C.T., Dovey O. Lezina L. Luo J.L., Gant T.W. ,Barlev N., Bradley A., Cowley S.(2010). Lysine specific demethylase 1 regulates the embryonic transcriptome and CoREST stability.Mol Cell. Biol.,30, 4851-63

Fox IJ, Paucar AA, Nakano I, Mottahedeh J, Dougherty JD, Kornblum HI (2004) Developmental expression of glial fibrillary acidic protein Mrna in mouse forebrain germinal zones: implications for stem cell biology. Brain Res Dev Brain Res 153:121–125.

Garcia-Bassets I, Kwon YS, Telese F, Prefontaine GG, Hutt KR, Cheng CS, Ju BG, Ohgi KA, Wang J, Escoubet-Lozach L, Rose DW, Glass CK, Fu XD, Rosenfeld MG (2007) Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. Cell 128:505–518.

Garriga-Canut M, Schoenike B, Qazi R, Bergendahl K, Daley TJ, Pfender RM, Morrison JF, Ockuly J, Stafstrom C, Sutula T, Roopra A (2006) 2-Deoxy-D-glucose reduces epilepsy progression by NRSF-CtBPdependent metabolic regulation of chromatin structure. Nat Neurosci 9:1382–1387.

Gatta R, Mantovani R (2008) NF-Y substitutes H2A-H2B on active cell cycle promoters: recruitment of CoREST-KDM1 and fine-tuning of H3 methylations. Nucleic Acids Res 36:6592– 6607.

Gaudilliere, B., Konishi, Y., de la Iglesia, N., Yao, G., and Bonni, A. (2004).A CaMKII-NeuroD signaling pathway specifies dendritic morphogenesis.Neuron 41, 229–241.

Grabowski PJ, Black DL (2001) Alternative RNA splicing in the nervous system. Prog Neurobiol 65:289 –308.

Greer PL, Greenberg ME.,(2008).From synapse to nucleus: calcium-dependent gene transcription in the control of synapse development and function. Neuron.59:846-60.

Gutierrez H., Davies AM. (2007). A fast and accurate procedure for deriving the Sholl profile in quantitative studies of neuronal morphology. Neurosci Methods. *163:24-30*

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES.(2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNA in mammals. Nature,458, 223-7

Hakimi MA, Bochar DA, Chenoweth J, Lane WS, Mandel G, Shiekhattar R (2002) A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. Proc Natl Acad Sci U S A 99:7420 –7425.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B.(2007)Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet.39, 311-8

Hua, J.Y., and Smith, S.J. (2004). Neural activity and the dynamics of central nervous system development. Nat. Neurosci. 7, 327–332.

Humphrey GW, Wang Y, Russanova VR, Hirai T, Qin J, Nakatani Y, Howard BH (2001) Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. J Biol Chem 276:6817– 6824.

Jin L, Kryukov K, Clemente JC, Komiyama T, Suzuki Y, Imanishi T, Ikeo K, Gojobori T. (2008)The evolutionary relationship between gene duplication and alternative splicing. Gene, 427,19-31

Jugloff DG, Jung BP, Purushotham D, Logan R, Eubanks JH (2005) Increased dendritic complexity and axonal length in cultured mouse cortical neurons overexpressing methyl-CpG-binding protein MeCP2.Neurobiol Dis 19:18 –27.

Kahl P, Gullotti L, Heukamp LC, Wolf S, Friedrichs N, Vorreuther R, Solleder G, Bastian PJ, Ellinger J, Metzger E, Schüle R, Buettner R. (2006). Androgen receptor coactivators lysine-specific histone demethylase 1 and four and a half LIM domain protein 2 predict risk of prostate cancer recurrence. Cancer Res.,66,11341-7

Karytinos A, Forneris F, Profumo A, Ciossani G, Battaglioli E, Binda C, MatteviA (2009) A novel mammalian flavin-dependent histone demethylase. J Biol Chem 284:17775–17782.

Kawaai K., Yoshino T.K., Urakubo T., Taniguchi N., Kondoh Y.,Tashiro H.,Ogura A.,Tashiro T.(2010) Analysis of Gene Expression Changes Associated With Long-Lasting Synaptic Enhancement in Hippocampal Slice Cultures After Repetitive Exposures to Glutamate. J. Neuros. Res. 88:2911–2922

Kim A.H., Bonni A. (2008) Cdk1-FOXO1: a mitotic signal taked center stage in post-mitotic neurons. Cell Cycle, 7, 3819-22

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME.(2010) Widespread transcription at neuronal activity-regulated enhancers. Nature 465: 173-4

Kopelman NM, Lancet D, Yanai I (2005) Alternative splicing and gene duplication are inversely correlated evolutionary mechanisms. Nat Genet 37:588 –589.

Kouzarides T (2007) Chromatin modifications and their function. Cell 128:693–705

Langmead B., Trapnell C. Pop M., Salzberg SL.(2009)Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 2009, 10:R25.1-25.9

Lakowski B, Roelens I, Jacob S (2006) CoREST-like complexes regulate chromatin modification and neuronal gene expression. J Mol Neurosci 29:227–239.

Lee H, Dean C, Isacoff E.(2010). Alternative splicing of neuroligin regulates the rate of presynaptic differentiation. J Neurosci., 30, 11435-46

Lee MG, Wynder C, Cooch N, Shiekhattar R (2005) An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature 437:432– 435.

Lee SH, Sheng M (2000) Development of neuron-neuron synapses. Curr Opin Neurobiol 10:125–131.

Lemercier C., Legube GI., Caron C., Louwagie M., Garin J., Trouche D., Khochbin S. (2003). Tip60 Acetyltransferase Activity Is Controlled by Phosphorylation. JBC 278, 4713-18

Leslie AG (1999) Integration of macromolecular diffraction data. Acta Crystallogr D Biol Crystallogr 55:1696 –1702.

Lessard J, Wu JI, Ranish JA, Wan M, Winslow MM, Staahl BT, Wu H, Aebersold R, Graef IA, Crabtree GR (2007) An essential switch in subunit composition of a chromatin remodeling complex during neural development.Neuron 55:201–215.

Licatalosi DD, Darnell RB (2006) Splicing regulation in neurologic disease. Neuron 52:93–101.

Lin L, Shen S, Jiang P, Sato S, Davidson BL, Xing Y. (2010)Evolution of alternative splicing in primate brain transcriptomes.Hum Mol Genet. ,19,2958-73

Lipscombe D (2005) Neuronal proteins custom designed by alternative splicing. Curr Opin Neurobiol 15:358 –363.

Lunyak VV, Burgess R, Prefontaine GG, Nelson C, Sze SH, Chenoweth J, Schwartz P, Pevzner PA, Glass C, Mandel G, Rosenfeld MG.(2002). Co-repressor dependent silencing of chromosomal regions encoding neuronal genes, Science 298,1747-52

Maric D., Maric I.,Chang YH.,Barker JL.(2003)Prospetive cell sortingof embryonic rat neural stem cells and neuronal and glial progenitors reveals selective effects of basic fibroblast growth factor and epidermal growth factor on self-renewal and differentiation.J.Neurosci. 23,240-251

McAllister AK (2000) Cellular and molecular mechanisms of dendrite growth. Cereb Cortex 10:963–973

Metzger E, Imhof A, Patel D, Kahl P, Hoffmeyer K, Friedrichs N, Müller JM, Greschik H, Kirfel J, Ji S, Kunowska N, Beisenherz-Huss C, Günther T, Buettner R, Schüle R.(2010). Phosphorylation oh histone H3T6 by PKC beta(I) controls demethylation at histone H3K4.Nature, 464,792-6

Metzger E, Wissmann M, Yin N, Müller JM, Schneider R, Peters AH, Günther T, Buettner R, Schüle R.(2005). LSD1 demethylates repressive histone marks to promote androgen-receprot-dependent transcription. Nature, 437,436-9

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D 53:240 –255.

Musri M., Carmona M.C., Hanzu F., Kalian P., Gomis R., Parrizas M. (2010) Histone demethylase LSD1 regulates adipogenesis. JBC, 285,30034-41

Nadeau H., Lester H.A.,(2002) RSF Causes cAMP-Sensitive Suppression of Sodium Current in Cultured Hippocampal Neurons. *J* Neurophysiol 88: 409-421

Needleman SB., Wunsch CD. (1970). A general method applicable to the search for similarities in the amino acid sequences of two proteins. Journal of Mol Biol 48, 443–53

Nott A., Watson PM., Robinson JD., Crepaldi L., Riccio A. (2008)S-nitrosylation of histone deacetylase 2 induces chromatin remodelling in neurons. Nature 455, 411-415

Nottke A, Colaia´covo MP, Shi Y (2009) Developmental roles of the histone lysine demethylases. Development 136:879–889.

Olave I, Wang W, Xue Y, Kuo A, Crabtree GR (2002) Identification of a polymorphic, neuron-specific chromatin remodeling complex. Genes Dev 16:2509 –2517.

Palm K, Belluardo N, Metsis M, Timmusk T (1998) Neuronal expression of zinc finger transcription factor REST/NRSF/XBR gene. J Neurosci 18:1280 –1296.

Palm K.Metsis M., Timmusk T. (1999) Neuron-specific splicing of zinc finger transcription factor REST/NRSF/XBR is frequent in neuroblastomas and conserved in human, mouse and rat. Molecular Brain Research 72 , 30–39

Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H, Hurst LD (2007) Splicing and the evolution of proteins in mammals. PLoS Biol 5:e14.

Pfenning A.R., Kim T.K., Spotts J.M., Hemberg M., Su D., West A.E. West (2010) Genome-Wide Identification of Calcium-Response Factor(CaRF) Binding Sites Predicts a Role in Regulation of Neuronal Signaling Pathways. Plos One 5,1-12

Qian X, Shen Q, Goderie SK, He W, Capela A, Davis AA, Temple S (2000) Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. Neuron 28:69–80.

Reynolds A, Leake D, Boese Q, Scaringe S, Marshall WS, Khvorova A.(2004).Rational siRNA design for RNA interference. Nat Biotechnol.22:326-30

Rizk G. , Lavenier D. (2010). GASSST: global alignment short sequence search tool. Bioinformatics. 26: 2534–2540.

Romorini S, Piccoli G, Jiang M, Grossano P, Tonna N, Passafaro M, Zhang M, Sala C (2004) A functional role of postsynaptic density-95-guanylate kinase-associated protein complex in regulating Shank assembly and stability to synapses. J Neurosci 24:9391–9404.

Sala C, Futai K, Yamamoto K, Worley PF, Hayashi Y, Sheng M. (2003). nhibition of dendritic spine morphogenesis and synaptic transmission by activity-inducible protein Homer1a**.** J Neurosci.,23,6327-37.

Sala C, Rudolph-Correia S, Sheng M (2000) Developmentally regulated NMDA receptor-dependent dephosphorylation of cAMP response element-binding protein (CREB) in hippocampal neurons. J Neurosci 20:3529 –3536.

Saleque, S., J. Kim, H. M. Rooke, and S. H. Orkin. (2007) Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1. Mol. Cell 27:562–572

Shalizi A, Bilimoria PM, Stegmüller J, Gaudillière B, Yang Y, Shuai K, Bonni A (2007) PIASx is a MEF2 SUMO E3 ligase that promotes postsynaptic dendritic morphogenesis. J Neurosci. ,27,10037-46.

Shimojo M, Paquette AJ, Anderson DJ, Hersh LB (1999) Protein kinase A regulates cholinergic gene expression in PC12 cells: REST4 silences the silencing activity of neuron-restrictive silencer factor/REST. Mol Cell Biol 19:6788–6795.

Shi Y (2007) Histone lysine demethylases: emerging roles in development, physiology and disease. Nat Rev Genet 8:829–833. Shi Y, Sawada J, Sui G, Affar el B, Whetstine JR, Lan F, Ogawa H, Luke MP, Nakatani Y, Shi Y (2003) Coordinated histone modifications mediated by a CtBP co-repressor complex. Nature 422:735–738.

Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119:941–953.

Shi YJ, Matson C, Lan F, Iwase S, Baba T, Shi Y (2005) Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell 19:857– 864.

Sholl DA (1953) Dendritic organization in the neurone of the visual motor cortices of the cat. J Anat 87:387– 406.

Schulte JH, Lim S, Schramm A, Friedrichs N, Koster J, Versteeg R, Ora I, Pajtler K, Klein-Hitpass L, Kuhfittig-Kulle S, Metzger E, Schüle R, Eggert A, Buettner R, Kirfel J.(2009).Lysine-specific demethylase 1 is strongly expressed in poorly differentiated neuroblastoma: implications for therapy**.** Cancer Res. ,69,2065-71

Sorek R, Ast G (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. Genome Res 13:1631–1637.

Stavropoulos P, Blobel G, Hoelz A (2006) Crystal structure and mechanism of human lysine-specific demethylase-1. Nat Struct Mol Biol 13:626–632.

Stead JD, Neal C, Meng F, Wang Y, Evans S, Vazquez DM, Watson SJ, Akil H (2006) Transcriptional profiling of the developing rat brain reveals that the most dramatic regional differentiation in gene expression occurs postpartum. J Neurosci 26:345–353.

Sui SJ, Mortimer JR, Arenillas DJ, Brumm J, Walsh CJ, Kennedy BP, Wasserman WW.(2005) oPOSSUM: identification of over-represented transcription factor binding sites in co-expressed genes. Nucleic Acids Res. 2005 Jun 2;33(10):3154-64. Print 2005.

Talavera D, Vogel C, Orozco M, Teichmann SA, de la Cruz X.(2007) The (in)dependence of alternative splicing and gene duplication. PLoS Comput Biol. ,3,

Tartaglia N, Du J, Tyler WJ, Neale E, Pozzo-Miller L, Lu B (2001) Protein synthesis-dependent and -independent regulation of hippocampal synapses by brain-derived neurotrophic factor. J Biol Chem 276:37585– 37593.

Tsai MC, Manor O., Wan Y.,Mosammaparast N., Wang JK, Lan F.,Shi Y.,Segal E.,Chang HI.(2010) Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. Science,329, 689-693

Tsankov AM, Thompson DA, Socha A, Regev A, Rando OJ. (2010)The Role of Nucleosome Positioning in the Evolution of Gene Regulation PLoS Biol., 8, 1-20

Ule J, Darnell RB (2006) RNA binding proteins and the regulation of neuronal synaptic plasticity. Curr Opin Neurobiol 16:102–110.

van Essen D, Zhu Y, Saccani S.(2010). A feed-forward circuit controlling inducible NF-κB target gene activation by promoter histone demethylation. Mol Cell, 39,750-60

Visel A.,Blow. M.J., Li Z.,Zhang T.,Akiyama J.A., Holt A., Frick I.P.,Shoukry M., Wright C., Chen F., Afzal V., Ren B., Rubin E.M., Pennacchio L.A.(2009) Chip-seq accurately predicts tissue-specific activity of enhancers. Nature, 457,854-860

Walton M, Henderson C, Mason-Parker S, Lawlor P, Abraham WC, Bilkey D, Dragunow M (1999) Immediate early gene transcription and synaptic modulation. J Neurosci Res 58:96 –106.

Wang, J., S. Hevi, J. K. Kurash, H. Lei, F. Gay, J. Bajko, H. Su, W. Sun, H.Chang, G. Xu, F. Gaudet, E. Li, and T. Chen. 2009. The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation. Nat.Genet. 41:125–129.

Wang J, Scully K, Zhu X, Cai L, Zhang J, Prefontaine GG, Krones A, Ohgi KA, Zhu P, Garcia-Bassets I, Liu F, Taylor H, Lozach J, Jayes FL, Korach KS, Glass CK, Fu XD, Rosenfeld MG (2007) Opposing LSD1 complexes function in developmental gene activation and repression programmes.Nature 446:882– 887.

Wang, Y., H. Zhang, Y. Chen, Y. Sun, F. Yang, W. Yu, J. Liang, L. Sun, X. Yang, L. Shi, R. Li, Y. Li, Y. Zhang, Q. Li, X. Yi, and Y. Shang.(2009). LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. Cell 138:660–672

Webster N., Jin JR.,Green S.,Hollis M.,Chambon P. (1998) The yeast UASg is a transcriptional enhancer in human HeLA cells in the presence of Gal4 trans-activator. Cell 52,169-178

Wissmann M, Yin N, Müller JM, Greschik H, Fodor BD, Jenuwein T, Vogler C, Schneider R, Günther T, Buettner R, Metzger E, Schüle R.(2007). Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression.Nat Cell Biol.9,347-53

Wiznerowic M., Trono D.(2003). Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. J.Virol. 77:8957-8961

Wolfertstetter F, Frech K, Herrmann G, Werner T (1996)Identification of functional elements in unaligned nucleic acid sequences by a novel tuple search algorithm. Comput Appl Biosci.12:71-80.

Wu JI, Lessard J, Olave IA, Qiu Z, Ghosh A, Graef IA, Crabtree GR (2007) Regulation of dendritic development by neuron-specific chromatin remodeling complexes. Neuron 56:94 –108.

Wu JI, Lessard J, Crabtree GR (2009) Understanding the words of chromatin regulation. Cell 136:200 –206.

Xia Z, Dudek H, Miranti CK, Greenberg ME (1996) Calcium influx via the NMDA receptorinduces immediate early gene transcription by a MAP kinase/ERK-dependent mechanism. JNeurosci 16:5425-5436

Xi H, Shulha JP, Lin JM, Vales TR,Bodine DM,McKay RD,Chenoweth JG, Tesar PJ,Furey TS, Ren B, Weng Z, Crawford GE (2007) Identification and characterization of cell type-specific and ubiquitous chromatin regulatory structures in the human genome.Plos Genetics 3,1377-1388

Xing Y, Wang Q, Lee C (2006) Evolutionary divergence of exon flanks: a dissection of mutability and selection. Genetics 173:1787–1791.

Yang M, Gocke CB, Luo X, Borek D, Tomchick DR, Machius M, Otwinowski Z, Yu H (2006) Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell 23:377–387.

You A, Tong JK, Grozinger CM, Schreiber SL (2001) CoREST is an integral component of the CoREST-human histone deacetylase complex. Proc Natl Acad Sci U S A 98:1454 –1458.

Zhou Z, Hong EJ, Cohen S, Zhao WN, Ho HY, Schmidt L, Chen WG, Lin Y, Savner E, Griffith EC, Hu L, Steen JA, Weitz CJ, Greenberg ME.(2006)Brain –specific phosphorylation of MeCP2 regulates activity-dependent BDNF transcription,dendritic growth and spine maturation. Neuron. 52:255-69.

Zibetti C., Adamo A. , Binda C., Forneris F., Toffolo M., Verpelli C., Ginelli E.,Mattevi A., Sala C Battaglioli E. (2010) Alternative Splicing Tunes LSD1/KDM1 Epigenetic Activity J. Neurosci., 30, 2521–2532

Zuccato C, Tartari M, Crotti A, Goffredo D, Valenza M, Conti L, Cataudella T, Leavitt BR, Hayden MR, Timmusk T, Rigamonti D, Cattaneo E (2003) Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. Nat Genet 35:76–83.

Zuffrey R., Nagy D.,Mandel RJ, Naldini L.,Trono D. (1997) Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. Nat.Biotech. 15:871-5.

Zufferey R, Dull T, Mandel RJ, Bukovsky A, Quiroz D, Naldini L, Trono D(1998) Self inactivating lentiviral vectors for safe and efficient in vivo gene delivery. J Virol.72:9873-80.