Nuclear DYRK1A: new insights to its role within the nucleus

Chiara Di Vona

TESI DOCTORAL UPF / 2013

Barcelona, June 2013

Nuclear DYRK1A: new insights to its role within the nucleus

Chiara Di Vona

Memòria presentada per optar al grau de Doctora per la Universitat Pompeu Fabra.

Aquesta tesi ha estat realitzada sota la direcció de la Dra. Susana de la Luna al Centre de Regulació Genòmica (CRG, Barcelona), dins del Programa de Genes i Malaltia.

Chiara DI Vona **Susana de la Luna**

Index

Abstract

The view on how protein kinases regulate gene expression have recently expanded to include not only transcription factors but also histones, chromatin remodelers or components of the basal transcription machinery, which are directly modified on genomic loci. For the shuttling kinase DYRK1A (dual-specificity tyrosine-regulated kinase), most of its nuclear-associated functions can be explained by DYRK1A cytosolic activities, questioning a role for DYRK1A within the nuclear compartment. In the present study, through an unbiased proteomic approach the first “DYRK1A nuclear interactome” have been generated. DYRK1A interacts with several components of the basal transcriptional machinery as well as with the pre-mRNA processing machinery. Moreover, evidences uncovering a new role for DYRK1A as a transcriptional regulator of specific target genes have been generated. Genome-wide DYRK1A-chromatin analysis shows that the kinase is recruited to RNA polymerase II proximal promoters, via a highly conserved palindromic sequence, and also to RNA polymerase III-dependent promoters. Growth-dependent induction of the expression of a subset of target genes (protein coding and tRNAs) depends on DYRK1A protein levels and/or activity. In addition, downregulation of DYRK1A leads to a reduction in cell size. DYRK1A could therefore work by sitting on promoters of specific genes and act on different components of the basal transcription and/or mRNA processing machinery to modulate gene expression.

Resultados recientes han puesto de manifiesto que la regulación de la expresión génica por proteína quinasas va mas allá de su modulación de la actividad de factores de transcripción, ya que tanto histonas como remodeladores de cromatina o componentes de la maquinaria basal de transcripción puedes ser sustratos de fosforilaciones directamente en regiones genómicas reguladoras. La proteína quinasa DYRK1A (dual-specificity tyrosine-regulated kinase) está presente tanto en el núcleo como en el citoplasma de células de mamífero, si bien la mayoría de sus actividades nucleares pueden ser explicadas por fosforilaciones que ocurren en el citosol, lo que ha planteado dudas sobre si esta quinasa posee funciones específicamente nucleares. En este trabajo, se ha definido el primer "interactoma" nuclear de DYRK1A mediante una aproximación proteómica no sesgada, que ha permitido mostrar que DYRK1A interacciona con componentes de la maquinaria basal de transcripción así como con complejos implicados en el procesamiento del pre-mRNA. Los resultados también han puesto de manifiesto un nuevo papel de DYRK1A como regulador de la transcripción de un grupo específico de genes relacionados con la traducción de proteínas. Análisis a nivel genómico de la presencia de DYRK1A en cromatina muestra que la quinasa es reclutada a regiones proximales de promotores dependientes de la RNA polimerasa II, mediante una secuencia palindrómica altamente conservada, así como a genes dependientes de la RNA polimerasa III. La inducción de la expresión de un grupo de estos genes diana (tanto codificantes como tRNAs) en respuesta a factores de crecimiento depende de DYRK1A. Además, la reducción en los niveles de DYRK1A provoca una reducción en el tamaño de las células. Los resultados permiten proponer un modelo por el que DYRK1A podría regular directamente la expresión de genes diana mediante la fosforilación, en regiones reguladoras promotoras, de diferentes componentes de la maquinaria basal de la transcripción y/o de los complejos proteicos implicados en el procesamiento del mRNA.

Introduction

The DYRK subfamily of protein kinases

DYRK1A belongs to a family of conserved protein kinases named dual-specificity tyrosine-phosphorylated and regulated kinases (DYRK), within the CMGC group of the eukaryote kinome, which also includes cyclin-dependent kinase (CDK), mitogen activated protein kinase (MAPK), glycogen synthase kinase (GSK), CDK-like kinase (CLK), serine-arginine-rich protein kinase, cdc2-like kinase, and RCK kinase families *(Manning et al., 2002)* (Fig. I1A).

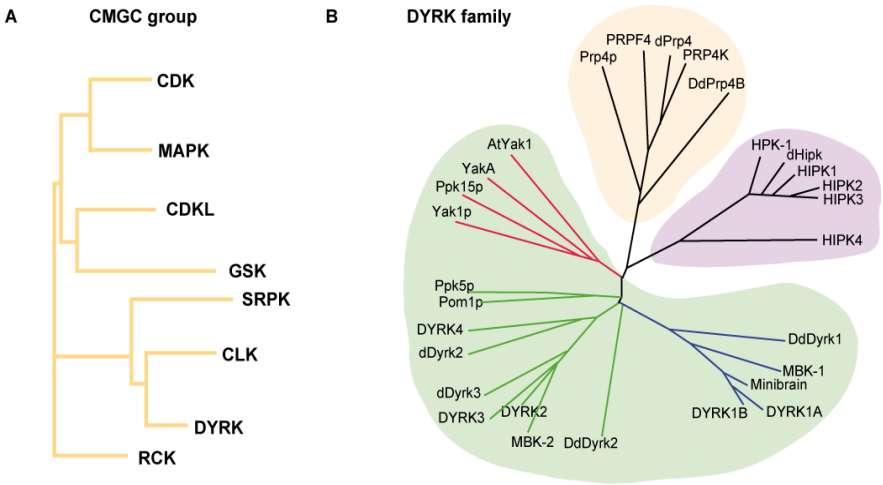


Figure I1: The DYRK family of proteins. **A)** Phylogenetic tree showing the evolutionary divergence of the different families within the CMGC group (see text for the name of the families). **B)** Unrooted phylogenetic tree showing the three subfamilies of the DYRK family, PRP4 (in orange), HIPK (in violet) and DYRK (in green) (adapted from *(Aranda et al., 2011)*.

Homology within the kinase domain allows for the division of the DYRK family into three subfamilies: DYRK kinases, homeodomain-interacting protein kinases (HIPKs) and pre-mRNA processing protein 4 kinases (PRP4s) (Fig. I1B). DYRK subfamily members are present all along the eukaryotic evolution and show a high degree of conservation *(Aranda et al., 2011)*. From a phylogenetic viewpoint, the subfamily can be classified into three classes: a first group includes kinases only present in yeast and plants such as *Saccharomyces cerevisiae* Yak1p or *Dictyostelium discoideum* YakA (red branch in Fig. I1B); a second class includes *Schizosaccharomyces pombe* Pom1p, *Caenorhabditis elegans* MBK-2, *Drosophila melanogaster* dDYRK2 and dDYRK3 (smell-impaired), and vertebrate DYRK2, DYRK3, and DYRK4 (green branch in Fig. I1B); and a third class includes *C. elegans* MBK-1, *D.*

melanogaster minibrain, and vertebrate DYRK1A and DYRK1B (blue branch in Fig. I1B). This thesis work is focused on the study of human DYRK1A. For this reason a review on this kinase follows. A recent review on all members of the DYRK subfamily can be found in *(Aranda et al., 2011)*.

Structure and mechanism of activation of DYRK kinases

Five members of the DYRK family are found in mammals: DYRK1A, DYRK1B, DYRK2, DYRK3 and DYRK4. The analysis of the genomic structure reveals that the five members arose by gene duplication events, which allows their classification into two subgroups: class I DYRKs, DYRK1A and its closest relative DYRK1B, and class II DYRKs, DYRK2, DYRK3 and DYRK4 (Fig. I2). At the protein level, conservation among the different member is restricted to the kinase domain and to a sequence in the N terminal part of the protein, named DYRK homology (DH)-box, which is characteristic of the subfamily *(Becker and Joost, 1999)*. The closest related DYRK1A-DYRK1B share additional homology outside the kinase domain in a nuclear localization signal (NLS) and a PEST motif within the N-terminal and C-terminal regions, respectively (Fig. I2). The class II DYRKs contain a functional domain in the N-terminal region known as NAPA-domain, which is required for catalytic activity *(Kinstrie et al., 2010)*. Evidence from structural data suggests that the DH-box and the NAPA-domain fold independently and are able to stabilize a catalytically active conformation through interactions with residues of the catalytic domain N-lobe *(Soundararajan et al., 2013)*.

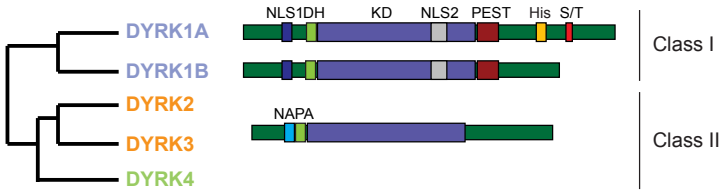


Figure I2: Protein structure and classification of human DYRKs. Several features can be distinguished within the primary structure of mammalian DYRKs: a central catalytic domain (KD), a functional bipartite nuclear localization signal (NLS1) at the non-catalytic N-terminus; a DYRK-homology box (DH); a complex nuclear localizations signal within the CMGC-extension in the catalytic domain (NLS2), a PEST region (PEST); a stretch of 13 consecutive His residues (His) and a region enriched in Ser and Thr residues (S/T). The two subgroups in which the kinases are classified are also indicated on the right. The functional NAPA-domain is only present in class II DYRKs.

DYRKs are known as dual-specificity kinases, because autophosphorylate on tyrosine and serine/threonine residues, while all the exogenous substrates described to date are phosphorylated in serine and threonine residues (an updated list is found in *(Aranda et al., 2011)*. In the case of DYRK1A, a consensus phosphorylation sequence was initially defined as RPX(S/T)P, suggesting that DYRK1A is a proline-directed kinase *(Himpel et al., 2000)*. However, the analysis of all the DYRK1A substrates indicates that not all their DYRK1A phosphorylation sites match perfectly the consensus *(Aranda et al., 2011)*. In fact, a larger screen of substrate recognition motifs for DYRK1A has shown that the kinase can accommodate peptides having smaller hydrophobic residues at the P+1 position (Pro, Ser, Ala, Val), with a preference for arginine at P-2, P-3 and P-4 (Fig. I3A) *(Soundararajan et al., 2013)*. Preference for P-3 arginine has been shown to be a feature of DYRK1A but not DYRK2 or DYRK4 that, together with the lack of overlap in the phosphosites identified in a substrate screen using peptides (Fig. I3B), suggest that substrate specificity exists within the DYRK subfamily members *(Papadopoulos et al., 2011)*.

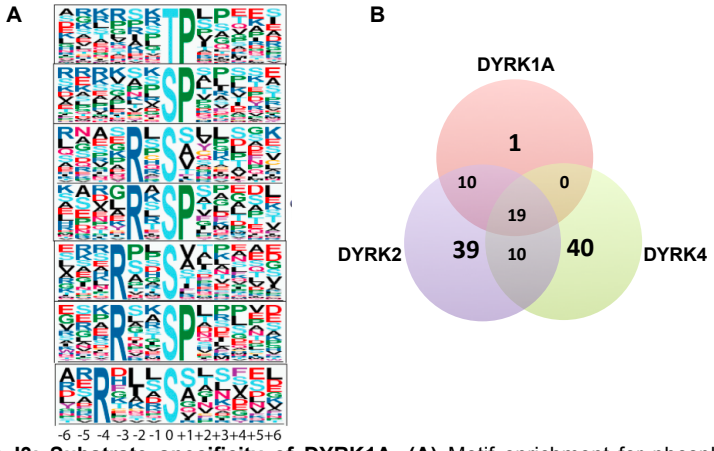


Figure I3: Substrate specificity of DYRK1A. (A) Motif enrichment for phosphosites found in a *in vitro* kinase substrate screen using naturally occurring substrates from HeLa cells *(Soundararajan et al., 2013)*. **(B)** Venn diagram showing the distribution of the peptides phosphorylated exclusively by different DYRK family members *(Papadopoulos et al., 2011)*.

All the DYRK family members, including DYRK1A, present a conserved Tyr-X-Tyr motif in the activation loop of the catalytic domain *(Becker and Joost, 1999)*. Phosphorylation of the second Tyr-residue of this motif is essential for DYRK1A kinase activity *(Himpel et al.,*

2001), a requisite that has been proven for other mammalian DYRKs *(Lee et al., 2000; Li et al., 2002; Papadopoulos et al., 2011)*, and other DYRK subfamilies as HIPKs *(Saul et al., 2013; Siepi et al., 2013)*. The mechanism of activation of DYRKs resembles that of the MAPKs, although, unlike MAPKs, Tyr-phosphorylation in the DYRK activation loop is an autophosphorylation event and does not involve an upstream activating kinase *(Himpel et al., 2001)*. The current model, based on experimental data from *Drosophila* DYRKs, suggests that Tyr-autophosphorylation within the activation loop is a "one-time only" intramolecular event that occurs while the protein is translated and that the tyrosine kinase activity is transient *(Lochhead et al., 2005)*. However, data with human DYRK kinases have shown that a non-phosphorylated enzyme retains some kinase activity *(Adayev et al., 2007)*, although it is very likely limited to tyrosine autophosphorylation events *(Himpel et al., 2000; Soundararajan et al., 2013)*.

Several DYRK family members have been shown to act as priming kinases for GSK3, both at position P-4 (classical GSK3 priming) and at distant sites from the GSK3 phosphorylated residue (non-classical or discontinuous GSK priming). For instance, DYRK1A acts as a classical GSK3 priming kinase in the case of eI2Fβ, tau, MAP1B or CRY2 *(Kurabayashi et al., 2010; Scales et al., 2009; Woods et al., 2001a)*. The list of GSK3-priming activities for DYRK2 includes glycogen synthase, eI2Fβ, tau, CRMP4, c-Jun and Myc *(Cole et al., 2006; Skurat and Dietrich, 2004; Taira et al., 2012; Woods et al., 2001a)*. Both DYRK1A and DYRK2 act as a non-classical GSK3 priming kinases for nuclear factor of activated T-cells (NFAT) transcription factors *(Arron et al., 2006; Gwack et al., 2006)*.

Several pharmacological inhibitors of DYRK1A have been developed in recent years including, among others, indirubin derivatives, roscovitine derivatives, leucettines or the benzothiazole derivative INDY *(Demange et al., 2013; Myrianthopoulos et al., 2013; Ogawa et al., 2010)*, although only two of them have been used to study the function of this kinase: (-)-epigallocatechin-3-gallate (EGCG), one of the major polyphenolic components of green tea, and the alkaloid harmine, a classical inhibitor of monoamine oxidase *(Bain et al., 2007)*. EGCG is a potent DYRK1A inhibitor and functions as a noncompetitive inhibitor of ATP, but although it has been used *in vivo* as a DYRK1A inhibitor *(Guedj et al., 2009)*, its value to explore DYRK1A functions is limited because inhibits other DYRKs and many other kinases *(Bain et*

al., 2007). Harmine is a potent ATP competitive DYRK1A inhibitor *(Laguna et al., 2008)* and it has been used to elucidate the *in vivo* functionality of DYRK1A *(Kuhn et al., 2009; Laguna et al., 2008; Pozo et al., 2013; Seifert and Clarke, 2009)* However, as harmine also inhibits DYRK1B (5-fold less efficiently) and DYRK2 and DYRK3 (50-fold less efficiently) *(Gockler et al., 2009)*, its use must be complemented with additional techniques.

The protein kinase DYRK1A

DYRK1A maps in the human chromosome 21 specifically in the region 21q22.2 *(Guimera et al., 1999; Song et al., 1996)*. It has been proposed as a candidate gene responsible for some of the pathogenic phenotypes of Down syndrome (DS) individuals *(Epstein, 2006)*, in particular mental retardation and early onset of Alzheimer's disease, based on the phenotypes of DYRK1A overexpression mouse models *(Ahn et al., 2006; Altafaj et al., 2001; Branchi et al., 2004; Ryoo et al., 2008)*. DYRK1A gene is indeed overexpressed in DS embryonic brain at the RNA level *(Guimera et al., 1999)*, and in adult brain at the protein level at 1.5 fold *(Dowjat et al., 2007; Wegiel et al., 2004)*. Only one DYRK1A null mice has been generated *(Fotaki et al., 2002)*, with no conditional models available. DYRK1A null mutants present a general growth delay and die during midgestation (from E10.5 to E13.5). Heterozygous mice show decrease neonatal viability and a significant body size reduction from birth to adulthood. Neurobehavioral analysis in these animals revealed developmental delay, specific motor alterations and learning deficits *(Fotaki et al., 2002)*. In humans, DYRK1A haploinsufficiency has been associated with microcephaly, intrauterine growth retardation and general developmental delay *(Courcet et al., 2012; Fujita et al., 2010; Moller et al., 2008; Yamamoto et al., 2011)*. In addition, DYRK1A disrupting mutations have been proposed to be associated with autism spectrum disorders *(O'Roak et al., 2012)*. Even if deregulation of DYRK1A gene dosage affects mostly nervous system development and homeostasis, DYRK1A dosage imbalance has been also associated with several other pathologies such as cardiac hypertrophy, bone homeostasis or tumor growth *(Baek et al., 2009; Kuhn et al., 2009; Lee et al., 2009; Malinge et al., 2012; Raaf et al., 2010)* suggesting that the activity of this kinase is required for the right functioning of many organs and systems.

Gene and protein expression

DYRK1A is organized in 17 exons and is expressed as several transcripts, differing in their 5'-ends due to alternative promoter usage and first exon choice (schematic representation in Fig. I4A). Another alternative splicing event, widely represented in different tissues, exists and involves the use of an alternative acceptor splicing site in exon 4 (Fig. I4B), which gives rise to two protein isoforms differing in the inclusion/exclusion of nine amino acids in the N-terminus of the protein *(Guimera et al., 1997)*. However, no functional differences among these protein isoforms have been reported yet.

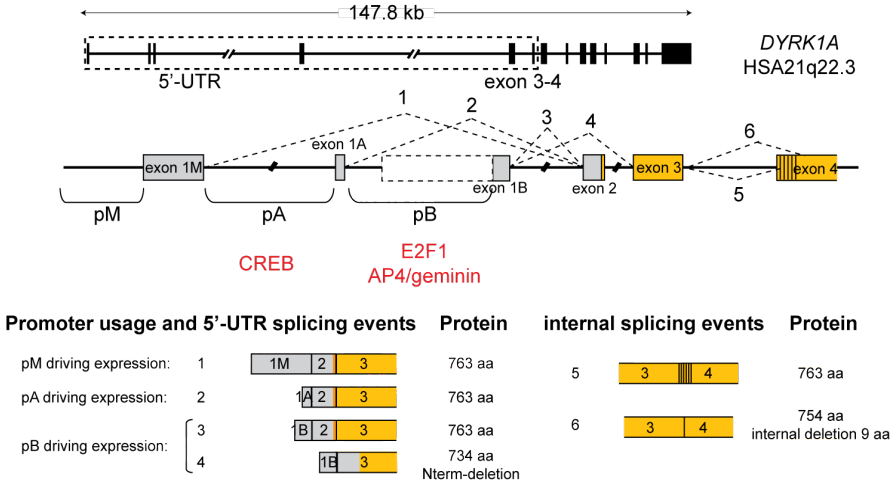


Figure I4: Genomic structure and splicing events of human DYRK1A. Schematic representation of human DYRK1A gene: boxes represent exons, while lines represent introns. Experimental evidences suggest that DYRK1A contains three putative promoter regions, promoter A (pA) before exon 1A, promoter B (pB) before exon 1B and promoter M (pM) before exon 1M, putative promoter regions. Transcription from pM and pB would result in transcripts with an identical coding potential, whereas transcripts generated from pA could potentially encode a protein lacking the first 29 amino acids. Gene size and chromosomal location are indicated. Adapted from *(Aranda et al., 2011)*.

DYRK1A is broadly expressed in both developing and adult tissues in mammals *(Becker et al., 1998; Guimera et al., 1999; Guimera et al., 1996)*, although the information about how expression is regulated is still quite scarce. It is known that the different *DYRK1A* promoters appear to have distinct responses to transcription factors, such as E2F1 or CREB (CRE binding protein) *(Impey et al., 2004; Maenz et al., 2008)* (Fig. I4). The transcription factor AP4 (activator protein 4) and its

corepressor partner geminin have been shown to repress DYRK1A transcription through recruiting HDAC3 (histone deacetylase 3) in non neuronal tissues *(Kim et al., 2006)*, while REST (RE1 silencing transcription factor) appears to regulate DYRK1A activity during neurodevelopment *(Lu et al., 2011)*. *DYRK1A* mRNA is upregulated in the brain of mice overproducing β-amyloid protein (Aβ), a major component of senile plaques in Alzheimer's disease, and in neuroblastoma cells when incubated with Aβ *(Kimura et al., 2007)*. Transcriptional activation has also been shown in osteoblastoma cells treated with RANKL (receptor activator of nuclear factor κB ligand), very likely through activation of NFATs *(Lee et al., 2009)*. Additionally, DYRK1A protein shows circadian oscillation in the mouse liver *(Kurabayashi et al., 2010)*, although it is not known whether the regulation occurs at the transcriptional or post-transcriptional levels. Finally, the control of DYRK1A expression via miRNAs has been proposed and miRNAs let-7b, miR-199b and miR-1246 reported as responsible *(Buratti et al., 2010; da Costa Martins et al., 2010; Zhang et al., 2011)*. It is worth to note that NFAT, the Aβ precursor protein and p53, which regulates the expression of miR-1246, are DYRK1A substrates *(Arron et al., 2006; Park et al., 2010; Ryoo et al., 2008)*, and thus the transcriptional/post-transcriptional effects described above could be part of regulatory loops.

DYRK1A regulation

One consequence of the mechanism of activation of all DYRKs is that the activation loop would be constitutively phosphorylated, and consequently the kinase would be constitutively active. However, given the extreme DYRK1A gene dosage sensitivity it is sensible to think that its activity has to be tightly regulated. This might happen not only by controlling protein amounts by regulatory mechanisms at the transcriptional/post-transcriptional level as discussed in the previous section, but also by actions at the post-translational level including the control of protein stability, the regulation of the catalytic activity, the regulation of subcellular localization and the accessibility to regulators or substrates (summarized in Fig. I5).

DYRK1A is predicted to have a short half-life, as it harbors a PEST motif in its non-catalytic C-terminal domain, although no regulatory mechanisms at this level have been reported so far. Unpublished

results from our group indicate that active DYRK1A has a longer half-life than the kinase inactive form, suggesting that DYRK1A might regulate its own stability. We have also observed that DYRK1A protein levels change during the cell cycle with lower levels in G1 that gradually increase during S phase until reaching the highest levels in G2/M (C. Di Vona and S. de la Luna, unpublished results). Whereas the increase in S phase appears to be transcriptional, likely a consequence of DYRK1A being an E2F1 target *(Maenz et al., 2008)*, the decrease in the M to G1 transition could involve protein stability.

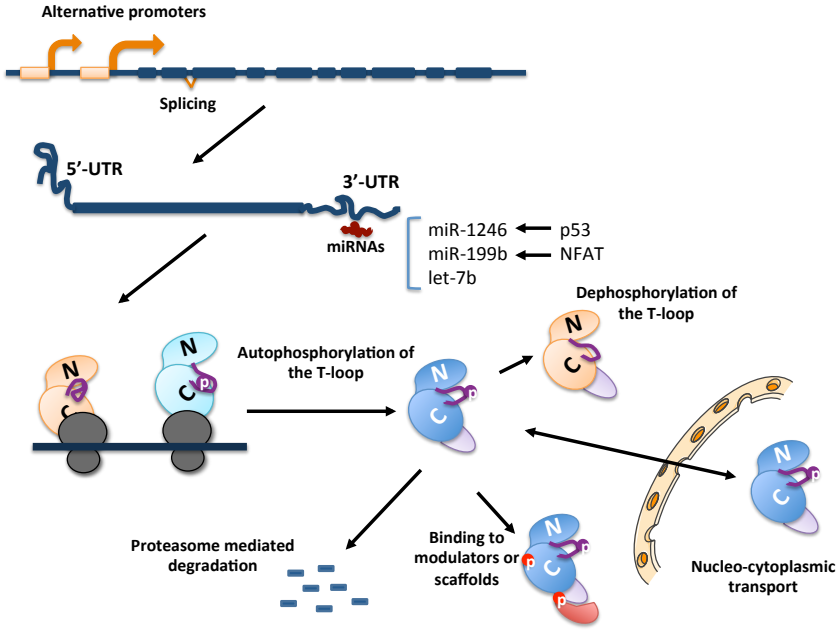


Figure I5: DYRK1A activity can be regulated at different levels. Schematic representation of the different steps at which DYRK1A amounts or activity can be regulated. Follow the text for specific explanations.

Two examples of mechanisms by which the modulation of DYRK1A kinase activity relies on the interaction with regulatory proteins have been already published. One of them has been described by our group and involves DYRK1A autophosphorylation on Ser520 residue outside the catalytic domain. This event triggers the association with 14-3-3β and induces a conformational change, resulting in increased DYRK1A catalytic activity *(Alvarez et al., 2007)*. Another example is the binding of SPRED1/2 (Sprouty-related protein with an EVH1 domain) to the kinase domain of DYRK1A; this interaction appears to compete for the

same binding site with other substrates and inhibit DYRK1A-dependent phosphorylation *(Li et al., 2010)*. Additionally, dephosphorylation of the activation loop should turn off DYRK1A kinase activity, but no specific phosphatases have been described yet.

Regulation of the subcellular localization may represent another way of controlling DYRK1A activity linked to substrate accessibility. Substrates for DYRK1A have been reported over recent years, including cytosolic and nuclear proteins (for review, see *Aranda et al., 2011)* (Fig. I6). The distinct nature of these substrates likely responds to the ability of DYRK1A to localize both to the nucleus and the cytoplasm, exerting its role in different cellular compartment (see next section). However, the knowledge on the regulatory mechanisms acting at this level is still very poor.

DYRK1A subcellular localization

DYRK1A localizes both in the cytoplasm and in the nucleus of different neuronal types in the human, mouse and chick central nervous system (CNS) *(Hammerle et al., 2003; Marti et al., 2003; Wegiel et al., 2004)*, while its expression is mostly restricted to the cytosol in glial cells *(Marti et al., 2003)*. Biochemical fractionation experiments indicate that DYRK1A is distributed in different pools in mouse brain cytosolic fraction: a soluble DYRK1A pool, a pool associated with the synaptic plasma membrane and a pool associated with vesicle-containing fractions *(Aranda et al., 2008; Murakami et al., 2009)*.

When exogenously expressed, DYRK1A is only localized in the nucleus of a wide variety of established cell lines *(Alvarez et al., 2003; Becker et al., 1998; Guo et al., 2010; Mao et al., 2002; Seifert et al., 2008; Sitz et al., 2004)*, in primary hippocampal neurons *(Sitz et al., 2004)* and in neural cells of the embryonic neocortex *(Yabut et al., 2010)*. A bipartite nuclear localization signal (NLS), at the non-catalytic N-terminus, and a complex NLS, within the CMGC-extension in the catalytic domain, appear to contribute to DYRK1A complete nuclear accumulation *(Alvarez et al., 2003)*; these NLSs are also conserved in the DYRK1A closest homologous DYRK1B and they are also responsible for the accumulation of DYRK1B in the nucleus (de la Luna's lab, unpublished results) (Fig. I2).

A switch from the nucleus to the cytoplasm has been described during Purkinje cells differentiation in chicken *(Hammerle et al., 2002)*, and results from our group indicate that DYRK1A has nucleocytoplasmic shuttling activity (de la Luna's group, unpublished results). All these data suggest that DYRK1A subcellular localization may be a control point to regulate substrate accessibility, although whether it is regulated at the nuclear import, nuclear export or nuclear/cytosolic retention levels it is still not known.

Within the nucleus, DYRK1A accumulates in a subnuclear compartment known as nuclear speckles (also known as splicing factor compartment (SFC) or interchromatin granule clusters); a stretch of 13 consecutive His residues in the C-terminus of DYRK1A is responsible of the targeting to this nuclear domain *(Alvarez et al., 2003)*. The presence of DYRK1A in the nuclear speckles has been also observed in physiological conditions *(Fernandez-Martinez et al., 2009)* de la Luna's group, unpublished results). Nuclear speckles are defined as 20-50 subnuclear dots that function as compartments of storage and assembly of components of the splicing machinery *(Spector and Lamond, 2011)*. Their main components are splicing factors (snRNPs and SR proteins), as well as transcription factors, 3'-RNA processing factors, translation factors, ribosomal proteins, a subpopulation of the RNA polymerase II (RNA Pol II) and some kinases and phosphatases *(Saitoh et al., 2004; Spector and Lamond, 2011)*. Moreover, speckles are known to be highly dynamic structures, with changes in number, shape and size depending on the transcriptional state and on the phase of the cell cycle. This dynamic behavior is associated to their activity as suppliers of splicing factors to sites of active transcription/pre-mRNA processing *(Lamond and Spector, 2003)*. The exchange rate of speckle factors is regulated by cycles of phosphorylation and dephosphorylation: in general and for the splicing factors, when phosphorylated, they are released into the nucleoplasm and recruited to the forming spliceosome; when dephosphorylated, they reassociate in nuclear speckles. The role of DYRK1A within the nuclear speckles is not yet clear, although it may be linked to the fact that several substrates of DYRK1A are splicing factors *(de Graaf et al., 2004; Ding et al., 2012; Galceran et al., 2003; Qian et al., 2011; Woods et al., 2001b; Yin et al., 2012)*. The overexpression of DYRK1A induces speckle disassembly in a subpopulation of cells, and this effect is dependent on the kinase activity of the protein *(Alvarez et al., 2003)*. It is worth noting that although several kinases accumulate in nuclear

speckles, only members of the CLK and SRPK families and PRPF4 have been reported to alter the normal appearance of nuclear speckles when overexpressed *(Colwill et al., 1996; Gui et al., 1994)*; *(Kojima et al., 2001; Kuroyanagi et al., 1998; Wang et al., 1998)*, and that these kinases are either DYRK kinases or close relatives of the family.

Functional significance for the presence of DYRK1A in the nucleus

The participation of DYRK1A in different cellular processes and signal transduction pathways has been mostly inferred from the activity of its interactors and substrates. The diversity of its substrates points at DYRK1A as a pleiotropic regulator of different cellular processes, such as transcription, splicing, metabolism, apoptosis and endocytosis, among others (summarized in Fig. I6).

A relevant group of DYRK1A targets are proteins involved in the regulation of transcription and gene expression programs. For instance, DYRK1A has been reported to directly phosphorylate NFAT, which in turns prime further phosphorylation by CK1 (casein kinase I) and GSK3 promoting NFAT translocation to the cytosol *(Arron et al., 2006; Guo et al., 2010; Gwack et al., 2006)*. DYRK1A promotes Gli1 nuclear, accumulation and enhances the transcriptional activity of full-length Gli1 but not of a natural N-terminally truncated splicing variant *(Mao et al., 2002; Shimokawa et al., 2008)*.

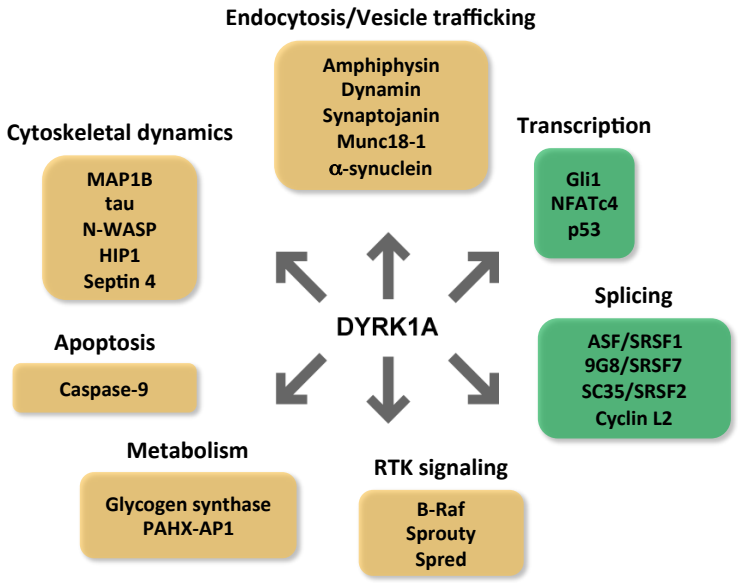


Figure I6: Substrates and multiple roles for mammalian DYRK1A. Substrates of DYRK1A are depicted in boxes, accordingly to the type of cell process in which they are involved, indicated on top of the boxes. Yellow boxes indicate proteins cytosolic or with cytosolic activities; green boxes indicate proteins with nuclear activities. A full list of substrates and references can be found at *(Aranda et al., 2011)*.

The fact that the list of DYRK1A substrates comprises both nuclear and cytosolic proteins (Fig. I6) likely reflects the ability of DYRK1A to shuttle between the nucleus and the cytoplasm. In addition, all "nuclear" DYRK1A substrates currently known are proteins capable of shuttling from the nucleus to the cytoplasm and vice versa, and in fact, the DYRK1A phosphorylation events on these proteins could well be occurring in the cytoplasm. Therefore, a role for DYRK1A within the nuclear compartment is still an issue to resolve.

Protein kinases and transcriptional regulation

Classical view: regulation of transcription factors

Upon exposure to changes in the environment or to developmental cues during differentiation, a cell reprograms transcription in its nucleus through a circuitry of signals that ultimately alters gene expression. Key elements of such signal-transducing cascades are protein kinases that, when at the end of such cascades (terminal kinases), trigger the necessary response by directly phosphorylating transcription factors, co-regulatory proteins, or chromatin-related proteins. Phosphorylation will alter thus the activity of the target protein at different levels including stability, nucleocytoplasmic transport or protein-protein interaction (Fig. I7). In this scenario, a nuclear activity for the kinase is not strictly required, and no stable association with the chromatin that harbors target genes of a given signaling pathway is expected.

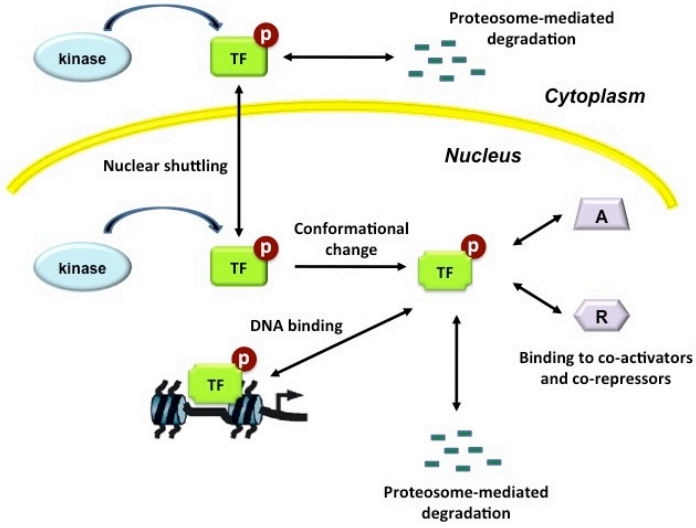


Figure I7: Protein kinases affect transcription at different levels: the classical view Schematic representation showing some of the effects that can be triggered by phosphorylation of transcription factors. TF, transcription factor; A, activator; R, repressor; p, phosphorylation.

This enzymatic view for terminal kinases has been the prevailing view until recently, when an alternative model is spreading out. In this model, terminal kinases are recruited to chromatin and play a more

direct role by acting directly on transcription complexes at target genes, and promoting thereby increase gene activity at other steps beyond mere enhancement of a transcription factor activity. Recent knowledge about known kinases able to physically occupy certain genes and thus regulating gene expression is summarized below.

Protein kinases as components of transcription factor complexes

The best-characterized example of how kinases can bind regulatory regions within gene promoters is the yeast MAPK Hog1p (high-osmolarity glycerol 1p). This example clearly illustrates how kinases can directly connect the response to an environmental signal to changes in the transcription of target genes. The current model proposes that the transcription factor Hot1p recruits Hog1p to the promoter regions to phosphorylate either additional transcription factors, chromatin remodeling complexes or nucleosomal components (Fig I8). In the absence of signal, Sko1p binds to promoter elements as part of a repressor complex with Cyc8p and Tup1p. Exposure to hyperosmotic stress induces the recruitment of Hog1p to Sko1p and leads to the phosphorylation of Sko1p (Alepuz et al., 2001). This in turn promotes the recruitment of SAGA and SWI/SNF chromatin-modifying factors resulting in transcriptional activation *(Proft and Struhl, 2002)*.

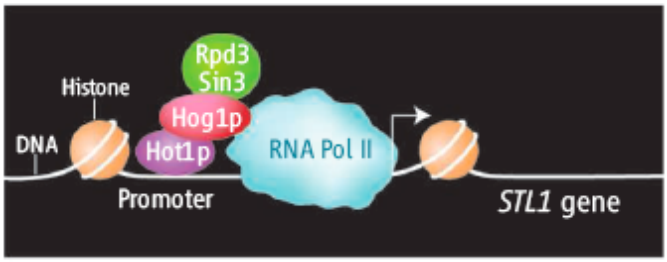


Figure I8: The protein kinase Hog1p is recruited to promoter regions. Schematic representation of the recruitment of Hog1p to chromatin using the osmoresponsive target *STL1* gene as example (adapted from *(Edmunds and Mahadevan, 2006)*.

Further work showed that Hog1p not only functions as a kinase at such genes, but also forms an integral component of transcription complexes involved in the recruitment of transcription factors, components of the general transcription machinery, RNA Pol II, and chromatin remodeling/modifying activities. In addition, Hog1p is also present on coding regions of genes whose expression is induced upon

osmotic shock and travels with the elongating RNA Pol II pointing to a role on pre-mRNA elongation (reviewed in *De Nadal et al., 2004)*. In *S. cerevisiae*, the physical and functional association of signaling kinases with chromatin has been further extended to several other kinases including the kinases of the mating pheromone signaling pathway Fus3p and Kss1p, the two catalytic subunits of protein kinase A, Tpk1p and Tpk2p, and the MAPK Mpk1p *(Kim and Levin, 2011; Pokholok et al., 2006)*. Each of these kinases bind to different type of genes and regions of genes, likely reflecting the mechanisms by which they are recruited and/or the functions that they carry out. Of note, the Mpk1p roles on the activation of stress-induced genes are not dependent on its catalytic activity but on its ability to recruit initiation factors to target promoters or by blocking the recruitment of the termination complex to the elongating RNA Pol II *(Kim and Levin, 2011)*.

The mammalian homolog of Hog1p, the MAPK p38, has been shown to be also recruited to stress-induced genes *(Alepuz et al., 2003)*. The recruitment requires p38 activity and its interaction with the transcription factors ELK1, AP1 or NF-κB, depending on the promoter context and/or the stimuli, but the targets of p38 activity on these genomic regions are not known *(Ferreiro et al., 2010)*. The presence of p38 to the enhancer region of muscle-specific genes during myogenic differentiation has been observed *(Simone et al., 2004)* and linked to an enhanced recruitment of the SWI-SNF complex at these elements *(Serra et al., 2007)*. Similarly to its yeast ortholog Hog1p, the presence of active p38 at the open reading frames of target genes also suggests the involvement of this kinase in transcription elongation *(Ferreiro et al., 2010)*, although no molecular targets for this activity have been found yet.

Considering the conservation of MAPK signaling pathways through evolution, it is highly likely that mammalian MAPKs might be components of transcriptional complexes, and this is the case not only for the MAPK p38 but also for extracellular signal-regulated kinases 1 and 2 (ERK1 and ERK2) and the c-Jun N-terminal kinases (JNK). Thus, the expression of a subset of genes during the differentiation of stem cells into neurons depends on the recruitment of JNK1/3 to the target promoters where they phosphorylate H3-S10 *(Tiwari et al., 2012)*. Moreover, JNK1 is recruited to the *c-Jun* promoter in response to glucocorticoids *(Bruna et al., 2003)*, and to hormone-respondent genes upon estrogen treatment of breast cancer cells. The formation of

promoter-bound complexes appears to require the estrogen receptor (ER) and, in some cases, tethering proteins such as AP-1 *(Sun et al., 2012)*. The MAPK ERK2 is also recruited to ER-binding sites across the genome in response to estrogen *(Madak-Erdogan et al., 2011)* and to AP-1 reporter genes in a phorbol ester-dependent manner *(Benkoussa et al., 2002)*. Other stimuli such as glucose, interleukin-1β or neural growth factor stimulate the binding of ERK1/2 to the insulin gene and other promoters, in a process that depends on the catalytic activity of the kinase and on the presence of the transcription factors targeted by ERK1/2 in the distinct signaling cascades *(Lawrence et al., 2008)*. For instance, ELK1 appears to be the tethering factor for ERK2 to its genomic loci in human embryonic stem cells *(Goke et al., 2013)*. Moreover, it was recently shown that ERK2 is able to repress the expression of a subset of IFNγ-induced genes that contains a specific consensus sequence to which ERK2 was shown to bind directly *in vitro (Hu et al., 2009)*. Finally, there is evidence for a direct role of ERK5 in transcriptional activation. ERK5 contains a potent transcriptional activation domain within its extended C-terminus. Activated ERK5 binds to the transcription factor MEF2D and is required for maximal transcriptional activation of MEF2D *(Kasler et al., 2000)*.

Other kinases, not belonging to the MAPK modules, have recently entered into the category of chromatin-associated kinases. Thus, PIM1 is able to phosphorylate H3-S10 on the nucleosome at MYC-target promoters, where it is recruited via complex formation with the dimer MYC-MAX, and thus contribute to their transcriptional activation *(Zippo et al., 2007)*. In the case of PKC-θ, the kinase forms a multisubunit complex that includes RNA Pol II, MSK1, the demethylase LSD1 and the adaptor molecule 14-3-3 at the regulatory regions of inducible immune response genes in activated human T cell *(Sutcliffe et al., 2011)*.

One of the targets for signaling kinases on genomic regions is chromatin itself, as part of the post-translational modifications of chromatin proteins that contribute to rapid alterations in nucleosomal structure. In particular, phosphorylation of serine residues 10 and 28 in histone H3 (H3-S10 and H3-S28, respectively) and serine 6 in the nonhistone chromatin protein HMGN1 are key events of the induction of immediate-early genes, such as *c-Fos* and *c-Jun*, in response to a wide collection of extracellular stimuli including growth factors, stressors and neurotransmitters (reviewed in *Hazzalin and*

Mahadevan, 2002). These phosphorylation events are mediated by MSK1 and MSK2 (mitogen- and stress-activated protein kinases 1 and 2), with are downstream of the ERK or the p38 MAPKs *(Dyson et al., 2005)*. The transient recruitment of MSK1/2 to chromatin requires different tethering proteins such as the transcription factors CREB and ATF1 in the case of ERK-signaling *(Shimada et al., 2010)* or active progesterone receptor in response to this hormone *(Vicent et al., 2006)*. It is currently accepted that phosphorylation of H3-S10 and H3-S28 recruits 14-3-3 proteins, which in turn mediate the recruitment of the chromatin remodeler SWI/SNF to induce the subsequent remodeling of promoter nucleosomes to promote transcription initiation *(Drobic et al., 2010)*; displacement of repressive complexes has been also associated to the phosphorylation of H3-S10 *(Vicent et al., 2013)*. Apart from MSK1/2, and as mentioned above the presence on chromatin regions of other protein kinases, such as JNK1/3 and PIM1, has been associated to the increased phosphorylation of H3-S10 and subsequent transcriptional activation *(Sun et al., 2012; Tiwari et al., 2012; Zippo et al., 2007)*.

All these evidences open up the possibility that terminal kinases might have dual functions, apart from its ability to act as histone kinases: a structural role, by mediating crucial protein-protein interactions within various transcription complexes or acting as a recruiter platform for components of the transcriptional initiation or elongation machineries, and an enzymatic role, by phosphorylating target proteins in such complexes to positively or negatively regulate their activities (Fig. I9).

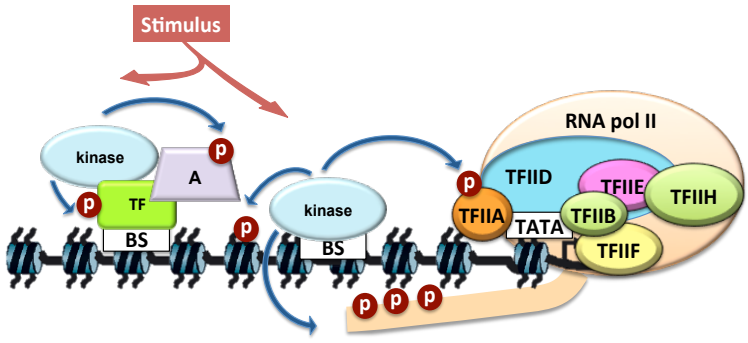


Figure I9: Protein kinases affect transcription at different levels: the chromatin view. Schematic representation showing some of the effects that can be triggered by phosphorylation at chromatin promoter regions.

The next chapter is dedicated to briefly introduce basic concepts of eukaryotic transcription in mammals, highlighting features of some components of the transcriptional machinery which are more relevant for the interest of this Thesis work.

Eukaryotic transcription

Gene regulation by the transcriptional machinery is central to most cellular responses including those directed to maintain cellular homeostasis in response to extracellular cues, and those involved in the regulation of differentiation and developmental programs. Eukaryotic gene expression involves several steps that start with the synthesis of RNA, an activity that is conducted by three multisubunit RNA polymerase (Pol) enzymes - I, II, and III - which are in charge of the synthesis of different classes of cellular RNAs *(Roeder and Rutter, 1969*, *1970)*. RNA Pol I synthesizes the 25S ribosomal RNA (rRNA) precursor, which encodes the 18S, the 5.8S, and the 28S RNA molecules of the ribosome in eukaryotes (reviewed in *Russell and Zomerdijk, 2005)*. RNA Pol II is dedicated to the synthesis of messenger RNAs (mRNAs), long non-coding RNAs (lncRNAs), most small nuclear RNAs (snRNAs) and microRNAs (miRNAs). Finally, RNA Pol III synthesizes transfer RNAs (tRNAs), 5S ribosomal RNA (rRNA), and a few other small RNAs (for a recent review, see *White, 2011)*. Next sections will deal only with RNA Pol II-dependent processes.

Basal transcription machinery

RNA Pol II is composed of 12 distinct subunits forming a complex of more than 500 kDa that is present in cells in low abundance *(Shilatifard et al., 2003)*. Early purification studies in yeast and mammals *(Chambon, 1975; Schwartz and Roeder, 1975)* led to the elucidation of the 12-subunit composition of RNA Pol II *(Hampsey, 1998)*. Apart from the core subunits, other factors are involved in the accurate transcription by RNA Pol II, which can be classified into three groups: general (or basic) transcription factors (GTFs), promoter-specific activator proteins (activators), and coactivators. GTFs are necessary and can be sufficient for accurate transcription initiation in vitro (reviewed in *Orphanides et al., 1996)*. Such factors include RNA Pol II itself and a variety of auxiliary components, including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH. In addition to these “classic” GTFs, in vivo transcription also requires Mediator, a highly conserved, large multisubunit complex that was originally identified in yeast, and that enables a response to regulatory factors (reviewed in *Malik and Roeder, 2005)*.

The events leading to efficient transcription are summarized as preinitiation complex (PIC) formation, initiation, promoter clearance, elongation and finally, transcription termination (reviewed, for instance, in *Saunders et al., 2006)*.

Transcription initiation and promoter clearance

The first step in PIC assembly is binding of TFIID, a multisubunit complex consisting of TATA-box-binding protein (TBP) and a set of tightly bound TBP-associated factors (TAFs). This is followed by GTFs assembling on the core promoter in an ordered fashion to form a PIC, which directs RNA Pol II to the transcription start site (TSS) (Fig. I10). Transcription then proceeds through a series of steps, including promoter melting, clearance, and escape, before a fully functional RNA Pol II elongation complex is formed.

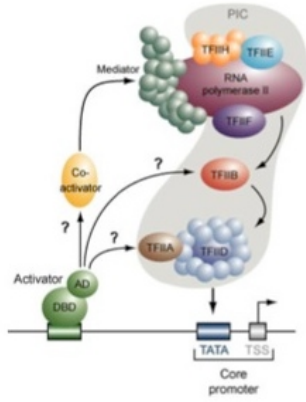


Figure I10: Assembly of the pre-initiation complex. Schematic representation of PIC formation by the stepwise recruitment of the general transcription machinery. For abbreviations, see the text. AD, activation domain; DBD, DNA binding domain. Adapted from *(Maston et al., 2006)*.

The earliest stage of transcription is marked by instability of the transcription complex and a tendency to release the RNA, in the form of short RNA products: RNA Pol II is thus viewed to be in a mode of abortive initiation, composed by a stalled RNA Pol II–DNA complex and short RNA products. The presence of all NTPs allows RNA Pol II to clear the promoter. During this stage, the PIC is partially disassembled: a subset of GTFs remains at the promoter, serving as a scaffold for the formation of the next transcription initiation complex *(Yudkovsky et al., 2000)*. After

RNA Pol II escapes from the promoter, a scaffold structure, composed of TFIID, TFIIE, TFIIH, and Mediator, remains on the core promoter *(Hahn, 2004)*; subsequent reinitiation of transcription then only requires re-recruitment of RNA Pol II-TFIIF and TFIIB.

Phosphorylation cycle of the RNA polymerase II C-terminal domain

Promoter clearance coincides with the beginning of another cycling event within the transcription cycle: phosphorylation of the CTD. The largest catalytic subunit of RNA Pol II (RPB1), contains multiple repeats of the heptapeptide sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser in its C-terminal domain (CTD) *(Hirose and Ohkuma, 2007)*. The number of these repeats increases with genomic complexity: 26 in yeast, 32 in *C. elegans*, 45 in *Drosophila*, and 52 in mammals. The CTD is the site of multiple phosphorylations, with up to five potential phosphorylation sites in the consensus heptapeptide repeat (Tyr1, Ser2, Thr4, Ser5 and Ser7).

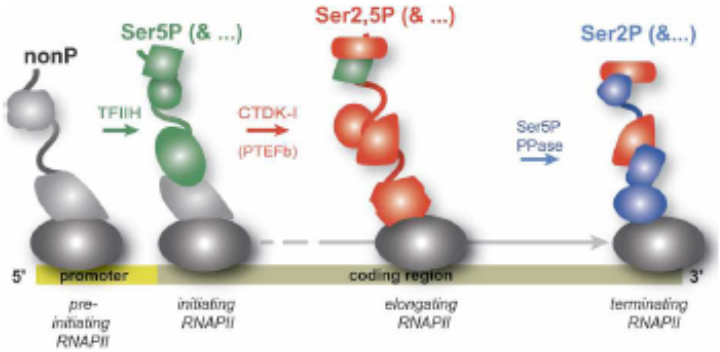


Figure I11: The phosphorylation states of the CTD are associated with different steps of transcription. Schematic representation of the different phosphorylation states of the RNA Pol II (grey oval) CTD (curve line, represented with diverse color depending on the phosphorylation). Each phosphorylation status is associated to a specific transcription step (see text for details). Proteins bound to a type of repeat are indicated in the same color as the repeat. Adapted from *(Phatnani and Greenleaf, 2006)*.

Studies using functional assays together with antibodies specific to one or the other form of RNA Pol II demonstrated that RNA Pol II in the PIC was unphosphorylated (IIA) *(Laybourn and Dahmus, 1989)*, whereas transcription-competent RNA Pol II was heavily phosphorylated on its CTD (IIO) *(Christmann and Dahmus, 1981)*, mainly at Ser2 and Ser5

residues. Recently, phosphorylation on Ser7 has also been identified *(Chapman et al., 2007; Egloff et al., 2007)*. CTD phosphorylation at Ser5 correlates with transcription initiation and early elongation (promoter clearance), whereas Ser2 phosphorylation is associated with RNA Pol II farther away from the promoter (Fig. I11). The phosphorylation pattern of these serine residues is therefore essential for regulating the transition through various stages of RNA Pol II transcription (initiation, elongation, termination). Moreover, phosphorylation also regulates the activity of the CTD as a scaffold for pre-mRNA-processing factors involved in capping, splicing and cleavage/polyadenylation. For this reason, the CTD is thought to couple transcription and processing *(Hsin and Manley, 2012)*.

The three major kinases that target the RNA Pol II CTD are the cyclin-dependent kinases CDK7, CDK8, and CDK9 (reviewed in, for instance, *Prelich, 2002)*. These enzymes are evolutionarily conserved from yeast to mammals, and all three are components of different protein complexes. CDK7, and its partner cyclin H, is a subunit of TFIIH *(Orphanides et al., 1996)*, and is responsible for Ser5 phosphorylation subsequent to the formation of the first phosphodiester bond *(Akoulitchev et al., 1995)*. CDK8, together with its partner cyclin C, targets Ser5, associates with the Mediator complex and functions in transcriptional events prior to elongation *(Maldonado et al., 1996)*. Finally, CDK9 is part of elongation factor P-TEFb (see below) *(Zhu et al., 1997)*.

CTD-mediated processes: P-TEFb dependent elongation

P-TEFb (positive transcription elongation factor b) was originally identified based on its ability to stimulate DRB-sensitive transcription of long transcripts *in vitro (Marshall et al., 1996; Marshall and Price, 1992)*. The heterodimeric P-TEFb consists of the CDK9 kinase that associates either with cyclin T1, cyclin T2a, cyclin T2b, or cyclin K *(Peng et al., 1998)*. The elongation activity of P-TEFb is dependent on its kinase activity (reviewed in *Price, 2000)*, and studies using the highly specific P-TEFb inhibitor flavopiridol demonstrated that P-TEFb primarily targets Ser2 at actively transcribed genes in cells *(Ni et al., 2004)*.

P-TEFb exists mainly in two different forms: as a “large” inactive complex and as a “small” active complex. P-TEFb is maintained in an inactive state through the association with the 7SK small nuclear RNA (7SK snRNA) *(Nguyen et al., 2001; Yang et al., 2001)* and with either hexamethylene

bisacetamide inducible protein (Hexim) 1 or 2 *(Yik et al., 2003)*. Upon certain stimuli the "active" P-TEFb, composed only by CDK9 and the respective cyclin T associated, is released from the large complex and recruited to responsive promoters via different activators or co-activators to facilitate elongation and relieve pausing. This is the case of the bromodomain protein Brd4, which has been shown to bind to acetylated histones and target P-TEFb to actively transcribed genes *(Mochizuki et al., 2008)*.

CTD-mediated processes: pre-mRNA splicing

Splicing of mRNA precursors takes place in a large macromolecular complex called the spliceosome, which is composed of small nuclear ribonucleoprotein particles (snRNPs) and non-snRNP proteins including members of the serine/arginine-rich (SR) protein family (for recent reviews in different aspects of the splicing process, see *Han et al., 2011; Hoskins and Moore, 2012; Irimia and Blencowe, 2012)*. Both biochemical and *in vivo* studies have provided support for the existence of functional interactions between RNA Pol II, especially the CTD, and the splicing apparatus. RNA Pol IIO, but not RNA Pol IIA, has been found to associate with splicing factors, and this isoform has also been detected in active spliceosomes *(Kim et al., 1997; Mortillaro et al., 1996; Yuryev et al., 1996)*. The current working model, built from observations from many groups, proposes that the hyperphosphorylated CTD of elongating RNA Pol II function in splicing, serving as a platform upon which processing factors bind, thus helping to promote efficient and accurate splicing by targeting necessary factors to transcription sites (Fig. I12) (for a recent review see *Hsin and Manley, 2012)*.

CTD-mediated processes: cleavage/polyadenylation and termination

The final step in the cycle is transcript termination. At this stage, the mRNA is cleaved, polyadenylated, and transported to the cytoplasm, where it will be translated *(Proudfoot et al., 2002)*. Polyadenylation of mRNA takes place in two steps: endonucleolytic cleavage of the mRNA precursor followed by poly(A) addition to the 3'-end of the cleavage product. This process requires multiple protein factors, including cleavage/polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), two cleavage factors, CFI and CFII, and poly(A) polymerase

(for a recent review, see *Proudfoot, 2011)*. It has been shown that CPSF and CstF bind the CTD, and both are present in an RNA Pol II holoenzyme preparation. Furthermore, in a reconstituted transcription assay, CPSF was shown to transfer from TFIID to RNA Pol II concomitant with initiation. The CTD functions thus to help recruit polyadenylation factors to sites of RNA Pol II transcription, increasing their local concentration and thereby facilitating efficient processing (Fig. I12) (recently reviewed in *Hsin and Manley, 2012)*.

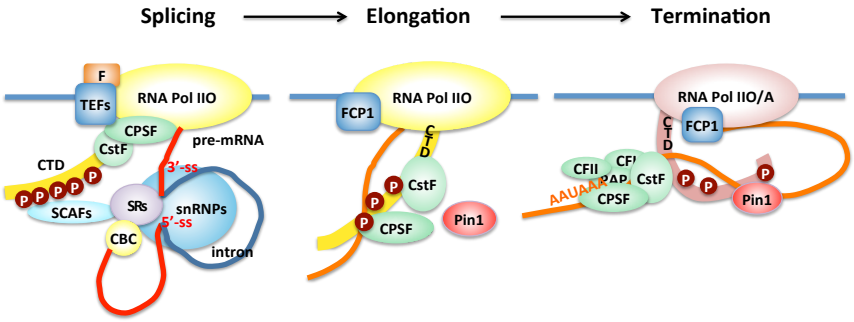


Figure I12: Coupling transcription elongation with pre-mRNA processing, cleavage and polyadenylation. RNA Pol IIO (hyperphosphorylated, in yellow) is associated to the splicing machinery. Introns (blue line) are spliced from exons (red lines) in a process that occurs co-transcriptionally. The CTD helps to recruit all the factors necessaries for the correct processing of the pre-mRNA (capping and splicing). After termination, the phosphatase FCP1 dephosphorylates the CTD to allow reinitiation of a new transcription cycle. CBC, cap-binding complex; CFI and CFII,, cleavage factor I and II; CPSF, polyadenylation factor; Cstf, cleavage stimulation factor; F, general transcription factor F; Pin 1, peptidylprolyl cis/trans isomerase Pin1; SCAF, SR-like CTD-associated factors; snRNPs, small nuclear ribonucleoproteins; SR, Ser/Arg-rich alternative splicing proteins; TEFs, transcription elongation factors.

Activators of transcription

The assembly of a PIC on the core promoter is sufficient to direct only low levels of accurately initiated transcription from DNA templates *in vitro*, a process generally referred to as basal transcription. Transcriptional activity is greatly stimulated by a second class of factors, termed activators. In general, activators are sequence-specific DNA-binding proteins whose recognition sites are usually present in sequences upstream of the core promoter. A property of activators is that they can stimulate transcription synergistically, a phenomenon in which the regulatory effect of multiple factors working together is greater than the sum of the activities driven by each factor individually. Activators work, at

least in part, by increasing PIC formation through a mechanism thought to involve direct interactions with one or more components of the transcriptional machinery *(Orphanides et al., 1996; Ptashne and Gann, 1997)*. Activators may also act by promoting a step subsequent to PIC assembly, such as initiation, elongation, or reinitiation *(Lee and Young, 2000)*. Finally, activators have also been proposed to function by recruiting activities that modify chromatin structure *(de la Serna et al., 2005; Lemon and Tjian, 2000)*.

Chromatin structure can influence transcription by preventing or facilitating the interaction between the transcriptional machinery with promoter DNA. Nucleosomal units are in fact subjected to a series of modifications, such as acetylation, methylation and phosphorylation, which can regulate the binding of an activator and PIC assembly (for recent reviews on the subject see *Campos and Reinberg, 2009; Suganuma and Workman, 2011)*. Actually, a strong relationship between the chromatin landscape and transcription exists, and it is reflected into the distinct placement of the different chromatin modifications along the gene body. For example, high levels of histone 3 acetylation and try-methylation at Lys4 are found at transcription start sites and strongly correlate with active gene expression, while try-methylation of Lys36 is found in the body of actively transcribed genes *(Barski et al., 2007; Bernstein et al., 2002; Bernstein et al., 2005)*. In contrast, elevated levels of histone 3 Lys27 methylation correlates with gene repression *(Roh et al., 2006)*.

The activity of an activator may be modulated also by co-activators or co-repressors. Typically, these effectors do not exhibit intrinsic sequence-specific DNA binding; instead, they are recruited by protein-protein interactions with one or more DNA-bound activators. Co-activators function in many of the same ways as activators, such as by stimulating PIC assembly or modifying chromatin towards an active. The specific set of coactivators/corepressors present in a cell can play a major role in determining the regulatory response, as they can modify an activator's ability to positively or negatively regulate transcription *(Lemon and Tjian, 2000)*.

Transcriptional cis-regulatory elements

Genes transcribed by RNA Pol II typically contain two distinct families of *cis*-acting transcriptional regulatory DNA elements: a promoter, which is

composed of a core promoter and nearby (proximal) regulatory elements, and distal regulatory elements, which can be enhancers, silencers, insulators, or locus control regions (reviewed in *Maston et al., 2006)*. The promoter typically spans less than 1 kb upstream the TSS, while the distal regulatory elements can be located up to 1 Mb from the promoter. The distal elements may contact the promoter through a mechanism that involves looping out the intervening DNA (Fig. I13).

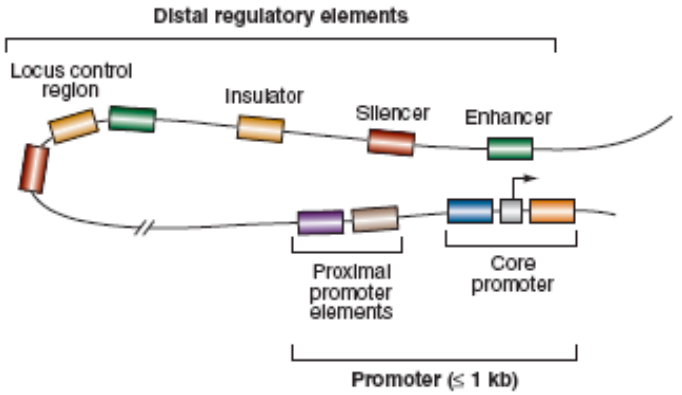


Figure I13: Transcriptional regulatory elements. See the text for details. Adapted from *(Maston et al., 2006)*.

The core promoter serves as the docking site for the basic transcriptional machinery and PIC assembly, and defines the position of the TSS as well as the direction of transcription. The first described core promoter element was the TATA-box, the binding site for the TBP subunit of TFIID *(Breathnach and Chambon, 1981)*. In addition to the TATA-box, metazoan core promoters can be composed of numerous other elements, including: Initiator element (Inr), which encompasses the TSS *(Corden et al., 1980)*, Downstream Promoter Element (DPE) *(Burke and Kadonaga, 1996)*, Downstream Core Element (DCE) *(Lewis et al., 2000)*, TFIIB-Recognition Element (BRE) *(Lagrange et al., 1998)*, Motif Ten Element (MTE) *(Lim et al., 2004)* and the X Core Promoter Element 1 (XPCE1) *(Tokusumi et al., 2007)* (Fig. I14).

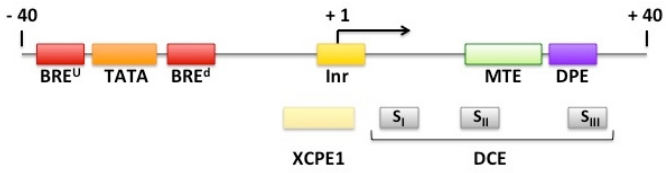


Figure I14: Core promoter elements for a RNA Pol II promoter. Scheme depicting the main core promoter elements (in boxes) within mammalian promoters. Positions respect their genomic location respect to TSS (+1). For details, see the text. Adapted from *(Juven-Gershon and Kadonaga, 2010)*.

With the exception of the BRE, which is specifically recognized by TFIIB, all other core promoter elements are TFIID-interaction sites: TAF6 and TAF9 contact the DPE, TAF1 and TAF2 contact the Inr, and TAF1 contacts the DCE *(Lee et al., 2005; Smale and Kadonaga, 2003)*. Core promoters are diverse in their content and organization, and therefore PIC assembly does not depend on a single nucleation point, such as a TATA-box; rather, many of the core promoter elements interact with TFIID and stabilize PIC assembly. Additionally, although TBP is present at many TATA-less promoters, various core promoters may interact with TFIID complexes having different subunit compositions *(Chen and Manley, 2003; Muller and Tora, 2004)*.

An interesting feature of around 60% of human genes is that their promoter falls near a CpG-island *(Venter et al., 2001)*, a relatively short stretch of DNA (500 bp to 2 kb) that has a high G+C nucleotide content and a high frequency of the CpG dinucleotide compared to bulk DNA; in fact, the presence of a CpG-island is the most reliable indicator for predicting the presence of a gene *(Ioshikhes and Zhang, 2000)*. Although the elements that are responsible for CpG-islands core promoter function remain poorly defined, some correlations have been observed between the presence of CpG-islands and certain core promoter elements: TATA-boxes are more common in promoters that do not have a CpG-island nearby, whereas BREs are more common in promoters associated with CpG-islands *(Gershenzon and Ioshikhes, 2005)*.

Finally, the proximal promoter region typically contains multiple binding sites for activators or repressors. Specificity in gene regulation is provided by the fact that transcription factors bind DNA preferentially at

characteristic sequence motifs. Although an activator/repressor can bind to a wide variety of sequence variants, in certain instances the precise sequence of a binding can impact the regulatory output by directing a preference for certain dimerization partners over others, or affecting the structure of a bound activator in a way that alters its activity *(Lefstin and Yamamoto, 1998; Remenyi et al., 2004)*. Work from many laboratories during the last decades has led to the generation of a catalogue of consensus sequences for a limited subset of transcription factors, which can be consulted through several public databases, such as TRANSFAC or Jaspar *(Knuppel et al., 1994; Sandelin et al., 2004; Wingender, 1994)*. In recent years, however, the availability of chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) approaches has allowed the detection, at a genome-wide level, of the binding locations for many individual transcription factors (see for instance, results from the ENCODE project in *(Wang et al., 2012)*, which has, and will, widened our knowledge on proximal regulatory sequences.

Material and Methods

Plasmids

Backbone vectors

- pCDNA-3: mammalian expression vector (Invitrogen).
- pCG4-DBD: DNA binding domain (DBD) from *S. cerevisiae* transcription factor Gal4 (amino acids 1-147) cloned into pcDNA3 *(de la Luna et al., 1999)*.

Expression plasmids for Gal4-DBD fusion proteins

- pG4DBD-DYRK1A: expression plasmid for human DYRK1A (isoform 754 amino acids) N-terminally fused to Gal4 DBD, obtained by digestion of pHA-DYRK1A *(Alvarez et al., 2003)* with EcoRI/XbaI and then ligated into the EcoRI/XbaI sites of pCG4-DBD (described in section 1.1).
- pG4DBD-DYRK1A^{KR}: expression plasmid for human DYRK1A with a mutation in the ATP-binding site (Lys179 to Arg) that generates a catalytically inactive protein kinase fused to Gal4 DBD; obtained by digestion of pHA-DYRK1A-K179R *(Alvarez et al., 2003)* with EcoRI/XbaI and then ligated into the EcoRI/XbaI sites of pCG4-DBD.
- pG4DBD-CDK9: expression plasmid for human CDK9 fused to Gal4 DBD, kindly provided by C. Suñe (Instituto de Parasitologia y Biomedicina "Lopez-Neyra"-CSIC, Granada, Spain) *(Majello et al., 1999)*.
- pG4DBD-CDK9^{dn}: expression plasmid for human kinase inactive CDK9 bearing a single amino acid change (Asp167 to Asn) in its catalytic domain, fused to Gal4 DBD, kindly provided by C. Suñe *(Majello et al., 1999)*.
- pG4DBD-DYRK1B: expression plasmid for human DYRK1B N-terminally fused to Gal4 DBD, obtained by digestion of pHA-DYRK1B with EcoRI and then ligated into the EcoRI site of pCG4-DBD.
- pG4DBD-DYRK2: expression plasmid for human DYRK2 (601 amino acid isoform) N-terminally fused to Gal4 DBD, obtained by digestion of pHA-DYRK2 with EcoRI and then ligated into the EcoRI site of pCG4-DBD.
- pG4DBD-DYRK3: expression plasmid for human DYRK3 (588 amino acid isoform) N-terminally fused to Gal4 DBD, obtained by

digestion of pHA-DYRK3 with EcoRI and then ligated into the EcoRI site of pCG4-DBD.
- pG4DBD-DYRK4: expression plasmid for human DYRK4 (520 amino acid isoform) fused to Gal4 DBD, obtained by digestion of pHA-DYRK4 with EcoRI and then ligated into the EcoRI site of pCG4-DBD.
- pG4DBD-VP16: transactivation domain of the herpes simplex virus protein VP16 (amino acids 411-490) fused to Gal4 DBD *(Sadowski et al., 1988)*.

Reporters plasmids

- pG5HIV-Luc: a reporter plasmid containing the Firefly luciferase gene under the control of the human immunodeficiency virus 1 (HIV-1) 3'-long terminal repeat and five Gal4 binding sites *(Montanuy et al., 2008)*, kindly provided by C. Suñe.
- pG4-E1b-Luc: reporter plasmid containing the Firefly luciferase driven by the adenovirus E1b minimal promoter and five Gal4 binding sites *(de la Luna et al., 1999)*.
- pG4-E1b-TATAless-Luc: reporter plasmid generated form pG4-E1b-Luc, in which the TATA-box was deleted by site-directed mutagenesis using the following primer: 5'-GGCTCGCCTCTGCAGACTAGACAGATTC-3'.
- -pCMV-RNL: Renilla luciferase expression vector (Invitrogen), in which the luciferase gene is under the control of the cytomegalovirus early promoter.

Plasmids for lentivectors preparation

- pCMV-VSV-G: lentiviral packaging vector that expresses the vesicular stomatitis virus G envelope protein *(Stewart et al., 2003)*. Obtained from Addgene (Plasmid #8454).
- pCMV-dR8.91: second generation packaging plasmids containing gag, pol, and rev genes *(Zufferey et al., 1997)*, kindly provided by D. Trono (Laboratory of Virology and Genetics, École Polytechnique Fédérale de Lausanne, Switzerland).
- pshControl: MISSION® pLKO.1-puro Non-Target shRNA Control Plasmid (Sigma #SHC016).
- pshDYRK1A-1: MISSION® pLKO.1-puro shRNA DYRK1A Plasmid (Sigma # TRCN0000022999).
- pshDYRK1A-2: MISSION® pLKO.1-puro shRNA DYRK1A Plasmid (Sigma # TRCN0000199464).

- pshKAISO-38: MISSION® pLKO.1-puro shRNA DYRK1A Plasmid (Sigma # TRCN0000017838).
- pshKAISO-40: MISSION® pLKO.1-puro shRNA DYRK1A Plasmid (Sigma # TRCN0000017840).

Techniques for DNA manipulation

DNA purification and sequencing

For small-scale purification, DNA was extracted from bacterial minicultures using the QIAGEN Plasmid Mini Kit (Qiagen) following manufacturer's instructions. Plasmid DNA purification for cell transfections was done by using the QIAGEN Plasmid Kit (Qiagen), following manufacturer's instructions.

For sequencing, DNA samples (500 ng/reaction) were sequenced at the Genomic Sequencing Service (Universitat Pompeu Fabra-PRBB, Barcelona) using the Big-Dye terminator 3.0 reagent (Applied Biosystems; 2 µl/reaction), 3.2 pmoles of primer (final vol, 10 µl), and the following PCR sequencing conditions: initial denaturation step (1 min, 94ºC) and 28 cycles of denaturation step (30 s, 95ºC), primer annealing (30 s, 55ºC) and extension step (4 min, 60ºC). PCR reactions were purified using Sephadex G-50 (Pharmacia) columns. Briefly, 800 µl of Sephadex-G50 equilibrated in H2O were added to a Centristep column and centrifuged at 2,600*xg* for 1 min. After discarding the supernatant, the column was washed with 10 µl of H2O; the PCR reaction was added and recovered by centrifugation. Finally, DNAs were dried in a Speed-vac for 10 min without heat.

Site-directed mutagenesis

Site-directed mutagenesis has been used to generate small deletions. The commercial kit QuickChange Multi Site-Directed Mutagenesis Kit (Stratagene) was used following manufacturer's instructions. The mutagenic primers were purchased with a phosphorylated 5'-end (Bonsai Technologies). All the mutants have been checked by DNA sequencing of the complete ORF sequence.

Cell Culture

Cell lines

The mammalian cell lines used in this work are as follows:
- HEK-293: epithelial cell line derived from human embryonic kidney.
- HeLa: epithelial cell line derived from human cervical adenocarcinoma.
- T98G: fibroblastic cell line derived from human glioblastoma.

All the cell lines were supplied by the American Type Culture Collection (http://www.atcc.org). Cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Invitrogen) containing 10% fetal bovine serum (FBS; Invitrogen) and supplemented with antibiotics (100 units/ml penicillin and 100 µg/ml streptomycin, Invitrogen) at 37ºC and in a 5% CO2 atmosphere.

For induction of G_0/G_1 growth arrest, T98G cells were plated at 40% confluency. One day after plating, cells were washed twice with phosphate-buffered saline (PBS) to remove residual FBS and incubated in serum-free medium for 48 h.

Cells serum-starved for 48 h were stimulated with insulin (100 nM, gift from M. Beato, CRG) for the times indicated in the Figure Legends before cell lysis.

Cell transfection

Cells were transfected with 1.3 µg (22-mm, MW12), 5 µg (60-mm plates) or 15 µg (100-mm plates) of DNA by the calcium phosphate precipitation method *(Graham and van der Eb, 1973)*. The optimal pH of the HEPES buffer for each cell line was tested by transfection with a green fluorescent protein-expressing construct and quantification of fluorescent cells. The DNA-calcium phosphate precipitate was removed after 16 h of incubation by washing the cells with PBS and adding fresh complete medium. Cells were washed with PBS at 48 h after transfection and processed according to the purpose of the experiment.

Preparation of lentivirus stocks and infection

The CRG has all the permits to work with lentivirus in a biological contention level 2 (A/ES/05/I-13 and A/ES/05/14). To generate a viral stock, HEK-293T cells were seeded at a density of $2x10^6$ in 100-mm

plates and transfected with 15 µg of pCMV- VSV-G, 10 µg of pCMV-dR8.91 packaging construct and 15 µg of the pLKO-based plasmids by the calcium phosphate precipitation method. Twenty-four h after transfection, fresh DMEM medium was added, and the lentivirus containing supernatant was harvested at 48 h and 72 h after transfection. The two supernatants were pooled together, spinned at 2,000*xg* for 10 min and filtered through a 0.45 µm filter (Millipore). Viral particles were concentrated by centrifugation (26,700 rpm in a Beckman Coulter centrifuge rotor SW32Ti for 90 min at 4ºC). The supernatant was discarded by inversion and the viral pellet was resuspended in PBS. Lentivector stocks were titered by infecting HEK-293 cells (10^5 cells in 60-mm plates) with a serial dilution of the viral stock. After 48 h in culture, infected cells were selected with 1 µg/ml puromycin (Sigma; 2 mg/ml in H_2O) during 8 days, and cell colonies were identified by staining with methylene blue (Sigma). The virus titer was calculated as follows: [number of cell colonies (TU) / viral volume used in the infection (il)]. The mean value was calculated considering the different viral dilutions.

For infection, cells were seeded at a density of $7x10^5$ in 60-mm plates in DMEM medium containing 5 µg/ml hexadimethrine bromide (Polybrene, Sigma). The virus was added to the medium and removed 24 h after infection. Cells were processed 96 h after infection for mRNA expression quantification or ChIP analysis. If required, puromycin was added at 1 µg/ml puromycin for 3 days.

FACs analysis of cell cycle parameters

Cells were harvested in PBS-5 mM EDTA (10^6 cells in 300 µl), fixed and permeabilized overnight with 700 µl 100% ethanol added dropwise at 4 °C. Cells were then washed twice in PBS, resuspended in 1 ml of PBS containing 0.5 µg/ml Ribonuclease A (Sigma), 50 µg/ml propidium iodide (Sigma) and 3.8 mM sodium citrate, and further incubated overnight at 4ºC. Samples were then analyzed with a FACSCan flow cytometer and the Cell Quest software (Becton Dickinson).

Cell volume determination

Cell volume determination was performed using the Beckman Coulter Z2 Cell and Particle Counter. In this system, the absolute cell volume of vital, non-fixed cells is measured in femtoliters. For these analyses,

logarithmically growing cells were harvested by trypsinization and cell pellets were resuspended in cold PBS (10^6 cells/ml). About 100 µl of the cell suspension were diluted in 10 ml of isotonic dilution buffer (COULTER Balanced Electrolyte Solution) and measured in triplicates. Discrimination between vital and dead cells/cell debris was carried out by using cell-type-dependent cursor settings.

Techniques for protein analysis

Preparation of cell lysates

For total cell lysates preparation, $1x10^6$ cells were resuspended in 150 µl 2x sample loading buffer (100 mM Tris-Cl pH 6.8, 200 mM dithiothreitol [DTT], 4% [w/v] sodium dodecil sulfate [SDS], 20% [v/v] glycerol, 0.2% [w/v] bromophenol blue) and heated for 10 min at 98ºC. Soluble extracts were prepared by incubating cells in lysis buffer A (50 mM HEPES, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% [v/v] Nonidet P-40 [NP-40], protease inhibitor cocktail [PIC; Roche Diagnostics], and phosphatase inhibitors [2 mM sodium orthovanadate, 30 mM sodium pyrophosphate, and 25 mM sodium fluoride]) for 15 min at 4ºC, followed by centrifugation at 13,000*xg* for 30 min at 4ºC.

Fractionation of HeLa cells was performed using NE-PER (Nuclear and Cytoplasmic Extraction Reagent; Pierce) according to the manufacturer's protocol.

Protein quantification was performed with the commercial kit BCA Protein Assay Kit (Pierce), following manufacturer's instructions.

Western blot analysis

Protein samples were resolved on SDS-polyacrylamide gels (SDS-PAGE) of different acrylamide percentage (depending on the molecular weight of the protein to study) at 30 mA in 1x running buffer (25 mM Tris-base, 200 mM glycine, 0.1% [w/v] SDS). Proteins were transferred onto Hybond-ECL nitrocellulose membranes (Amersham Biosciences) at 400 mA for 1 h in 1x transfer buffer (25 mM Tris-HCl pH 8.3, 200 mM glycine, 20% [v/v] methanol). Protein transfer was checked by staining with Ponceau S (Sigma).

Table MM1: Properties and working dilution of the primary and secondary (HRP-conjugated) antibodies used.

Primary antibody	Host	Working dilution	Commercial brand
Anti-Brd4	Rabbit	1:1000	Santa Cruz (H-250;sc-48772)
Anti-CDC5L	Mouse	1:1000	BD Transduction (#612362)
Anti-CDK9	Rabbit	1:1000	Santa Cruz (C-20; sc-484)
Anti-Cyclin T1	Goat	1:1000	Santa Cruz (T-18; sc-8127)
Anti-DYRK1A	Mouse	1:1000	Abnova (H00001859-M01)
Anti-G4DBD	Rabbit	1:500	Santa Cruz (DBD; sc-577)
Anti-GADPH	Mouse	1:1000	Millipore (MAB374)
Anti-HDAC2	Rabbit	1:1000	Abcam (ab7029)
Anti-HEXIM1	Rabbit	1:1000	Abcam (ab25388)
Anti-Histone H3	Rabbit	1:500	Abcam (ab1791)
Anti-KAISO	Mouse	1:1000	Santa Cruz (6F8; sc-23871)
Anti-Lamin B1	Rabbit	1:500	Santa Cruz (H-90; sc-20682)
Anti-p70S6K	Rabbit	1:500	Cell Signaling (#9202)
Anti-p70S6K-p(Thr389)	Rabbit	1:500	Cell Signaling (#9205)
Anti-PLRG1	Goat	1:2000	Abnova (PAB6081)
Anti-hPrp19	Rabbit	1:1000	Bethyl (A300-101A)
Anti-RNA Pol II	Rabbit	1:1000	Santa Cruz (N-20; sc-899)
Anti-RpAp48	Rabbit	1:1000	Abcam (ab79416)
Anti-SPF27	Mouse	1:1000	Novus (H00010286-M01)
Anti-TAF15	Rabbit	1:2000	Gift from L. Di Croce (CRG)
Anti-α-Tubulin	Mouse	1:5000	Sigma (T6199)
Anti-Vinculin	Mouse	1:10000	Sigma (V9131)

Secondary antibody	Host	Working dilution	Commercial brand
Anti-mouse	Rabbit	1:2000	Santa Cruz
Anti-rabbit	Goat	1:2000	Santa Cruz
Anti-goat	Rabbit	1:2000	Santa Cruz

Transferred membranes were blocked with 10% (w/v) skimmed milk (Sigma) diluted in TBS (10 mM Tris-HCl pH 7.5, 100 mM NaCl) - 0.1% (v/v) Tween-20 (TBS-T) for 30 min at room temperature (RT), and later incubated with the primary antibody diluted in 5% (w/v) skimmed milk in TBS-T for 1 h at RT or 16 h at 4ºC (Table MM1). Four washes of 10 min each in TBS-T eluted non-bound primary antibody, and then the secondary antibody (horseradish peroxidase (HRP)-conjugated, Table MM1), was added diluted in 5% skimmed milk in TBS-T and incubation proceed at RT for 45 min. After four 10 min-washes in TBS-T, membranes were revealed with SuperSignal West Pico Chemiluminescent Substrate (Pierce), and finally exposed in a LAS-3000 image analyzer (Fuji PhotoFilm, Tokyo, Japan) with the LAS3000-pro software.

Immunoprecipitation assays

For co-immunoprecipitation assays, 700 µg of HeLa Nuclear Extracts (HNE, CIL Biotech) were incubated overnight at 4ºC with protein G-Magnetic beads (Invitrogen) pre-bound with 5 µg of mouse monoclonal antibody anti-DYRK1A (mDYRK1A; Abnova H00001859) or mouse IgGs (Santa Cruz sc-2025). When indicated, a rabbit polyclonal anti-DYRK1A (rDYRK1A; Abcam ab69811) was used and rabbit IgGs (Santa Cruz sc-2027) as controls. Beads were washed four times with lysis buffer A, adding 0.1% NP-40 in the three initial washes. Samples were resuspended in 6x sample loading buffer, resolved by SDS-PAGE and analyzed by immunoblotting or used for *in vitro* kinase assays.

In vitro kinase assays

DYRK1A immunocomplexes were incubated for 20 min at 30ºC in 20 µl of phosphorylation buffer (50 mM HEPES, pH 7.4, 10 mM MgCl2, 10 mM MnCl2, 1 mM DTT and 50 µM ATP) in the presence of 2.5 µCi [γ^{32}P-ATP] (3000 Ci/mmol). ^{32}P incorporation was determined by resolving samples in SDS-PAGE, and exposing the dried gel to a Phosphoimager screen or to a film. The amount of DYRK1A present in the immunocomplexes was normalized, as quantified by immunoblot with anti-DYRK1A antibody. Kinase activity of DYRK1A protein was determined with the peptide substrate DYRKtide *(Himpel et al., 2000)*. To determine the catalytic activity of endogenus DYRK1A, DYRK1A-immunocomplexes were incubated for 20 min at 30ºC in 20 µl of phosphorylation buffer containing 200 µM DYRKtide and 1 µCi [γ^{32}P-ATP] (3000 Ci/mmol). Reaction aliquots were dotted onto P81 Whatman phosphocellulose paper, and, after washing extensively with 5% orthophosphoric acid, counts were determined in a liquid scintillation counter (LS6500 Multi-Purpose Scintillation Counter, Beckman Coulter). Each data point was determined in triplicate.

Mass spectrometry analysis

Mass spectrometry (MS) analysis has been performed by Cristina Chivas and Henrik Molina at the Joint CRG/UPF Proteomics Unit. Endogenus DYRK1A immunocomplexes (3 mg of HNE immunoprecipitated with 10 µg of anti-DYRK1A monoclonal antibody; parallel control

immunoprecipitations were performed with 10 µg of normal mouse IgGs) were resolved by SDS-PAGE and stained with Coomassie blue. Bands were excised (together with the equivalent band position in the control immunoprecipitations), and proteins subjected to in-gel digestion using trypsin (Promega). For LC-MS/MS analysis, peptides were pre-concentrated and desalted using Zorbax 300SB-C18 cartridges (Agilent Technologies). Peptides were eluted for MS/MS analysis, via a homemade RP C18 column: 75-µm × 10-mm, 3-µm particle size (Reprosil), using a gradient delivered at 300 nL/min and increasing from 3% to 45% solvent B in 28 min: solvent A: 0.1% formic acid; solvent B: 0.1% formic acid in acetonitrile (1200-series, Agilent Technologies). Peptides were analyzed on a LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific) equipped with a Proxeon nano-ESI source. An electrospray voltage of 2,200 V and a capillary voltage of 14 V at 180°C were used. MS/MS data were queried against IPI_HUMAN database (version 3.52) using Mascot v2.2 (Matrix Science).

Glycerol gradient sedimentation

3 mg of HNE were applied to a 10 ml 10–40% glycerol gradient in 20 mM HEPES pH 7.9, 150 mM KCl, 0.2 mM EDTA, 0.1 % NP-40, plus PIC. After centrifugation at 33,000 rpm in a SW41 Ti rotor (Beckman) for 20 h at 4°C, 24 fractions (500 µl-each) were collected from the top of the gradient and 30 µl of each fraction was analyzed by SDS-PAGE and Western blot. Molecular mass standards (Gel Filtration High Molecular Weight Calibration Kit, Amersham) were used to identify fractionation profiles within the gradient; detection of the marker proteins was done by silver staining of SDS-PAGE gels using standard protocols.

Gel size exclusion chromatography

Size exclusion chromatography was performed on Superdex 200 HR 10/30 (Pharmacia) (column volume 24 ml). A Pharmacia fast-performance chromatography system was used. Briefly, 3 mg of HNE were separated at 4°C with a flow rate of 0.5 ml/min while 0.5-ml fractions were collected. The elution buffer was 20 mM HEPES (pH 7.9), 100 mM KCl, 1 mM DTT, 20% (v/v) glycerol, and 0.2 mM EDTA. An aliquot (30 µl) of each fraction was then subjected to SDS-PAGE and Western blot analysis. Molecular mass standards (Gel Filtration High Molecular Weight Calibration Kit, Amersham) were used to calibrate the column.

Reporter assays

HEK-293 cells were seeded into 12-well (22-mm) plates ($8x10^4$ cells/well). After 24 h, cells were transfected by the calcium phosphate precipitate method with 100 ng/sample of the reporter (pG5HIV-Luc, pG5E1b-Luc or pG5E1b-LucTATAless), along with the effector plasmids as indicated in the Figure legends. The total DNA concentration (4 µg/sample) was kept constant by supplementing with pcDNA3. Cells were harvested in 150 µl Passive lysis buffer (Promega). Luciferase activity was determined with the Dual-Luciferase Reporter Assay System (Promega). pCMV-RNL (80 ng/sample) encoding Renilla luciferase was used to normalize transfection efficiencies. Each data point was determined in triplicate.

Electrophoretic Mobility-Shift Assay (EMSA)

To generate the double strand DNA probe, 50 pmol of single-stranded DNA-oligonucleotides were annealed, and 1.75 pmol were labeled with $\gamma[^{32}P]$ATP using T4 polynucleotide kinase (New England Biolabs). Unincorporated nucleotides were removed by centrifugation through Sepharose G-25 resin. The binding reactions were performed at 20°C for 20 min with 0.03 pmol of the labeled probe, 5–10 µg of nuclear cell extracts, 30 ng of poly-dI-dC (Roche Diagnostics) as a non-specific competitor, in a final volume of 10 µl of Gel Shift Binding Buffer (250 mM Tris-HCl pH7.5, 250 mM NaCl, 5 mM $MgCl_2$, 5 mM DTT, 1 mM EDTA, 50% glycerol). For the “competition” assays, a 40, 80 or 160-fold molar excess of cold wild type or mutated double-stranded DNA-oligonucleotides were added. For the “supershift” assay, 2 µg of either mouse IgG (Santa Cruz), anti-DYRK1A form Abnova (mDyrk1A) or anti-DYRK1A form Abcam (rDyrk1A) antibodies were pre-incubated with the extracts 15 min on ice before addition of the DNA probe. Binding reactions were analyzed on 4% acrylamide gels in 0.5X Tris/Borate/EDTA (TBE) buffer on ice.

The DNA-oligonucleotides probes used were (consensus sequence wild type and mutated are highlighted in red):

DYRK1A-WT-F: 5’-GGCCTAAGACTCTCGCGAGACACCGTCTAG-3’
DYRK1A-WT-R: 3’-CCGGATTCTGAGAGCGCTCTGTGGCAGATC-5’

DYRK1A-MUT-F: 5'-GGCCTAAGACAGGTGTACAACACCGTCTAG-3'
DYRK1A-MUT-R: 3'-CCGGATTCTGTCCACATGTTGTGGCAGATC-5'

Genome-wide association study of DYRK1A-bound chromatin regions

Chromatin immunoprecipitation (ChIP) assay

ChIP assays were performed from approximately 10^7 HeLa or T98G per experiment. Briefly, formaldehyde was added to the culture medium to a final concentration of 1% for 20 min at RT. Crosslinking was quenched with 0.125 M glycine for 5 min. Crosslinked cells were washed twice with TBS, resuspended in 1 ml of Buffer I (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40 plus PIC) and incubated on ice for 10 min. Cells were centrifuged (800x*g* for 5 min) and the cell pellet was resuspended in 1.2 ml of Buffer II (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0, 1 mM phenylmethylsulfonyl fluoride and PIC) and incubated for additional 10 min on ice. Samples were centrifuged (800x*g* for 5 min) and the cell pellet was then resuspended in 1.2 ml of IP Buffer (100 mM Tris-HCl pH 8.6, 0.3% SDS, 1.7% Triton X-100, and 5 mM EDTA). Chromatin was sonicated to an average size of 0.2–0.5 kb using Bioruptor (Diagenode), and the soluble material was quantified by BCA. For each immunoprecipitation, 600 µg of total protein extract diluted in 1 ml of IP Buffer was used, while a 60 µg-aliquot (10%) was kept to be used as input DNA. Samples were immunoprecipitated overnight at 4°C with antibodies specific for DYRK1A (5 µg, rabbit polyclonal, Abcam ab69811), KAISO (10 µg, rabbit polyclonal, Bethyl A300-000A) or mouse IgG (5 µg, Santa Cruz, sc-2025). Immune complexes were recovered by adding 40 µl of protein A or protein G sepharose beads (depending on the antibody species and Ig isotype) and incubated for additional 4 h at 4°C. Beads were washed with three successive 1 ml-washes of Low salt Buffer I (50 mM HEPES pH 7.5, 140 mM NaCl, 1% Triton X-100 plus PIC), and one wash with High salt Buffer II (50 mM HEPES at pH 7.5, 500 mM NaCl, 1% Triton X-100 plus PIC). Elution of DNA from beads was achieved by using Elution Buffer (1% SDS, 100 mM $NaHCO_3$) for 1 h at 65ºC with constant agitation (1,000 rpm) twice. The crosslinking was reverted by an additional incubation of the DNA at 65ºC overnight at 1,000 rpm. The eluted material

was phenol/chloroform extracted and ethanol-precipitated. ChIP DNA was resuspended in 50 µl of H_2O for further analysis.

For the ChiPSeq analysis, DYRK1A-bound chromatin from 10^8 cells were prepared. Library construction was carried out at the Genomics Unit of the CRG according to established protocols at the Unit. Libraries were sequenced to a length of 36 bp on an Illumina GAIIx sequencer, and around $25x10^6$ reads were obtained for each library with more than 97% aligned reads in all cases.

Quantitative PCR (q-PCR)

For ChIP-qPCR, a 1/10 dilution of ChIP DNA was used; a 1/100 dilution of input DNA was used as standard for normalization. PCR reactions (10 µl) were performed with SYBR Green (Fermentas) in 384 well plates using the Roche LC-480 equipment (Roche Applied Science). Denaturation was at 95º for 5 min followed by 50 cycles of 15 sec 95º, 20 sec 60º, 20 sec 72º. Ct (threshold cycle) were calculated for each sample using the Relative Quantification 2nd Derivative Maximum method with the Lightcycler 480 1.2 software (Roche). No PCR products were observed in the absence of template. The sequences of the primers used for this work are given in Table MM2. All primer sets gave narrow single melting point curves that were checked at the end of each run.

For RT-qPCR, total RNA was isolated with the RNeasy Kit extraction (Quiagen), treated with DNase I (Ambion, 2U/µl), and RNA quantified with Nanodrop, Total RNA (500 ng) was subjected to cDNA synthesis using Superscript II Reverse Transcriptase (RT) (Invitrogen) as recommended by the manufacturer's instructions. qPCR analysis was performed using 1/10 dilution of cDNA as template in 20 µl reaction volumes using SYBR green (Roche), as described in the previous paragraph. A "no-RT" control was included in each experiment to ensure that PCR products were not from contaminating DNA.

Table MM2: Sequences of oligonucleotides used for ChIP-qPCR experiments.

Region (gene)	Forward primer	Reverse primer
chr19_800	ACGCCCGGCTAATTTTTGT	TGAGGTCAGGGGTTCAACA
chr19_1000	GCCACACATGCTCCTAGCAC	GCTCTGTTCGTTCAGCCTTC
chr19_1200	GAAGGCTGAACGAACAGAGC	CGTACGGAGTCCATCCTGTT
chr19_1400	GGTTTAAGTGCAGCCTGTCAA	ACGTAAGGAGAAACGCAGCA
chr19_1800	GAGGCTCAAGCGTTTTAGGA	CCGTTTCCACAGAATTACCC
RBM39	AATTTGAGCGGCCGAAGTAT	GAATGGGGGATGGGAATATC
RPS11	GCTGAAGGCTGGTCACATCT	GGGCACTGTGAAGGACTGAC
ASXL1	AGCATCGCCTCCCAGAAT	CACCGACCTCAGCTAGGAAC
CDK12	TGATAAGCAGGGGAATGAGG	CTCCCTCACACAGACCCAGT
DENR	ACGCTCCGCAATTTTTCTC	CTCCCGCGAGACAATGAG
LUC7L2	CCGGGAGGGAATGTAATGTA	CTCCTCCCGCCCCTTTAC
CDC5L	CTCTGCCACTCGGTGACG	CGCATGTCCAAAACAGAATG
SMEK2	GGGAGTACTTCGGCGAGAC	GGGAGATCGCGAGAACCT
RPL12	GCGGACAAGCCAGATATAGG	CTGCCCACAACAAACATGG
RBM15	GAAAAAGGGGGTCGAAAGAG	AACCCGTCCTTTAAACCACTA
EIF5	GCCCAAACCGACAAAGAAC	GTCTCGCTTCCTTCCTCTCC
chr19_cov0	ATCTTGGTTCACCGCAACCT	AATTAGCTGGGTGTGGTGGT
chr20_cov0	GCTTGGCCAACAAGGTAAAA	CTCTCTGCAACCTCCACCTC

Table MM3: Sequences of oligonucleotides used for RT-qPCR experiments.

Gene	Forward primer	Reverse primer
RPS11	GCTTCAAGACACCCAAGGAG	TGACAATGGTCCTCTGCATC
LUC7L2	TCGTCAACGAATCAAATTCAG	ATCCGCTCTTAAAGCCAGGT
ASXL1	TGCAGGTCATAGAGGCAGAA	GAGCGTGAAAAGGCTGATTC
CDC5L	AGCTGCCCAAAGAGACAATG	TGGCTTCAGAAAGCATCTCA
DENR	CGATTACCCACTTCGAGTCC	CAGCTTCTTGTTTGGGTGAA
RPS6	GCTAGCAGAATCCGCAAACT	CTGCAGGACACGTGGAGTAA
CDK12	TCCCACCCTTATTACCTGGA	AGCTCTGGAGGGAGAGGAAG
RICTOR	TTGCAGTTGGAAATGATGGA	TGGCAATCGATCACATTTTT
RNU6	CTGGCTTCGGCAGCACATATAC	TATCGAACGCTTCACGAATTTGC
chr6.tRNA2.MetCAT	CTGGGCCCATAACCCAGAG	TAGCAGAGGATGGTTTC
chr8.tRNA10.MetCAT	CTCGTTAGCGCAGTAGGTAGC	GGTCGAACTCACGACCTTC

Each sample was assayed in triplicate, and values calculated as described above. All results were normalized to GAPDH or Actin expression. The sequence of the primers used in this work is provided in Table MM3. All primer sets gave narrow single melting point curves that were checked at the end of each run.

Bioinformatic analysis and Databases

All the bioinformatic analysis on the ChiP-Seq data regarding quality control, read mapping and peak identification was done by our

collaborators Daniela Bezdan and Stephan Ossoswki at the CRG. ChiP-Seq mapping was done to the human reference genome hg19.

For peak annotation an in-house gene and promoter annotation pipeline (ChIPanno) developed by S. Ossowski has been used. ChIPanno uses gene, transcript, promoter and TSS annotation data from Ensembl, SwitchGear-TSS (http://www.switchgeargenomics.com/), National Center for Biotechnology Information (NCBI) RefSeq, and University of California Santa Cruz (UCSC) genome browser database *(Landolin et al., 2010; Meyer et al., 2013; Trinklein et al., 2003)* via the UCSC Table Browser *(Karolchik et al., 2004)*. A promoter definition provided by the annotation tool Homer *(Heinz et al., 2010)* was also included.

Bioinformatic analysis of DYRK1A-associated regions, including chromosomal and genome distribution of ChIP regions, peak profile around TSS, histone modification profile of DYRK1A-associated regions, conservation of identified palindromic consensus –TCTCGCGAGA- and *in silico* expression analysis of DYRK1A-associated targets, has been done by K. Islam at N. Lopez Bigas' lab (Biomedical Genomics Group, University Pompeu Fabra, Barcelona, Spain). When required, a brief description has been included in the corresponding Figure Legends or the text.

For *de novo* identification of motifs in peak regions and associated promoters the MEME suite has been used, in collaboration with D. Bezdan *(Bailey et al., 2009)* (http://meme.nbcr.net/meme/).

All the analysis regarding the overlap between DYRK1A- and KAISO ChIP-Seq were done by D. Bezdan. BED files of KAISO ChIP-Seq were kindly provided by Jie Wang (Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, USA) *(Wang et al., 2012)*.

Gene Ontology analysis of DYRK1A-associated genes was performed using Ingenuity Pathway Analysis (IPA) or DAVID Bioinformatic Resources *(Huang da et al., 2009)* (http://david.abcc.ncifcrf.gov/). Microarray data was obtained from the Gene Expression Omnibus at NCBI (http://www.ncbi.nlm.nih.gov/geo/). Gene Set Enrichment Analysis (GSEA) was done using the Molecular Signatures Database (MSigDB) (http://www.broadinstitute.org/gsea/index.js).

Results

DYRK1A is present in the nucleus as an active kinase

As mentioned in the Introduction section, the protein kinase DYRK1A has been detected in both the nucleus and the cytosol of different types of mammalian cells by immunostaining techniques. However, little is known about the enzymatic activity and the functional roles of DYRK1A within the nuclear compartment *in vivo*.

To verify the presence of DYRK1A in both the cytosolic and nuclear compartments, HeLa cells were subjected to nuclear-cytoplasmic fractionation and the purity of each fraction assessed by the use of marker proteins (GAPDH for the cytosol and lamin B1 for the nuclear extracts; see Fig. R1A). Western blot analysis of protein extracts from both fractions revealed that, although most of DYRK1A accumulates within the cytoplasm, a small pool of the kinase is detected specifically in the nuclear compartment (Fig. R1A).

We first checked whether DYRK1A behaves as an active kinase within the nuclear compartment. To assess DYRK1A kinase activity on exogenous substrates, activity was tested using an optimized peptide substrate for DYRK1A, DYRKtide *(Himpel et al., 2000)*. Nuclear DYRK1A was immunoprecipitated from commercial HeLa nuclear extracts (HNE) using a specific antibody, and then immunocomplexes were subjected to an *IVK* assay: ^{32}P incorporation into DYRKtide was measured during a 30 min reaction in parallel reactions with control immunoprecipitates. As shown in Fig. R1B, DYRK1A is immunopurified from nuclear extracts as an active kinase.

The extracts from the IVK assay using DYRK1A-specific complexes immunoprecipitated from HNE were also analyzed by SDS-electrophoresis and autoradiography (Fig. R1C). The results showed the presence of a strong radioactive signal at around 95 kDa, corresponding most probably to autophosphorylated DYRK1A. Several radioactive signals were also observed at different molecular weights when DYRK1A immunocomplexes were compared to IgG control immunoprecipitates (Fig. R1C, lane 1 vs 3). To assure that the phosphorylation pattern was dependent on DYRK1A and not on other kinases present in the immunoprecipitates, we took advantage of harmine, an inhibitor of DYRK1A kinase activity *(Bain et al., 2007)*.

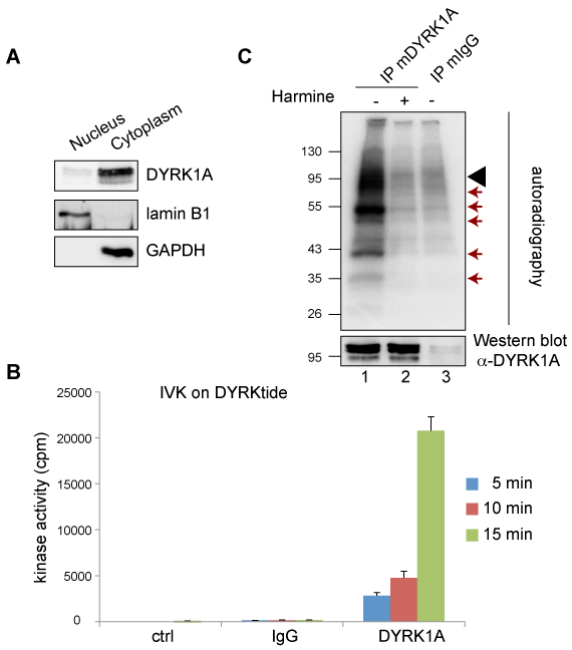


Figure R1: Nuclear DYRK1A behaves as an active kinase. (A) HeLa cells were fractionated into cytoplasmic/membrane and nuclear fractions. Equal percentage of each fraction was analyzed by Western blot with antibodies to the indicated proteins. **(B)** DYRK1A kinase activity was determined in DYRK1A-immunocomplexes using DYRKtide as substrate in a radioactive *IVK* assay, at three different time points (indicated on the right). Immunocomplexes obtained with normal IgG were used as a control (IgG); radioactivity incorporated on DYRKtide was also measured in absence of any protein complex to measure background signal (ctrl). **(C)** Anti-DYRK1A and anti-IgG immunoprecipitates (lane 1 and 3, respectively) from HNE were subjected to an *IVK* assay in the presence of radioactive ATP. Proteins were separated by SDS-PAGE and analyzed by autoradiography. DYRK1A-immunoprecipitates were also pre-incubated with harmine (10 μM) for 30 min before preforming the *IVK* assay (lane 2). Equivalent aliquots of each immunoprecipitate were analyzed by Western blot with an anti-DYRK1A antibody to check for equal loading. The black arrow point to the position of DYRK1A in the gel. The red arrows indicate protein bands that become phosphorylated in the DYRK1A-immunocomplexes.

When the IVK assays were performed in the presence of harmine, the phosphorylation profile of DYRK1A-immunocomplexes was indistinguishable from the one obtained with the IgG-control immunocomplexes (Fig. R1C, lane 2 vs 3), suggesting that the phosphorylation observed in the DYRK1A-immunoprecipitates is due to specific DYRK1A activity. All these results allow us to conclude that nuclear DYRK1A acts as an active kinase and furthermore, they suggest that there are specific substrates for DYRK1A within the nucleus.

DYRK1A is part of macromolecular complexes within the nucleus

The existence of DYRK1A substrates in DYRK1A-specific immunocomplexes led us to hypothesize that DYRK1A might participate in different nuclear complexes. To test this hypothesis, HNE were subjected to centrifugation in a 10 to 40% glycerol gradient that will separate protein complexes upon the basis of size. Aliquots from different fractions of the gradient were resolved on SDS-PAGE, and the distribution of DYRK1A along the fractions was assessed by Western Blot. The detection of DYRK1A in fractions corresponding to its molecular weight (Fig. R2A, lanes 1-7) is an indication of the existence of a soluble monomeric pool of DYRK1A. However, a pool of DYRK1A was also found in high molecular weight fractions (Fig. R2A, lanes 8-22), suggesting that DYRK1A could be part of macromolecular protein complexes.

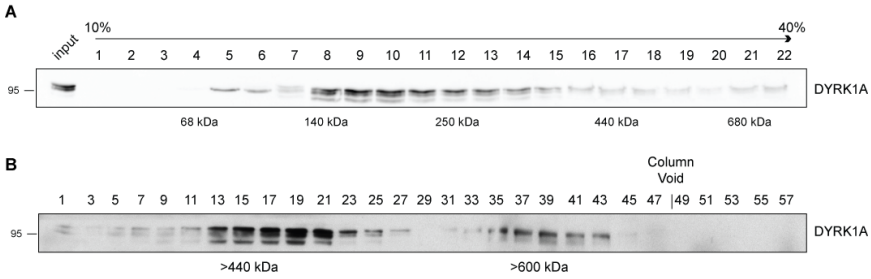


Figure R2: DYRK1A fractionates with high molecular weight complexes within the nucleus. (A) HNE were fractionated by sedimentation through a 10–40% glycerol gradient. Twenty-two protein factions were collected then from the top of the gradient (numbers are indicated above), and 30 µl of each analyzed by Western blot with an anti-DYRK1A antibody. The molecular weights corresponding to the indicated fractions were assigned according to the fractioning profile of molecular mass standards analyzed in parallel. **(B)** HNE were subjected to gel size exclusion chromatography using a Sephadex 10/300 column. Aliquots of the indicated fractions were analyzed by Western blot with an anti-DYRK1A antibody. The molecular weight of the two DYRK1A elution peaks is indicated below the panel. The experimentally determined exclusion limit of the columns is also indicated (Void).

To further confirm this result, and to obtain a better resolution of DYRK1A distribution in high molecular weight complexes, HNE were fractionated by size exclusion chromatography using a Sephadex10/300 column. The distribution of DYRK1A in the column indicated that it accumulates in two different pools of high molecular weight complexes (Fig. R2B): one of them bigger than 400 kDa and another one of more

than 600 kDa, indicating that DYRK1A is participating in more than one high molecular weight complexes within the nucleus.

The DYRK1A nuclear interactome

With the aim of characterizing the possible macromolecular complexes containing DYRK1A, we implemented an approach based on affinity purification coupled to MS identification. Fig. R3A shows the experimental design: i) HNE were immunoprecipitated either with an antibody specific for DYRK1A or with normal IgGs, used as control; ii) the immunoprecipitated protein complexes were subjected to SDS-PAGE and Coomassie staining; iii) selected bands from specific immunoprecipitates were cut from the gel in parallel with their respective band in the control immunoprecipitation; and iv) the bands were subjected to standard gel-based protein identification by LC-MS/MS at the Proteomics Unit of the CRG/UPF following established protocols at the Core Facility. Four biological replicates of DYRK1A-immunocomplexes were subjected to MS analysis.

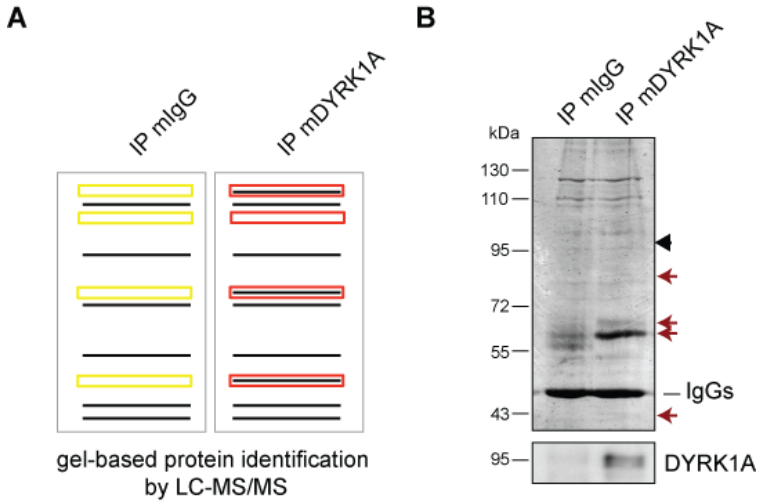


Figure R3: Mass spectrometry analysis of DYRK1A-associated proteins. (A) Scheme for the experimental design as explained in the text. **(B)** Coomassie staining of mDYRK1A- and mIgG- purified immunocomplexes. Black arrow indicates the band corresponding to DYRK1A, which was faintly detected only in the specific immunoprecipitate. Red arrows indicate some of the bands analyzed by MS. The band corresponding to the immunoglobulin heavy chains is marked (IgGs). The presence of DYRK1A in the specific immunoprecipitate was controlled by Western blot analysis with an anti-DYRK1A antibody (in the bottom panel).

We first tested several available commercial anti-DYRK1A antibodies and found a mouse monoclonal (mDYRK1A) that behaved as the most efficient in immunoprecipitating DYRK1A from HNE (data not shown). The presence of DYRK1A, as a faint enriched band, in the specific immunoprecipitates was confirmed by MS analysis in all experiments, and the enrichment of certain bands was clearly observed in the DYRK1A immunoprecipitates (Fig. 3RB, red arrows). Despite this observation, several common bands were present in both immunoprecipitates, likely due to unspecific trapping of proteins as a result of the use of highly concentrated HNE to enrich for the low abundant DYRK1A protein in the nuclear fraction.

To reduce the number of false positive hits, the putative DYRK1A interacting candidates resulted after MS analysis were manually filtered with a series of criteria. First, true positives should be present only in the specific DYRK1A-IP and not in the IgG-IP; second, they should migrate at the expected size according to their molecular weight; third, they should appear at least in two out of four experiments; last, they should not belong to the list of usual contaminants as described by Lamond's group in this type of experiments *(Trinkle-Mulcahy et al., 2008)*.

We identified a total of 195 proteins that passed the selection criteria and thus have been considered as part of the putative DYRK1A interactome for subsequent analysis. Assuring the good quality of the identification and selection process, several already described DYRK1A partners were identified, as 14-3-3 *(Alvarez et al., 2007; Kim et al., 2004)* and DCAF7 *(Skurat and Dietrich, 2004)*. Ingenuity Pathway Analysis (IPA) of DYRK1A nuclear interactome revealed a strong enrichment of proteins related to RNA processing and gene transcription (Table R1).

Table R1: IPA analysis of the DYRK1A nuclear interactome

Molecular and Cellular Functions	p-Value	Counts[1]
RNA Post-Transcriptional Modification	6.87E-23	32
Gene Expression	6.18E-09	55
Cell Cycle	5.32E-07	33
DNA Replication and Repair	5.32E-07	26
Cellular Assembly and Organization	1.44E-06	28
Canonical Pathways		
Cleavage and Polyadenylation of mRNA	6.84E-10	6
Assembly of RNA Polymerase Complex	1.05E-06	4
DNA Methylation	9.17E-04	3

[1]: number of proteins in the category

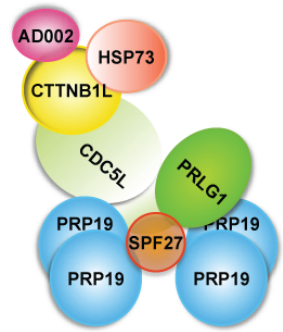
mRNA processing related DYRK1A-interacting proteins

The category in IPA analysis found to be the most enriched was "*RNA post-transcriptional modification*". Links between DYRK1A and the splicing process have been already established in the past years. In fact, several splicing factors have been described to be phosphorylated by DYRK1A including SRSF1/ASF *(Shi et al., 2008)*, SRSF2/SC35 *(Qian et al., 2011)* or SFRS7/9G8 *(Ding et al., 2012)*, although no evidences for interaction in physiological conditions have been reported for any of these factors. Some of these splicing factors, such as ASF and SC35, were found to be enriched in the MS analysis. However, a more consistent and stronger enrichment was observed for numerous components of the splicing machinery itself. Particularly, six out of the nine components of the hPrp19/CDC5L complex (Grote et al., 2010) were identified (Table R2 and scheme included).

Table R2: Composition of the hPrp19/CDC5L complex

Component	Gene	Protein[1]
Core	BCAS2	SPF27
	PLRG1	PLRG1
	PRPF19	hPrp19
	CDC5L	CDC5L
	CWC15	AD002
	CTNNBL1	CTNNBL1
	HSPA8	HSP73
Others	SRRM2	CWF21/SRm300
	XAB2	NTC90

[1]the proteins found in the screen are in pink



hPrp19/CDC5L complex organization

MS analysis of immunoaffinity-purified human Prp19/CDC5L complexes showed that they consist of seven core proteins: hPrp19, CDC5L, PRL1, AD002, SPF27, CTNNBL1 (β-catenin-like 1), and HSP73 (see scheme associated to Table R3) plus many other associated proteins *(Makarova et al., 2004)*.The hPrp19/CDC5L complex is part of the spliceosome, and it participates in the activation steps of the splicing reaction *(Ajuh et al., 2000; Grillari et al., 2005; Grote et al., 2010)*. When analyzed by sedimentation in a glycerol gradient, DYRK1A fractionates together with SPF27, PLRG1, hPrp19 and CDC5L (Fig. R4A), supporting the idea that they may be part of the same protein complex. The detected interaction between DYRK1A and the core components hPrp19 and PLRG1 was validated by co-immunoprecipitation experiments followed by Western blot using specific antibodies against selected proteins (Fig. R4B). Noteworthy, we did detect the presence of another core component of the hPrp19/CDC5L complex, CDC5L, which was not present in the proteomic screen (Fig. R4B), supporting the proposal that DYRK1A could associate with the hPrp19 complex *in vivo*.

Previous work from the group had shown that DYRK1A is able to bind single strand nucleic acids, both DNA and RNA, *in vitro* (E. Salichs, Doctoral Thesis, 2008). Thus, the possibility of DYRK1A being recruited to the hPrp19 complex through RNA molecules used as scaffolds was evaluated. To this aim, co-immunoprecipitation experiments were performed using HNE treated with RNAse. DYRK1A was able to interact with hPrp19 independently of the presence of RNA (Fig. R4C), suggesting that DYRK1A has to be recruited to the complex via protein-protein interactions, and furthermore, that the formation of the complex could be a step prior the association to the spliceosome.

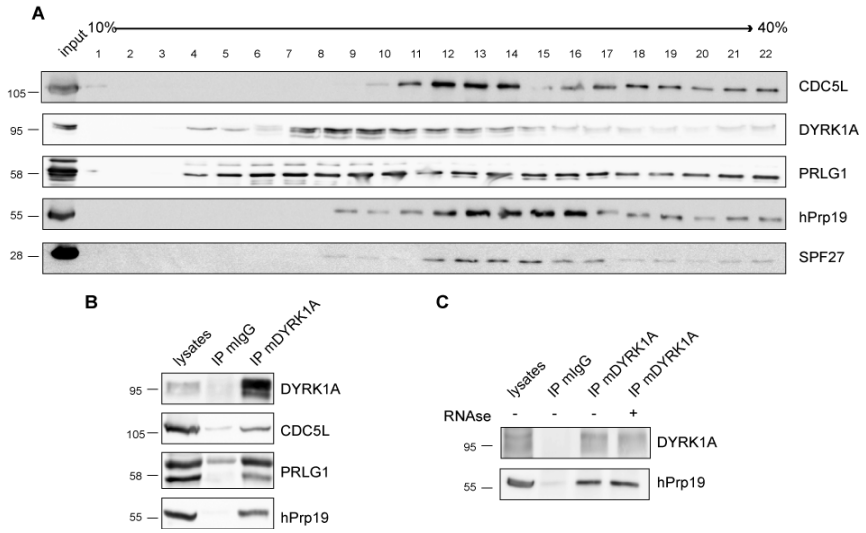


Figure R3: DYRK1A interacts with the hPrp19/CDC5L complex in a RNA-independent manner. (A) HNE were fractionated on a 10-40% glycerol gradient, and aliquots of the fractions analyzed by Western blot with antibodies to the indicated proteins. **(B)** HNE were immunoprecipitated with anti-mDYRK1A or normal mouse IgGs (mIgG), as a control and both the lysates (10%) and the immunocomplexes were analyzed by immunoblotting with antibodies to the indicated proteins. **(C)** HNE were treated with or without RNAse (1 mg/ml) and a coimmunoprecipitattion was performed as in **B**.

Processing of mRNA 3′-end is an essential maturation step that increases the stability of mRNA, facilitates its export from the nucleus to the cytoplasm, and enhances translation efficiency (reviewed in *Mandel et al., 2008)*. The process, which involves several steps including recognition of the polyadenylation signal, cleavage of the 3'-end of the pre-mRNA and addition of a polyA tail, requires the participation of several multisubnit complexes such as CPSF, CstF CFI_m and $CFII_m$ (schematic representation in Fig. R5). Several components of the CFI_m and the CPSF protein complexes, as indicated in Table R3, were found to be enriched in the DYRK1A proteomic screen.

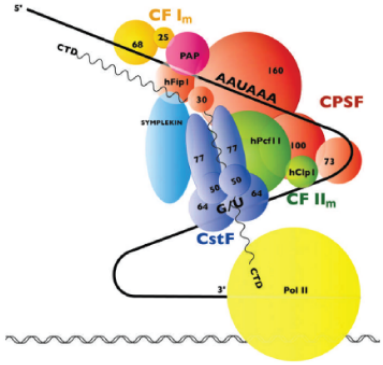


Table R3: 3'-end mRNA processing factors found in the DYRK1A nuclear interactome

Complex	Protein
CF Im	CF I 25
	CF I 68
CPSF	Symplekin
	hFIP1
	PABPN1
	FEN1
	CPSF160
	CPSF30

Figure R5: The mammalian pre-mRNA 3'-end processing machinery. Schematic representation of the protein complexes involved in pre-mRNA 3' processing. The elongating RNA Pol II is in yellow, while the other complexes are associated by colors: cleavage stimulation factor complex (CstF, in blue); cleavage factor I and II (CF I and CF II, in yellow and in green, respectively); cleavage and polyadenylation specificity factor (CPSF, in red); Symplekin (in light blue) and the poly(A) polymerase (PAP, in violet). Image adapted from *(Mandel et al., 2008)*. The proteins found in the DYRK1A proteomic screen are listed in the accompanied Table R3.

Unfortunately, due to technical problems related to the specificity of the antibodies available, none of these putative interactions could be validated *in vivo*. Further analysis is thus needed to rule out whether DYRK1A has a role in the regulation on any of these complexes *in vivo*.

Chromatin remodeling DYRK1A-interacting proteins

Another class of proteins that was enriched in the DYRK1A proteomic screen was "*Gene expression related factors*". In particular, three proteins related to chromatin organization and remodeling were identified: the class I histone deacetylase HDAC2, and the retinoblastoma-interacting proteins RbAp46/RBBP7 and RbAp48/RBBP4, which are usually part of co-repressor complexes. The three proteins, together with HDAC1, are the subunits of the core histone deacetylase (HDAC) complex *(Zhang et al., 1999)*. They are part of the catalytic core of many multiprotein transcriptional complexes *(Havugimana et al., 2012)*, such as the Sin3a and NurD complexes *(McDonel et al., 2009; Xue et al., 1998)* (see scheme in Fig. R6A). RbAp46 and RbAp48 also belong to the polycomb repressive PRC2/EED-EZH2 complex *(Kuzmichev et al., 2002)*, which promotes repression of homeotic genes during development, and to the nucleosome remodeling factor, NURF, complex *(Barak et al., 2003)*.

The interaction between DYRK1A and HDAC2 and RbAP48 was confirmed by immunoprecipitation experiments followed by Western blot with specific antibodies (Fig. R6).

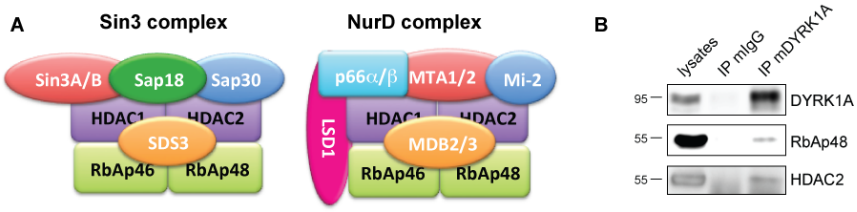


Figure R6: DYRK1A interacts with chromatin remodeling factors. (A) Schematic representation of the Sin3 and the NurD complexes. **(B)** HNE were immunoprecipitated with mDYRK1A antibody or normal mouse IgGs (mIgG) and both the lysates (10%) and the immunocomplexes analyzed by immunoblotting with antibodies against the proteins indicated.

Transcription initiation related DYRK1A-interacting factors

The majority of the factors belonging to the "*Gene expression*" category were however strictly related with the transcription process itself. As described in the Introduction section, initiation of transcription starts with the recognition of the promoter and the assembly of the PIC. Four out of sixteen components of the TFIID subunit of the PIC were present in the proteomic screen (TAF6/TAF2E/TAFII80, TAF9/TAF2G/TAFII32, TAF10/TAF2H/TAFII30 and TAF15/RBP56/TAF2N/TAFII68), together with the general transcription factor GTFII-I *(Havugimana et al., 2012)*. As shown in Fig. R7, TAF15, used as a marker of the PIC, co-immunoprecipitates with DYRK1A in nuclear extracts, supporting the proposal that DYRK1A could be part of gene-specific PICs *in vivo*.

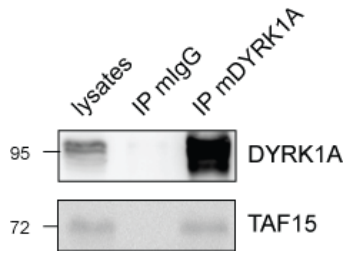


Figure R7: DYRK1A interacts with TAF15. HNE were immunoprecipitated with mDYRK1A antibody or normal mouse IgGs (mIgG) and both the lysates (10%) and the immunocomplexes analyzed by immunoblotting with antibodies against DYRK1A and TAF15.

Previous experimental evidence from the group indicated that DYRK1A interacts with and phosphorylates the CTD of the RNA Pol II *in vitro* (E. Salichs, Doctoral Thesis, 2008). Given the presence of components of the PIC in the DYRK1A immunocomplexes, we wonder whether the interaction with the RNA Pol II would be also occurring *in vivo*. First, we analyzed the distribution of the two proteins in a glycerol gradient sedimentation profile and observed that there is a pool of DYRK1A that distribute in the same high molecular fractions as RNA Pol II (Fig. R8A), suggesting that both proteins could be part of the same complexes *in vivo*. In fact, RNA Pol II is co-immunoprecipitated with nuclear DYRK1A (Fig. R8B).

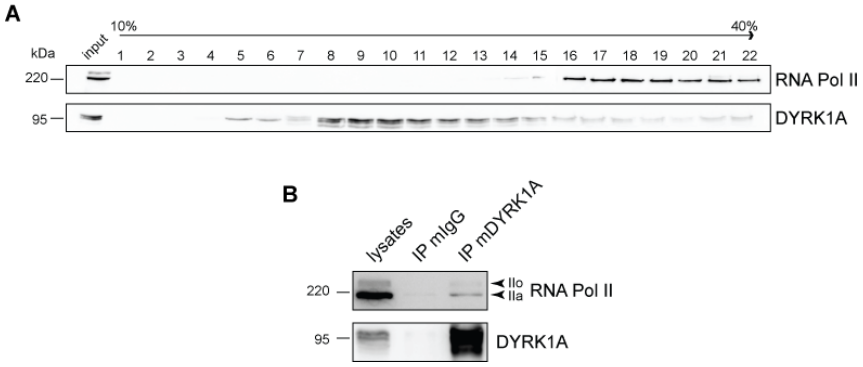


Figure R8: DYRK1A is associated with RNA Pol II-containing complexes *in vivo*. (A) HNE were fractionated on a 10-40% glycerol gradient, and aliquots of the different fractions were analyzed by Western blotting with antibodies specific for RNA Pol II and DYRK1A. **(B)** HNE were immunoprecipitated with mDYRK1A antibody or normal IgGs (mIgG) and both lysates and immunocomplexes were analyzed by immunoblotting with antibodies against DYRK1A and RNA Pol II (N20). Note that the RNA Pol II antibody used recognizes both forms of the polymerase, the hypo- and the hyper- phosphorylated (IIa and IIo, respectively). Both forms are present in DYRK1A-immunocomplexes.

Both the hypophosphorylated and the hyperphosphorylated forms of the RNA Pol II are detected in the DYRK1A immunoprecipitates, based on the presence of two bands of different electrophoretic mobility detected by an antibody targeting the N-terminus of RNA Pol II (Fig. R8B).Considering that DYRK1A interacts with the RNA Pol II, that it phosphorylates the CTD, and that it is able to interact with others components of the PICs, we propose that DYRK1A could somehow be functionally linked to the basal transcription process itself.

Transcription elongation related DYRK1A-interacting factors

After PIC formation and pausing, a set of different proteins participates in the release of the RNA Pol II, allowing its activation and the subsequent start of the transcription process. The major player of this process is pTEFb, a cyclin dependent kinase containing the catalytic subunit CDK9 and a regulatory subunit, which in mammals is one of several cyclin subunits, cyclin T1, T2 or K (for details see the Introduction section). Both CDK9 and cyclin T1 were identified as part of the DYRK1A nuclear interactome.

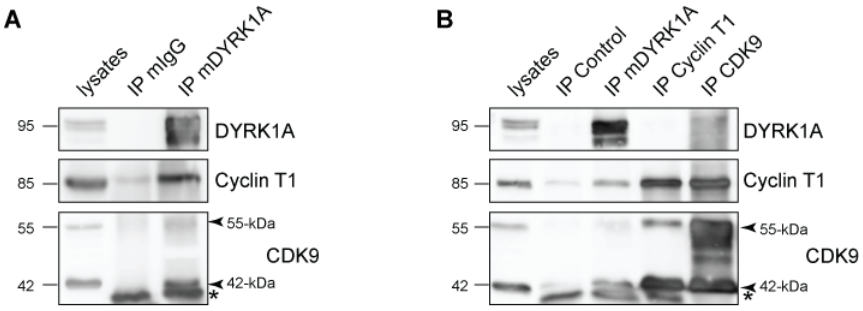


Figure R9: DYRK1A interacts with pTEFb. HNE were immunoprecipitated with mDYRK1A **(A, B)**, anti-Cyclin T1 or anti-CDK9 **(B)** antibodies or normal mouse IgGs (mIgG) **(A, B)** and both lysates (10%) and immunocomplexes were analyzed by immunoblotting with antibodies to the indicated proteins. Normal IgGs (IP Control) were used a control for specificity. .In **A** and **B,** the two CDK9 isoforms (55- and 42- kDa) were indicated with arrows. An asterisk indicates the immunoglobulin heavy chain.

The presence of CDK9 and cyclin T1 in the nuclear DYRK1A immunocomplexes was validated by co-immunoprecipitation experiments (Fig. R8A). There are two isoforms of CDK9 in mammalian cells, a major 42 kDa isoform and a minor 55 kDa isoform, which has a 117 residue amino terminal extension due to the use of an upstream transcriptional start site *(Shore et al., 2003)*. Both isoforms were detected in the immunoprecipitates (Fig. R9A). However, we could not detect any binding of DYRK1A to cyclin T1, and very faint to CDK9, in reverse co-immunoprecipitation assays (Fig. R9B). It could be possible that the antibodies used in the reverse co-immunoprecipitation interfere with the site of interaction for DYRK1A with any of the components of the cyclin T1/CDK9 complex, resulting in a loss of binding. Alternatively, the three proteins could be part of different sub-complexes, with DYRK1A being part of only a small subset of cyclin T1/CDK9 complex;

therefore, the enrichment of the DYRK1A-containing complexes is needed to detect the interaction, otherwise not appreciable.

To explore this possibility, the distribution of the CDK9/cyclin T1 complexes was analyzed by glycerol gradient sedimentation. pTEFb accumulates in two discrete waves, corresponding to the “active pTEFb” complex, or Complex I, identified by the presence of the activator protein Brd4, and to the “inactive pTEFb” complex or Complex II, which co-fractionates with its inhibitor HEXIM1 (Fig. R10). As shown in Fig. R10, DYRK1A is present in fractions corresponding to the two pTEFb type of complexes, although its accumulation is most noticeable in the fractions corresponding to the active pTEFb complexes.

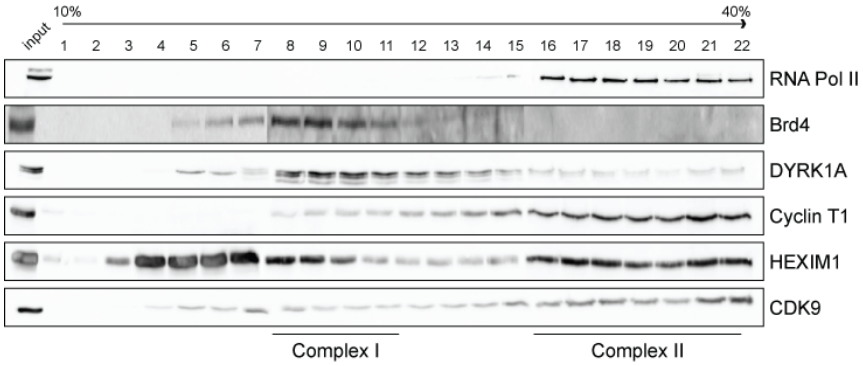


Figure R10: DYRK1A co-fractionates with pTEFb. (A) HNE were fractionated on a 10-40% glycerol gradient, and aliquots of the fractions were analyzed by Western blotting with antibodies specific for the proteins indicated. Two main CDK9-containing complexes were identified (indicated below the panels as Complex I and Complex II).

To understand if DYRK1A recruitment to pTEFb was differential within the two complexes, we tried to immunoprecipitate DYRK1A from the two different pools. Fractions corresponding to Complex I (fractions 7 to 12, around 100-300 kDa) and fractions corresponding to Complex II (fractions 16 to 22, 400-600 kDa) were collected and then immunoprecipitation with an anti-DYRK1A antibody was performed. Unfortunately, we were not able to immunoprecipitate DYRK1A from these pools, most probably due to the high concentration of glycerol within these fractions.

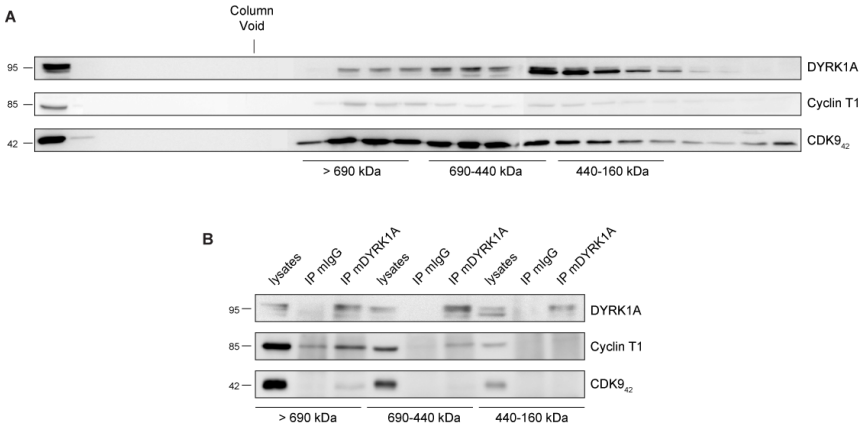


Figure R11: DYRK1A interacts with cyclin T1 in CDK9-free complexes. (A) HNE were subjected to gel size exclusion chromatography using a Sephadex 10/300 column. Aliquots from high molecular weight fractions were analyzed by Western blot with antibodies to the indicated proteins. The position of the main high molecular complexes are indicated with the corresponding profile of molecular weight markers elution (indicated below the panels). **(B)** Extracts corresponding to the three high molecular weight group of fractions indicated in **A** were immunoprecipitated with mDYRK1A antibody or normal mouse IgGs (mIgG) and both lysates (10%) and immunocomplexes were analyzed by immunoblotting with antibodies against DYRK1A, cyclin T1 and CDK9. Only the 42 kDa isoform of CDK9 is shown in both panels.

We thus performed gel size exclusion chromatography by implementing the Sephadex Column previously used to obtain the two different DYRK1A-containg protein complexes. Although the distribution of all proteins is slightly different from the pattern observed with the glycerol gradient separation, two complexes were also detected (Fig. R11A). High molecular weight fractions were then pooled accordingly with the distribution of molecular mass standards, and. DYRK1A binding to both CDK9 and cyclin T1 was tested by immunoprecipitation. The presence of the three proteins, DYRK1A-cyclin T1-CDK9, as suggestive of a ternary complex was observed only in those fractions corresponding to the higher molecular weight (Fig. R11B, left panel). Moreover, DYRK1A is able to recruit cyclin T1, but not CDK9, in fractions corresponding to complexes from 690-440 kDa (Fig.11B, central panel).

As mentioned previously, CDK9 is the catalytic subunit of pTEFb, responsible of the phosphorylation of the RNA Pol II CTD, while cyclin T1 is the regulatory subunit of the complex. Our data suggest that DYRK1A complexes containing cyclin T1 could exist, in which CDK9 could be substituted by DYRK1A that could act as the catalytic subunit

of a specific “pTEFb-like” complex, regulating the transcriptional elongation of a subset of target genes.

DYRK1A as a novel regulator of transcription

DYRK1A is able to induce transcription when recruited to promoters

Given the presence of DYRK1A in high molecular complexes containing transcription related proteins, and its ability to physically interact with RNA Pol II and with pTEFb, we decided to investigate whether DYRK1A can directly regulate transcription. To this aim, DYRK1A was targeted to a promoter by fusing it to a sequence-specific DNA binding domain (G4DBD; amino acids 1-147 from the yeast transcription factor Gal4), and the ability of the fusion protein to activate promoter elements containing appropriate GAL4-DNA binding sequences was tested in reporter assays.

We first used a luciferase reporter plasmid harboring the HIV-1 promoter preceded by 5 Gal4 binding sites driven luciferase transcription (G5HIV1-Luc; Fig. R12A); this reporter has been shown to be activated by a chimera G4DBD-CDK9 *(Montanuy et al., 2008)*. As shown in Fig. R12B, the fusion G4DBD-DYRK1A effectively activated transcription when tested on the G5HIV1-Luc reporter, even more efficiently than CDK9. In contrast, a fusion with a catalytically inactive DYRK1A (G4-DYRK1A^{KR}, ATP binding lysine residue 179 replaced by arginine) failed to activate the reporter construct (Fig. R12B). We therefore conclude that DYRK1A is able to activate transcription when recruited to a promoter and that the effect is dependent on DYRK1A kinase activity. Furthermore, the results suggest that the phosphorylation of either one of the transcription factors involved in the HIV-1 promoter activation and/or of a component of the basal transcription machinery could be responsible of the effect.

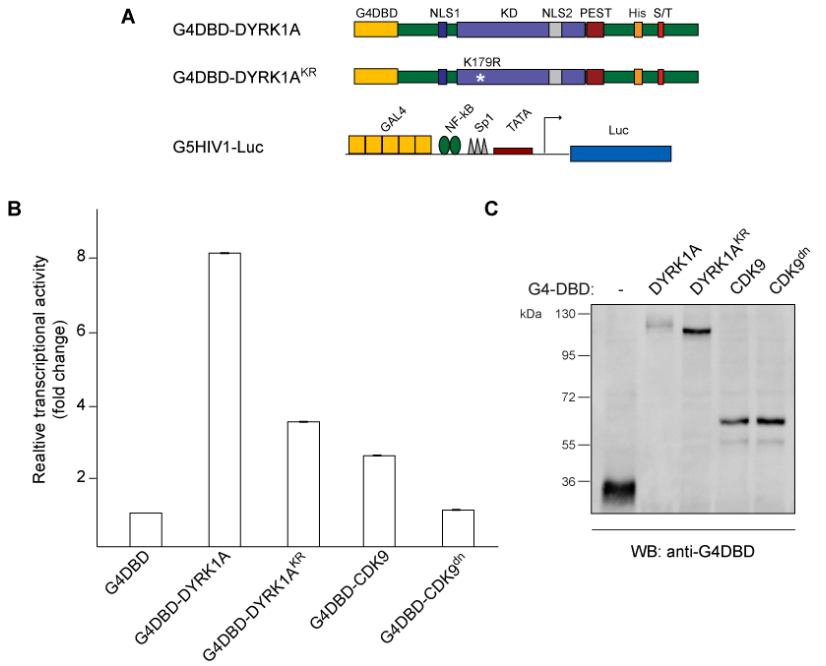


Figure R12: DYRK1A induces transcription in a kinase-dependent manner. (A) Schematic representation of the constructs used in the reporter assays. The effector plasmid G4DBD-DYRK1A directed the expression of the full-length DYRK1A protein fused to the Gal4 DNA binding domain (DBD) while G4DBD-DYRK1A^{KR} expresses a kinase-inactive DYRK1A. The target plasmid (G5HIV1-Luc) contained 5 Gal4 binding sites (Gal4), the TATA-box containing HIV-1 promoter, including NF-kB and Sp1 binding sites driving the expression of the luciferase reporter gene. **(B)** HEK-293 cells were co-transfected with the reporter G5HIV1-Luc together with expression plasmids encoding G4DBD fusions of DYRK1A wild type (G4DBD-DYRK1A), a kinase-death version of DYRK1A (G4DBD-DYRK1A^{KR}), CDK9 wild type (G4DBD-CDK9), or kinase death CDK9 (G4DBD-CDK9^{dn}) as indicated. A plasmid expressing unfused G4DBD was used to measure the basal activity of the reporter. Transfection also included the Renilla expression plasmid, pCMV-RNL. Luciferase activity was measured in triplicate plates and values were corrected for transfection efficiency as measured by Renilla activity. The graph represent transcriptional activity as fold change compared to the control G4-DBD value, set as 1 (average ± s.d.). Data corresponds to a representative experiment of 3 performed. **(C)** Soluble extracts from HEK-293 expressing the indicated G4DBD fusion plasmids were separated by SDS-PAGE and analyzed Western blot using a G4-DBD specific antibody. The position of marker proteins (in kDa) is indicated.

To distinguish between the two possibilities, we tested a reporter with only a minimal promoter region containing the TATA box without any other binding site for accessory transcriptional factors (Fig. R13A). As shown in Fig. R13B, DYRK1A was also able to induce transcription, suggesting that additional transcriptional factors are not needed for proper DYRK1A transcriptional activation. Notably, a clear reduction in the DYRK1A-dependent activation was observed upon deletion of the TATA-box (Fig. R13B); in contrast, this reduction was not apparent when the activation domain of the herpes virus VP16 protein, the

prototypical acidic trans-activator *(Lin et al., 1991)* was assessed (Fig. R13B). The data indicates therefore that the transcriptional activity of DYRK1A does not depend on specific sequences on the promoter. Moreover, the results indicate that DYRK1A-dependent transcriptional activation is more dependent upon the TATA-box than that of VP16. The activity of the fusion Gal4-VP16 as a trans-activator on a TATA-less promoter has been linked to its ability to recruit TFIIB and/or Mediator *(Hall and Struhl, 2002)*, therefore the results might suggest that DYRK1A should target a step that occurs at or downstream of TBP recruitment.

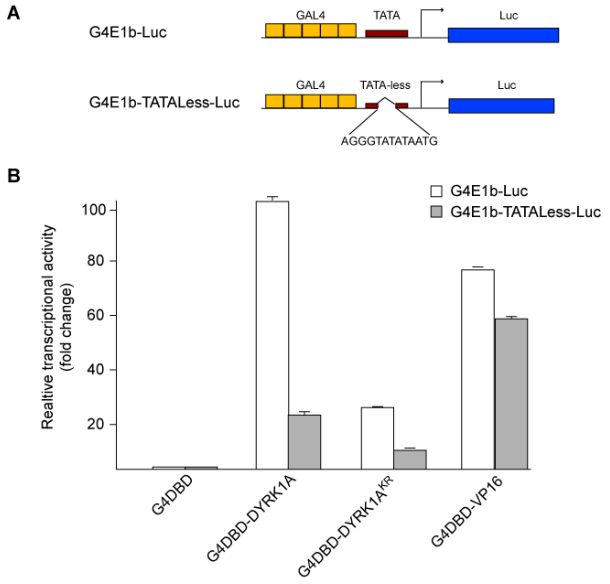


Figure R13: DYRK1A transcriptional activity depends on the presence of a TATA-box in the promoter. (A) Schematic representation of the target plasmids used. The target plasmid G4E1b-Luc contains 5 Gal4 binding sites upstream the adenovirus E1b minimal promoter; in the G4E1b-TATAless-Luc the sequence corresponding to the TATA-box, as indicated, was deleted. **(B)** HEK-293 cells were co-transfected either with G4E1b-Luc or G4E1b-TATAless-Luc reporters together with the expression plasmids encoding G4DBD fusions to DYRK1A wild type (G4DBD-DYRK1A), to kinase-death DYRK1A (G4DBD-DYRK1A^{KR}) or to the C-terminal 87 amino acids of herpes virus VP16 (G4DBD-VP16) as indicated. A plasmid expressing unfused G4DBD (G4DBD) was used as negative control. A Renilla expressing plasmid, pCMV-RNL, was included in all transfections. Luciferase activity was measured in triplicate plates and values were corrected for transfection efficiency as measured by Renilla activity. The graph represent transcriptional activity as fold change compared to the control G4-DBD value, set as 1 (average ± s.d.). Data corresponds to a representative experiment of 3 performed.

As mentioned in the Introduction, DYRK family consist of five different members in humans: DYRK1A, DYRK1B, DYRK2, DYRK3 and DYRK4. The family members are classified as Class I or Class II members

based on phylogenetic analysis (Fig. R14A) *(Aranda et al., 2011)*. To check if the transcriptional activity shown by DYRK1A is shared by others members of the DYRK family, one-hybrid experiments using G4-DBD fusion expression plasmids for all DYRK members were performed on the G4E1b-Luc reporter. The fusion proteins were expressed at similar levels (data not shown), but only DYRK1B was able to induce transcription to a similar extent as was observed for DYRK1A (Fig. R14B). Notably, data from our lab indicated that DYRK1A and DYRK1B are the only members of the human DYRK family that are able to bind the CTD of the RNA Pol II (E. Salichs, Doctoral Thesis, 2008), suggesting a functional link between DYRKs interaction with the CTD and the ability to activate transcription.

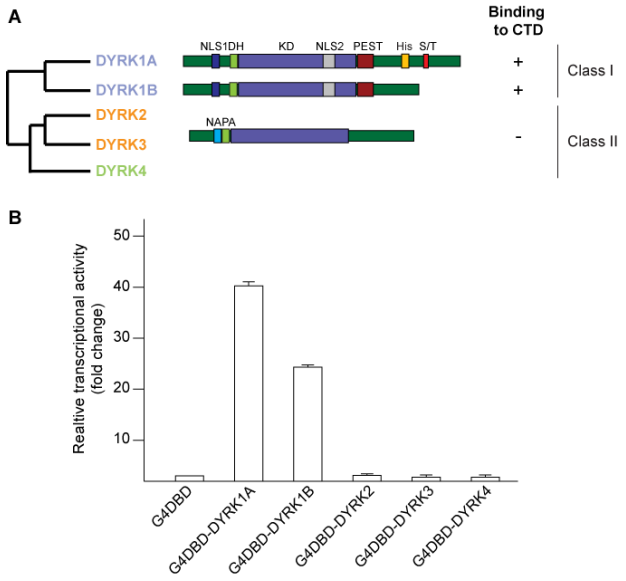


Figure R14: DYRK1A transcriptional activity is conserved in Class I DYRK family members. (A) Schematic representation of DYRK proteins. Left, the simplified tree shows the phylogenetic relationship among human DYRK family members; center, the primary structure of the different members is shown with the main domains as in Fig. I2; right, the ability of the different DYRKs to interact with the RNA Pol II CTD assessed by *in vitro* pull down assays is shown. **(B)** HEK-293 cells were co-transfected with the G4E1b-Luc reporter and pCMV-RNL plasmids together with expression plasmids encoding G4-DBD fusion of wild type versions of DYRK1A, DYRK1B, DYRK2, DYRK3 and DYRK4 as indicated. A plasmid to express unfused G4-DBD was used as negative control. Luciferase activity was measured in triplicate plates and values were corrected for transfection efficiency as measured by Renilla activity. The graph represent transcriptional activity as fold change compared to the control G4-DBD value, set as 1 (average ± s.d.).

Genome-Wide *in vivo* mapping of DYRK1A recruiting sites to chromatin

The proteomic screen allowed us to identify several transcription-related proteins associated to DYRK1A nuclear complexes, including some components of RNA Pol II initiation and elongation complexes. Our results also showed that DYRK1A behaves as a transcriptional activator, in a kinase-dependent manner, when recruited to promoter regions. These evidences led us to hypothesize that DYRK1A could work by sitting on the promoter of specific genes, and thus modulate gene expression by acting on different components of the basal transcription machinery.

Setting up of the workflow

The experimental approach chosen for testing our hypothesis was chromatin immunoprecipitation (ChIP) combined with massive parallel DNA sequencing of the bound DNA fragments (ChIP-Seq). ChIP is a technique for assaying protein–DNA binding *in vivo*, in which antibodies are used to select specific proteins together with the DNA fragments that are bound to these proteins. For that, the DNA-binding protein is cross-linked to the genomic DNA *in vivo* by treating cells with formaldehyde and the chromatin is sheared by sonication into small fragments, which are generally in the 200–600 bp range. An antibody specific to the protein of interest is used to immunoprecipitate the DNA–protein complex. Finally, the crosslinks are reversed and the released DNA is sequenced to determine those sequences bound by the protein.

The value of any ChIP data, including ChIP–Seq data, depends crucially on the quality of the antibody used. Thus, conditions for ChIP with DYRK1A available antibodies were first set up. As a *bona fide* indicator of having immunoprecipitated DYRK1A bound to chromatin, we checked for the presence of histone H3 in the immunoprecipitates. As shown in Fig. R15, DYRK1A was specifically immunoprecipitated from cross-linked HeLa cells by both of the antibodies used. However, histone H3 was only observed when the rabbit antibody was used, very likely indicating different specificities of the two antibodies.

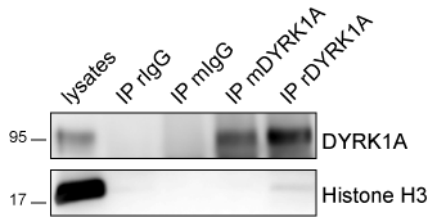


Figure R15 DYRK1A associates with chromatin. Chromatin cross-linking of HeLa cells followed by ChIP assays were performed, as described in Material and Methods, with a mouse antibody (mDYRK1A; Abnova clone 7D10) and a rabbit antibody (rDYRK1A; Abcam ab6981-100) against DYRK1A. Control immunoprecipitations were performed with mouse IgGs (mIgG) and rabbit IgGs (rIgG), Both the lysates (10% of input) and the immunoprecipitates were analyzed by Western blot with an anti-DYRK1A antibody and an anti-histone H3 antibody as indicated. The position of marker proteins (in kDa) is indicated.

ChIP-Seq experiments were performed using HeLa cells and the two DYRK1A antibodies tested. As control, DNA isolated from the same sonicated extract and immunoprecipitated using mouse or rabbit control IgGs was analyzed in parallel. Library construction was carried out at the Genomics Unit of the CRG according to established protocols. Libraries were sequenced to a length of 36 bp on an Illumina GAIIx sequencer. A preliminary analysis of the sequencing data indicated that only the rabbit anti-DYRK1A antibody efficiently and specifically immunoprecipitated DNA fragments when compared to the control IgG ChIP-Seq data (data not shown). Therefore, further analysis was done on the data obtained with the rabbit antibody. Libraries from two biological replicates were made and subjected to deep sequencing. As shown in Fig. R16A, a good correlation was observed for the biological replicates.

All the bioinformatic analysis regarding quality control, read mapping and peak identification were done by our collaborators Daniela Bezdan and Stephan Ossoswki at the CRG. To test the performance of the peak caller ("shore peak" caller), a putative DYRK1A binding site was selected, and the shape of the predicted peak tested by quantitative PCR using primers inside (positives) and outside (negatives) the predicted binding site region. No enrichment was observed in external regions, while a progressive enrichment was achieved when approaching the center of the DYRK1A binding region (Fig. R16B), supporting the accuracy of the peak caller used.

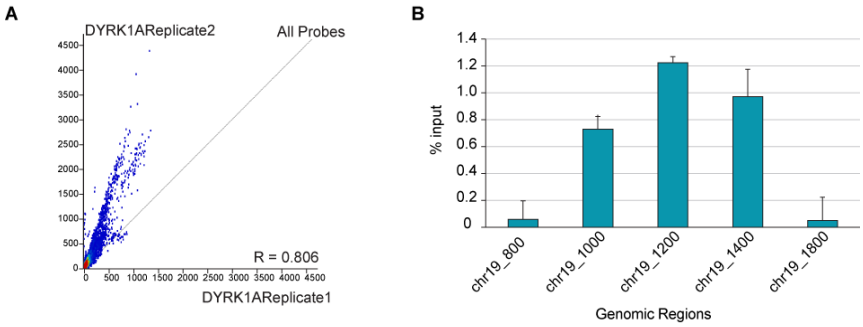


Figure R16: Validation of the ChIP-Seq sequencing procedure. (A) Scatter plot showing Pearson correlation coefficient (PCC) test with reads in defined genomic intervals that were used from two DYRK1A data-sets (DYRK1AReplicate1 and DYRK1AReplicate2) to test the consistency of the replicates. Principal Component Correlation (PCC) showed an R-value of 0.806. **(B)** ChIP-qPCR of a randomly selected DYRK1A associated region on chromosome 19. The results are represented as percentage of input recovery; error bars represent standard deviation from three technical replicates. The ChIP DNA used in the qPCR represents a biological replicate of those samples used in ChIP-Seq.

DYRK1A is recruited to promoter proximal regions

After peak calling, we identified 73 DYRK1A target regions with a high confidence score. No particular recruitment preference to a specific chromosome was observed for the DYRK1A ChIP (Fig. R17A). However, when we analyzed how the DYRK1A-ChIP peaks were distributed across the different genomic regions we found a strong enrichment within promoter regions (80% -57 peaks- versus 2.5% in the whole genome) (Fig. R17B). Examination of the enriched regions confirmed that the DYRK1A-peaks were located mostly within a 1000 bp region upstream of the transcription start site (TSS) of the corresponding genes (Fig. R17C).

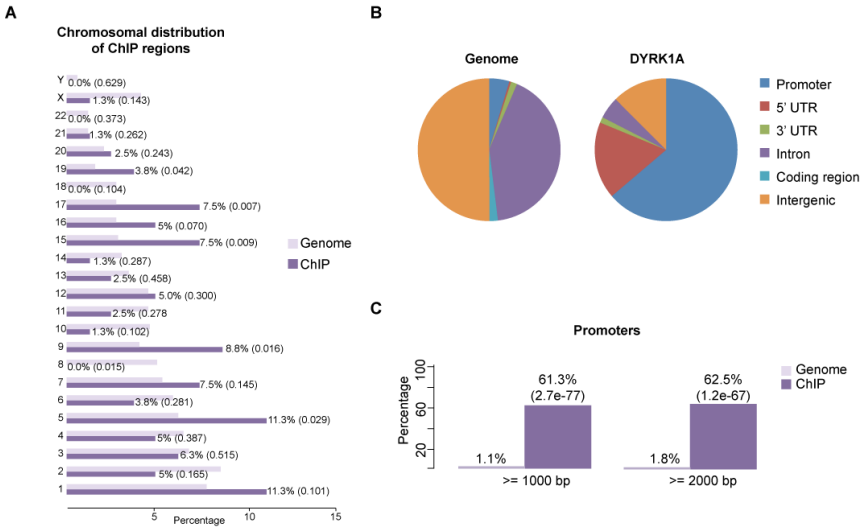


Figure R17: DYRK1A is recruited to promoters. (A) Distribution of ChIP regions over human chromosomes. The light violet bars represent the percentages of the whole tiled or mappable regions in the chromosomes (genome background) and the dark violet bars the percentages of the whole ChIP. These percentages are also marked right next to the bars. P-values for the statistical significance (hypergeometric test) of the relative enrichment of ChIP regions with respect to the genome background are shown in parentheses next to the percentages. **(B)** Graphical results illustrating how DYRK1A ChIP regions are distributed over some genomic features. The pie chart is mutually exclusive (no overlaps); thus, the sum of the percentage values is 100 %. `Intergenic' represents the percentage of ChIP regions that do not belong to any of the other genomic features. **(C)** Bar plot shows the relative enrichments of ChIP regions in promoters with respect to the genome background. The light violet bars represent the percentages of the tiled or mappable regions that are located in this genomic regions (genome background) and the dark bars the percentages of ChIP regions. P-values for the significance (hypergeometric test) of the relative enrichment of ChIP regions with respect to the genome background are shown in parentheses next to the percentages.

Because of the low number of DYRK1A peaks found in the ChIP-seq of HeLa cells, we decided to perform another ChIP-seq analysis on another cell line. We chose the glioblastoma cell line, T98G, in which a role for DYRK1A in quiescence had been previously shown *(Litovchick et al., 2011)*. To understand whether DYRK1A-recruitment to promoters was dependent on the proliferative status of the cells, we decided to perform the ChIP-Seq analysis in two different growth conditions: cells grown on complete media (proliferating cells) and cells growth-arrested by culturing them under serum starved conditions. As shown in Fig. R18A, serum deprivation induced a change in the cell cycle profile with cells mostly found in the G1 phase. No changes in DYRK1A protein levels were observed (Fig. R18B).

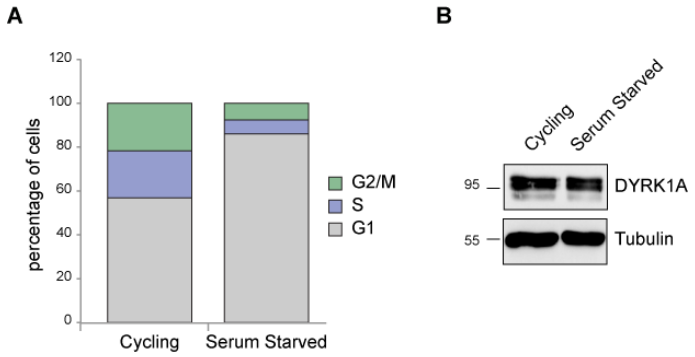


Figure R18: Serum starvation does not alter DYRK1A protein levels (A) Subconfluent T98G cells were maintained for 48 h in DMEM either without FBS (serum starved) or in the presence of 10% FBS (cycling). Cells were then collected and cell cycle profile of the entire populations was thus analyzed by FACs analysis. Quantitative representation of G1, S, and G2/M subpopulations is shown in the graph. **(B)** Cell extracts from the two different growth conditions were analyzed by Western Blot with an anti-DYRK1A antibody. An antibody to Tubulin was used as loading control.

DYRK1A-ChIP experiments were then performed using chromatin from proliferating and serum starved cells. In the case of proliferating T98G, 516 high confidence peaks were identified (Fig. R19A). However, when analyzing the data from the serum starved cells only 305 peaks passed the threshold (Fig. R19A). Of note, 85% of the peaks (263) found in serum starved conditions were also found in proliferating conditions, suggesting that DYRK1A binding is most probably not dependent on the growth conditions of the cell. Moreover, all the peaks previously found in HeLa cells were present within this common subset, indicating that DYRK1A binding to DNA is not cell type specific, at least for the small subset of regions identified in both screens (Fig. R19A). Notably, a strong enrichment within promoter regions was observed for DYRK1A peaks when considering both the subset specific for proliferating conditions (subset a) and the subset common to the two conditions (subset ab) (50% and 85% of total peaks, respectively). However, only 6 out of 62 (9%) of the serum starved specific peaks (subset b) corresponded to promoter regions (Fig. R19B).

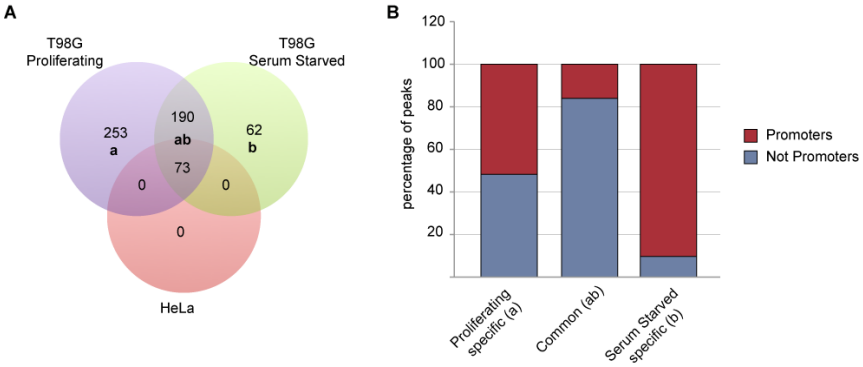


Figure R19: DYRK1A-peak enrichment in the different ChIP-Seq experiments. (A) Definition of peak subsets by intersection of peak sets called for HeLa, T98G proliferating and T98G Serum Starved shown as Venn diagram. Peak counts for each subset are shown. **(B)** Distribution of DYRK1A ChIP-Seq target regions in the different subsets.

The loss of DYRK1A recruitment to a specific subset of target promoters in serum starved conditions could have a biological meaning, that is, it would depend on a specific upstream signal that regulates DYRK1A recruitment. Alternatively, the observed results could be due to technical problems deriving from the different chromatin preparations. Further experiments are needed to distinguish between these two possibilities. The subsequent analysis in this Thesis work has been focused on the DYRK1A-peak subsets in which a promoter enrichment was observed (subsets a and ab in Fig. R19A).

As mentioned above, a strong enrichment within promoter regions was observed for the DYRK1A peaks, since they comprise more than 60% of all DYRK1A binding sites (Fig. R20A). Furthermore, 7% of DYRK1A associated regions overlapped with CpG islands, which although not associated with known promoters in the different databanks used in the annotation process, were mostly found in the proximity of an annotated TSS. Among the remaining sites, less than 10% corresponded to gene bodies, mostly within introns (7%), and the rest, 22% of the DYRK1A regions, were located in regions that do not belong to any of other genomic regions considered (named as "intergenic regions" in this work) (Fig. R20A). Indeed, when the relative position of the DYRK1A-enriched regions were plotted with respect to their closest TSSs, the majority of the DYRK1A peaks covered a narrow region 5' of the TSS (Fig. R20B), indicating that DYRK1A is associated to proximal promoter regions.

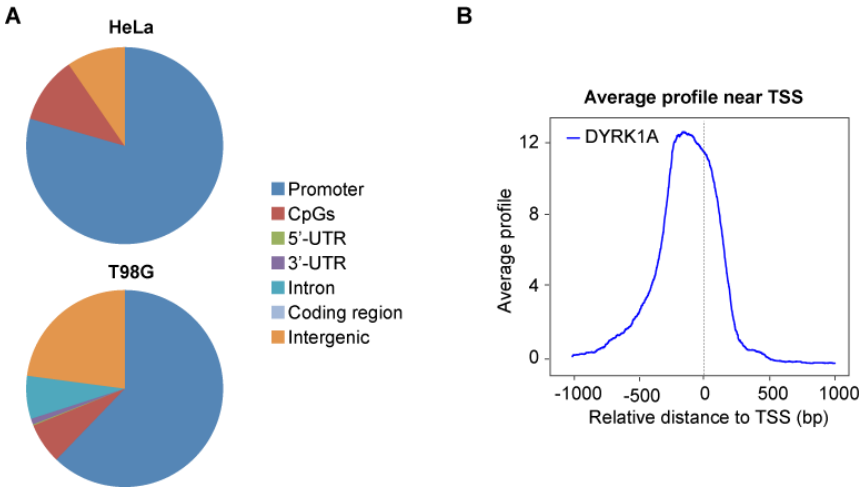


Figure R20: DYRK1A is recruited to proximal promoter regions. (A) Distribution of DYRK1A ChIP-Seq target regions in HeLa and T98G cells, accordingly to the annotation pipeline described in Fig. R17. **(B)** Distribution of DYRK1A enriched peaks near the TSS of all Ensembl genes. The density plot shows the distribution of the "peak center to SwitchGear-TSS" distance. Average peak coverage is shown in a bin size of 50 bp for a window of 1 kb upstream and 1 kb downstream from TSS.

For further validation of the interaction of DYRK1A with chromatin, we selected several DYRK1A target regions with different levels of enrichment and performed independent ChIP analysis coupled with quantitative PCR. DYRK1A binding was found to be markedly enriched at all the genomic regions tested, whereas no enrichment was observed on random genomic regions used as negative control (Fig. R21A). To assure the specificity of the DYRK1A antibody used, we performed experiments in cells in which the levels of DYRK1A were reduced by lentiviral transduction of a specific shRNA against DYRK1A. ChIP-qPCR analysis was then carried out in parallel with cells expressing a control shRNA and cells knocked down for DYRK1A. Knockdown of DYRK1A led to a marked reduction in the occupancy of target genes (Fig. R21B), confirming thespecificity of the ChIP data.

All these data indicate that DYRK1A is recruited to the promoter regions of a subset of genes, and further suggest DYRK1A involvement in their transcriptional regulation.

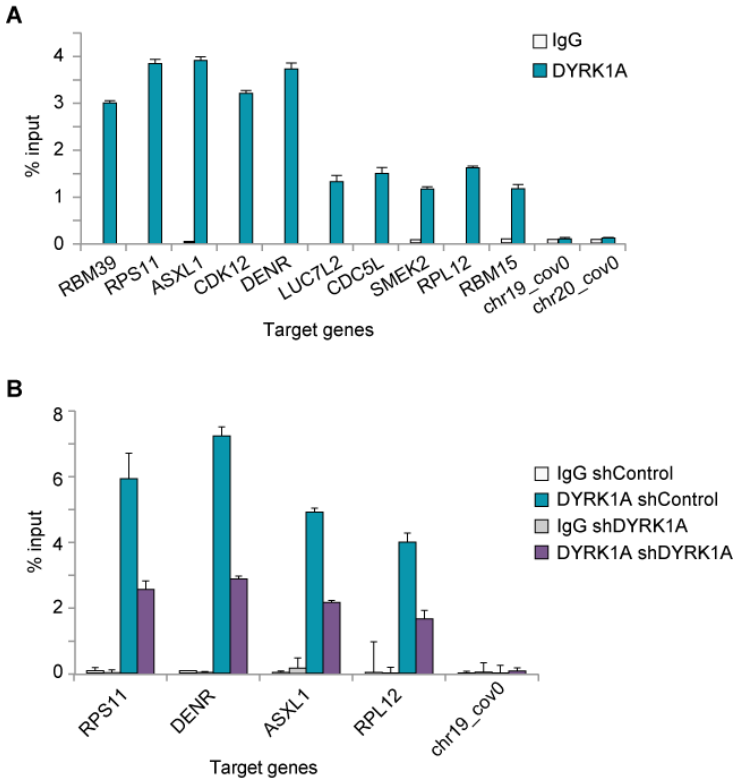


Figure R21: DYRK1A is bound to genomic regions *in vivo*. (A) ChIP-qPCR assays using anti-DYRK1A antibody and normal IgGs as negative control. Two random genomic regions (chr19_cov0 and chr20_cov0) were selected as additional negative controls. DYRK1A ChIP samples derived from biological replicates of those samples used in ChIP-Seq. The data is represented as percentage of input recovery; error bars represent standard deviations from three technical replicates. **(B)** Validation of DYRK1A target regions. ChIP-qPCR assays on samples from HeLa cells infected with a control lentivirus (shControl) or a lentivirus expressing a shRNA against DYRK1A (shDYRK1A) were performed as described in **A**. DYRK1A knocked down cells showed reduced binding.

DYRK1A is located in accessible and active chromatin regions

To investigate the relationship between promoter occupancy by DYRK1A and the transcriptional activity of its putative target genes, we first looked for overlap between DYRK1A peaks and the presence of several features that define gene activity such as RNA Pol II recruitment and open chromatin as indicated, for instance, by the presence of histone H3 lysine 27 acetylation (H3K27Ac) and H3 lysine 9 acetylation (H3K9Ac) *(Kim et al., 2005; Roh et al., 2005)*. Other chromatin marks generally used to profile gene expression and associated to histone

lysine methylation were also included *(Barski et al., 2007; Kim et al., 2005)*. To this aim, publicly available ENCODE ChIP-Seq data sets were used (this bioinformatic analysis was done by Khademul Islam at the Nuria López-Bigas' group, UPF, Barcelona). As shown in Fig. R22A, DYRK1A islands are free of nucleosomes as seen by the deep cleft in acetylated H3 variants and for the marked depletion in all histone marks at the center of the ChIP regions. These findings are an indication of nucleosome exclusion, which would facilitate the recruitment of proteins to DNA. Given the proximity of the DYRK1A-islands to the TSS, it is predictable that the open chromatin state will allow binding of the transcription machinery and of other factors required for the proper initiation of transcription.

Moreover, DYRK1A-bound promoters show a higher occurrence of activating chromatin modifications, when compared with genes negative for DYRK1A binding, such as i) high levels of histone H3 lysine 4 tri-methylation (H3K4me3) and histone H3 lysine 4 dimethylation (H3K4me2) surrounding TSS; ii) elevated histone H3 lysine 36 trimethylation (H3K36me3) downstream the TSS and along the gene body; and iii) high RNA Pol II occupancy (Fig. R22B). All these chromatin modifications are marks of actively transcribed genes *(Barski et al., 2007)*. Therefore, we conclude that genes bound by DYRK1A are maintained in an open chromatin state and are actively transcribed in basal conditions.

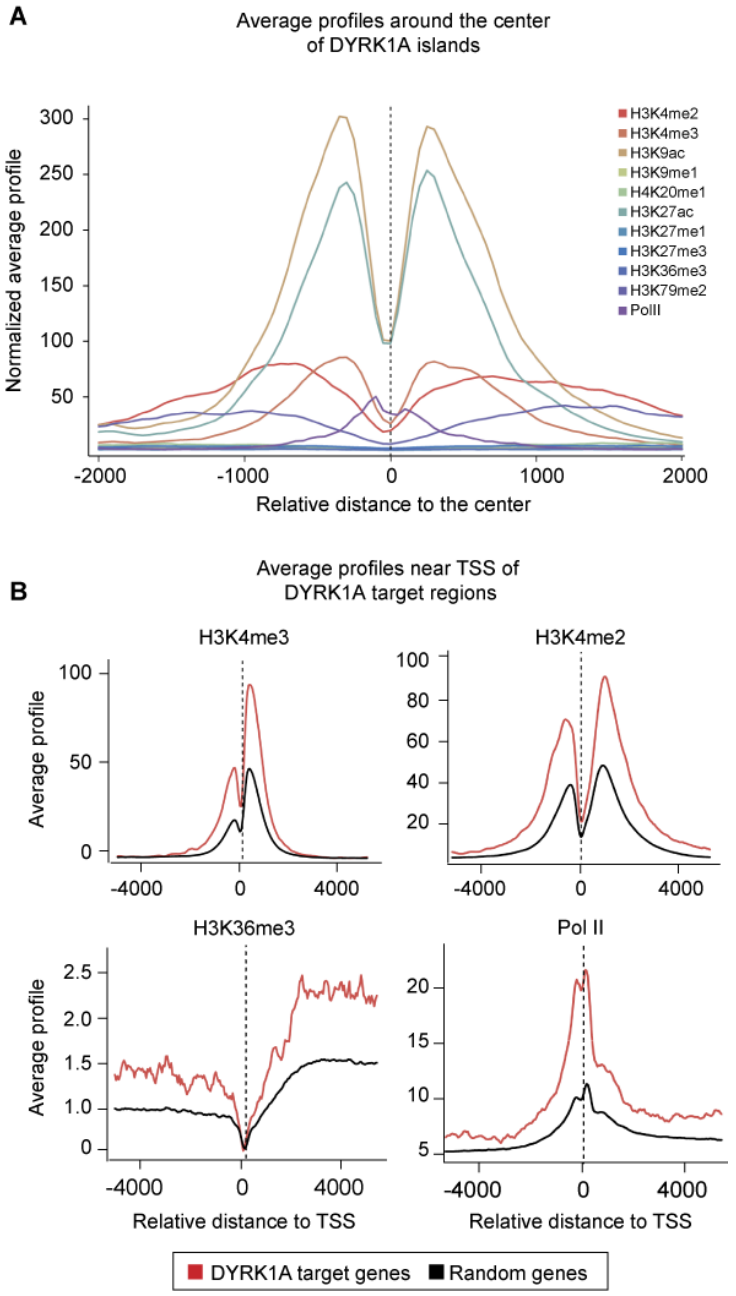


Figure R22: DYRK1A associated genes are actively transcribed. (A) Profiles of the histone modifications across DYRK1A islands as indicated on the right legend. **(B)** Profiles of the histone modifications indicated above each panel and of RNA Pol II (Pol II) across the TSS are shown for DYRK1A associated genes (in red) and for random Ensemble genes (in black).

DYRK1A is an activator of transcription in vivo

The analysis of the chromatin state of the DYRK1A-associated genes revealed a common pattern of activating chromatin marks supporting the hypothesis that DYRK1A recruitment to promoters could have a functional role in regulating transcription *in vivo*. To experimentally validate this hypothesis, the expression levels of DYRK1A-target genes were analyzed comparing wild type cells and cells in which DYRK1A was depleted through lentiviral transduction of specific shRNAs against DYRK1A (Fig. 23A). To control for off-target effects, two distinct shRNAs were used. Upon DYRK1A knockdown, the expression levels of its target genes were reduced, pointing thus to a positive role for DYRK1A in regulating gene expression (Fig. R23B).

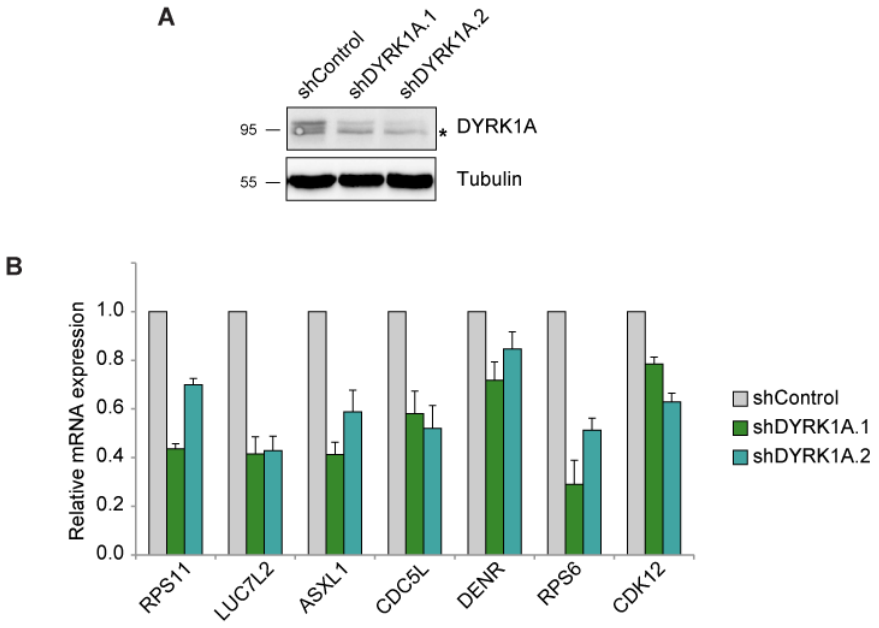


Figure R23: DYRK1A regulates expression of its target genes *in vivo*. T98G cells were transduced either with a control lentivirus (shControl) or with lentiviruses expressing two different shRNAs against DYRK1A (shDYRK1A.1 and shDYRK1A.2). (*) indicates an aspecific band recognized by the antibody.(**A)** Reduction in DYRK1A levels as shown by Western blot analysis. The asterisk points to a cross-reacting band. **(B)** Five days post infection RNA was prepared and the expression levels of randomly selected DYRK1A target genes were determined by RT-qPCR using specific primers. Data represent the average and standard deviation from technical triplicates.

It is widely accepted that genes controlled by a common regulator often show similar expression in specific sets of tissues *(Allocco et al., 2004; Yu et al., 2003)*, an assumption that is the basis of all approaches that

use mRNA expression data from microarray experiments to identify gene regulatory networks. Therefore, we reasoned that *DYRK1A* could share tissue expression specificity with its putative target genes. To test this hypothesis, we used expression data from the compendium of human tissue samples in GeneAtlas *(Su et al., 2004)* to investigate the co-expression profiles of DYRK1A with its putative target genes, and limited the subset of genes to those common to the HeLa and T98G ChIP experiments. *DYRK1A* expression levels are quite different, when the tissue samples and cell lines included in the Atlas are compared, with the highest level of expression shown by cell types of the hematopoietic lineage, including B cells, CD4+ and CD8+ T lymphocytes or monocytes, endothelial cells and fetal brain (Fig. R24A). Noteworthy, the correlation analysis indicated that DYRK1A target genes are also preferentially expressed in these cell types (*Z*-score >1.96) (Fig. R24A). Likewise, *DYRK1A* is expressed at low levels in heart or skeletal muscle, where expression of its target genes is also reduced.

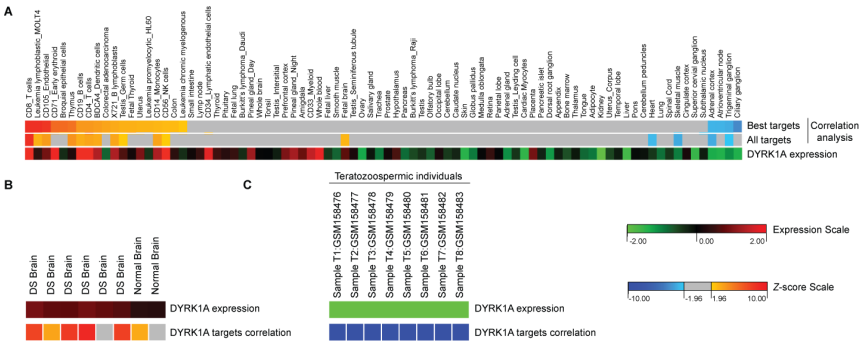


Figure R24: Expression and correlation of DYRK1A and its targets in human tissues. The graphs show the absolute (log_2) expression values for DYRK1A in different tissues and cell lines plotted as a color-coded heatmap, where red indicates higher expression and green indicates lower expression (see "Expression scale"). The correlation analysis is represented as *Z*-score values for preferential expression and underexpression of DYRK1A target genes delineated in a different colored heatmap, where red means overexpression of targets and blues means underexpression of targets; grey indicates no significant difference from expected value (see "*Z*-score scale"). This analysis was performed by K. Islam at the López-Bigas' group. **(A)** DYRK1A expression and correlation analysis in different tissues and cell lines for the GeneAtlas dataset (GEO GSE1133 dataset). The subset of genes common to the HeLa and T98G ChIP experiments were used for the correlation analysis ("All targets"); a reduced subset corresponding to those genes having DYRK1A peaks on the -300 to +1 region ("Best targets") was also used. **(B)** DYRK1A expression and correlation analysis in brain samples from six Down syndrome (DS) individuals and two controls (GEO dataset GSE5390). **(C)** Expression data is represented as fold change (log2) in each of the teratozoospermic samples respect to the average level of expression of DYRK1A in samples from 14 normospermic individual (GEO dataset GSE6872).

To further support these findings, we decided to analyze the pattern of expression of putative DYRK1A targets in conditions in which DYRK1A is either overexpressed or downregulated, using expression data from publicly available microarray experiments. For the first condition, we used data from Down syndrome brains where DYRK1A is known to be overexpressed around 1.5-fold *(Guimera et al., 1999)*. The data correspond to the GEO Dataset GSE5390 that includes postmortem brain tissue from Down syndrome individuals and non-trisomic controls *(Lockstone et al., 2007)*. For the second situation, we search the GEO Datasets for experiments showing a clear reduction in *DYRK1A* expression and found that the expression of the gene is highly reduced (log2 around -5) in sperm cells from males with consistent and severe teratozoospermia *(Platts et al., 2007)* (GSE6872). Interestingly, DYRK1A target genes positively correlate with DYRK1A expression: they are found to be upregulated in the condition in which DYRK1A is overexpressed (Fig. R24B), while they are strongly downregulated when DYRK1A is not expressed (Fig. R24C).

Based on all the data presented in this section, we propose a new role for DYRK1A as an activator of transcription *in vivo*.

DYRK1A is recruited to target promoters via a conserved DNA motif

De novo motif analysis reveals a putative DYRK1A binding sequence within target promoters

Although DYRK1A-associated promoters are enriched in activating chromatin marks, they are only a small subset if compared with all actively transcribed promoters in the whole genome. This suggests that chromatin structure alone cannot explain the specific chromatin localization of DYRK1A. To determine whether the genomic targeting of DYRK1A occurs by a mechanism involving transcription factors, we asked whether DYRK1A occupied sites were enriched for certain DNA motifs. MEME analysis on the 516 DYRK1A positive hits revealed one motif that was strikingly significantly enriched within DYRK1A associated regions (p-value = e-431), and that corresponded to the

palindromic sequence -TCTCGCGAGA- (Fig. R25A). The motif presents a high conservation score within placental mammals (Fig. R25B). Furthermore, the motif distribution was not homogenous within the genomic categories, since 288 out of 347 (83%) of the DYRK1A-associated promoters contained the consensus sequence, while only 6% of the intergenic regions did (Fig. R25C). All together, the results point to a potential role for the newly identified palindrome as a regulatory element.

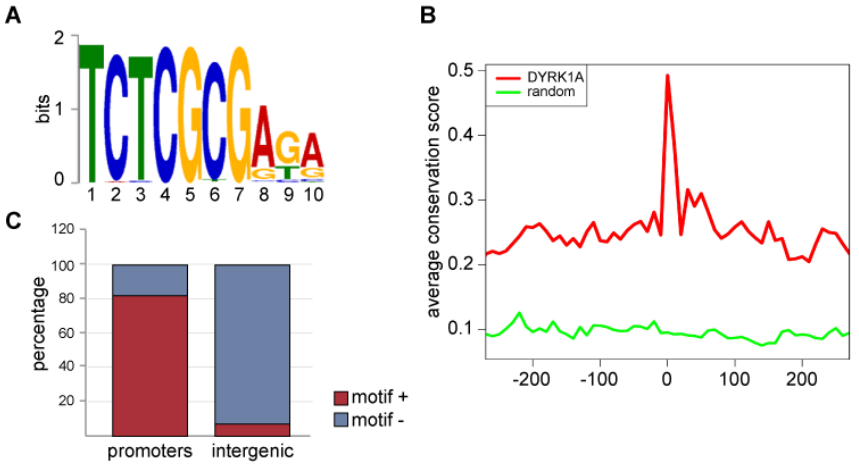


Figure R25: De novo motifs analysis identify a putative DYRK1A-recruiting DNA sequence. (A) Web logo of the significantly enriched motif identified in DYRK1A targets using MEME *(Bailey et al., 2009)*. **(B)** Average conservation of DYRK1A binding sites (red line) compared to random sequences of equal length (green). In X axis, 0 indicates the mid point of the DYRK1A peaks. Minus indicates upstream bases and plus value indicates down-stream bases of DYRK1A binding site sequences from center. **(C)** Histogram depicting the percentage of DYRK1A targets that contains -TCTCGCGAGA- motif, classified by their genomic annotation.

The same MEME search was made separately in each of the DYRK1A-peak subsets (a = proliferating specific; ab = common; b = serum starved specific). Identical parameter settings were applied to test for differences in motif composition, however no significant distances could be detected between the predicted position weight matrixes identified in the promoter enriched subsets (a and ab) (Fig. R26A and B). However, no significant motif enrichments were found instead in the 62 peaks of the serum starved-specific subset, except for two simple sequence repeat-like motifs (Fig. R26C) seem to be very common among the peaks. These results further suggest a possible role for this consensus

sequence as a promoter specific-recruiting platform for DYRK1A-containing complexes.

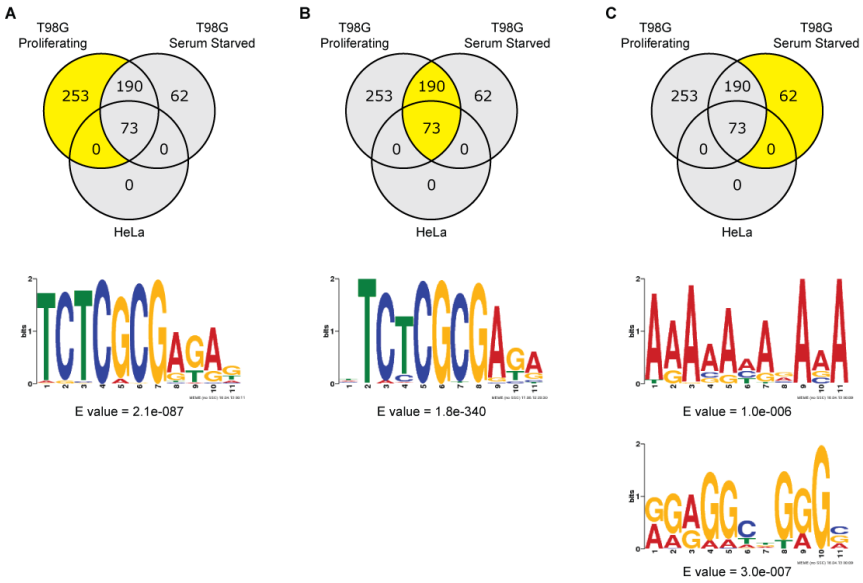


Fig. R26: TCTCGCGAGA motif is promoter specific. MEME motif search in the "a" **(A)**, "ab" **(B)** and "b" **(C)** subsets, respectively, as shown in yellow in the Venn diagrams (see Fig. R19 for details of subsets). Best motifs found and respective E values are shown below each subset motif.

To assess the functional role of the putative DYRK1A binding site, we first tested the ability of the DNA motif to interact with proteins in cell extracts. For that, electrophoretic mobility shift assays (EMSA) were performed using nuclear extracts prepared from HNE and a double strand oligonucleotide containing the palindromic consensus sequence identified by MEME. Two different band shifts were observed with increasing amounts of HNE, indicating the existence of two different protein complexes bound to the palindrome (Complex I and Complex II, respectively) (Fig. R27A). Competition assays with increasing amounts of the wild type unlabeled oligonucleotide blocked the formation of both complexes, while preincubation with increasing amounts of a mutant oligonucleotide failed to do it (Fig. R27B), supporting the specificity of the protein complexes bound to the DNA-containing motif.

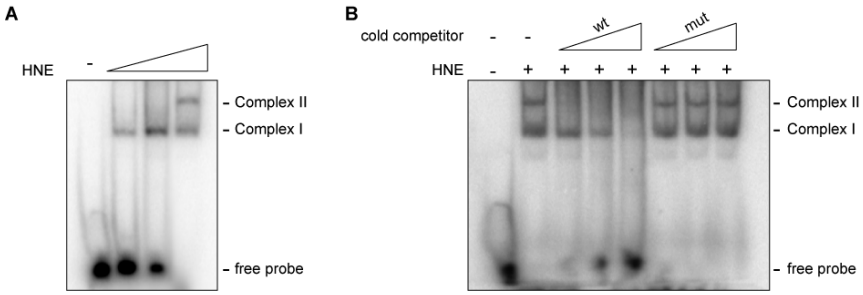


Figure R27: The palindromic sequence TCTCGCGAGA has protein binding activity. (A) EMSA was performed using increasing amounts of nuclear extracts from HeLa cells and a ^{32}P-labeled oligonucleotide probe containing the authentic sequence of DYRK1A sites from the RPS11 promoter region. DNA-protein complexes were resolved by electrophoresis on non-denaturing polyacrylamide gels. **(B)** A competition assay was performed by using increasing amounts (20-fold, 50-fold and 100-fold) of a wild type cold probe (wt) and increasing amounts of a probe in which the consensus sequence was mutated (mut). The position of the shifted band corresponding to Complex I and II is indicated as well as that of the free probe.

The positional distribution of the consensus TCTCGCGAGA-motif within the DYRK1A peaks was analyzed, as this could provide further detail as to whether DYRK1A may be directly recruited to the DNA or rather via more complex, indirect interactions. We observed an exact motif positioning around the peak center (Fig. R28A), pointing to a direct DYRK1A recruitment to the motif. To validate this hypothesis, we performed EMSA assays with a DYRK1A protein expressed and purified form bacteria as a GST-fusion protein. However, we could not detect any specific binding of the purified protein under the conditions used (data not shown).

To provide direct evidence that Complex I and/or II contained DYRK1A, an antibody-mediated supershift analysis was performed using two different antibodies that recognize DYRK1A in immunoprecipitation assays and therefore, in native conditions. As shown in Fig. R28B, pre-incubation with one of the antibodies completely prevented the formation of Complex I, while pre-incubation with the other resulted in an increased intensity of the band corresponding to Complex I and in the appearance of a higher molecular weight band likely corresponding to a supershift. The differences in behavior of the two antibodies might be due to their recognition of different epitopes within the protein. A control antibody failed to generate any change in mobility (Fig. R28B). These data suggest several things: first, DYRK1A is necessary for the formation of the DNA-protein complex, because when sequestered by

an antibody that does not recognize the kinase bound to DNA (as mentioned previously, the mDYRK1A antibody did not produce any ChIP-Seq data, but it efficiently binds to DYRK1A in immunoprecipitation assays using HNE) no band shift is observed; second, we can safely assume that DYRK1A is part of the protein complex bound to the consensus motif, although our data do not allow us to conclude whether its binding is direct or mediated by other proteins.

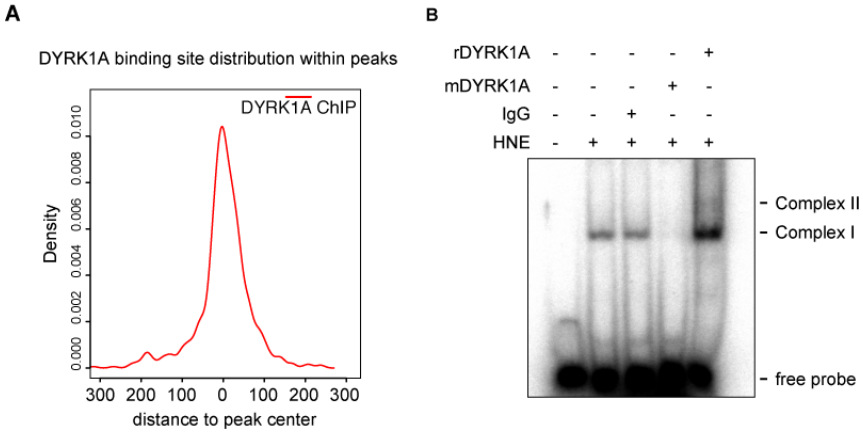


Figure R28: The palindromic sequence TCTCGCGAGA has protein binding activity. **(A)** Distribution of the TCTCGCGAGA motif in the DYRK1A peak regions**.** Peaks have been centered at x-axis = 0. Motif density has been plotted using Rgraphics combo density plot. **(B)** EMSA was performed by using HNE incubated with wild type probe in the absence (-) or presence (+) of the antibodies to DYRK1A indicated (mDYRK1A: mouse monoclonal; rDYRK1A: rabbit polyclonal; IgG: normal mouse IgGs).

At this point of my Thesis experimental work, a manuscript from Deplancke's group was published describing that the transcription factor KAISO/ZBTB33 binds directly to the TCTCGCGAGA-motif, recruiting the co-repressor SMRT (silencing mediator of retinoic acid and thyroid hormone receptor) to pre-adypocite specific proximal-promoter sites *(Raghav et al., 2012)*. In addition, KAISO was one of the 119 human transcription factors analyzed for chromatin distribution within the ENCODE Consortium project and whose results were made available very recently *(Wang et al., 2012)*. The ENCODE data revealed that the TCTCGCAGA-motif was enriched in three datasets, corresponding to that of KAISO, of the multitasking tumor suppressor BRCA1 and of the chromatin modifier CHD2 (chromodomain helicase DNA binding protein 2) *(Wang et al., 2012)*. The possibility was then considered that any of these proteins could be the scaffold bridging DYRK1A to DNA. Given

that BRCA1 and CHD2, unlike KAISO, lack any DNA-binding domain, they were not taken into account as putative DYRK1A scaffolds.

To explore whether KAISO is a potential mediator of DYRK1A recruitment to chromatin, publicly available ENCODE transcription factor ChIP-Seq data sets were used to check whether the subset of chromatin regions bound by KAISO corresponded indeed to DYRK1A-associated regions. The analysis revealed a marked co-occurrance down to the level of peak maxima of DYRK1A and KAISO-binding events (Fig. R29A). We found that 289 (54%) out of 528 sites primed by DYRK1A are also positive for KAISO binding (Fig. R29B, "total" column). Indeed, if only the subset of promoter regions was considered, the overlap goes up to 70% (Fig. R29B, "promoter" column), pointing to KAISO as a good candidate to be the recruiter of DYRK1A to its target promoters.

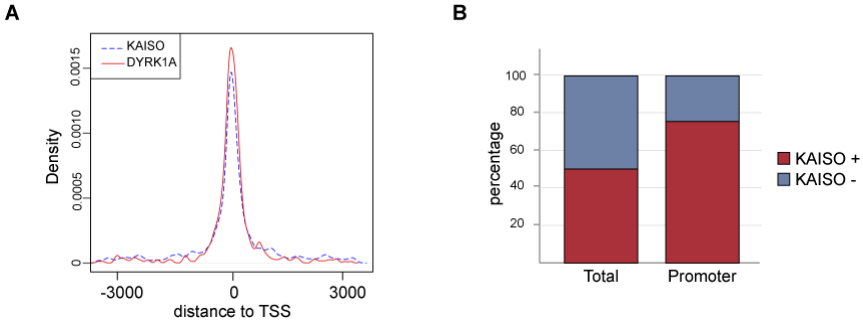


Figure R29: KAISO binding profile within DYRK1A associated peaks. (A) Density plots were generated to show the distribution of the peak center of the KAISO and DYRK1A peaks according to the TSS. Positions of TSS were taken from the UCSC genome browser (SwitchGear TSS Track). The offset have been set to +/- 3,500 bp. **(B)** Histogram depicting the percentage of DYRK1A targets (all targets and only promoters, first and second column respectively) which is found to be bound by KAISO as well.

We thus tested whether the two proteins interact by performing co-immunoprecipation experiments using HNE with antibodies against DYRK1A and KAISO. We have not been able to detect the presence of any of the proteins in the immunoprecipitates of the other one (Fig. R30A), and therefore, we conclude that KAISO does not interact with DYRK1A in soluble extracts. However, we cannot exclude that the interaction only happens in the specific context in which both proteins meet at DNA. In this situation and accordingly to the hypothesis of KAISO being the recruiter of DYRK1A to its target regions,

downregulation of KAISO would abrogate the binding of DYRK1A to its targets. To test this possibility, we performed DYRK1A ChIP of several common DYRK1A/KAISO targets in cells in which KAISO was depleted by lentiviral transduction of a shRNA, and found that DYRK1A binding to its targets was reduced (Fig. R30B), although not completely abrogated.

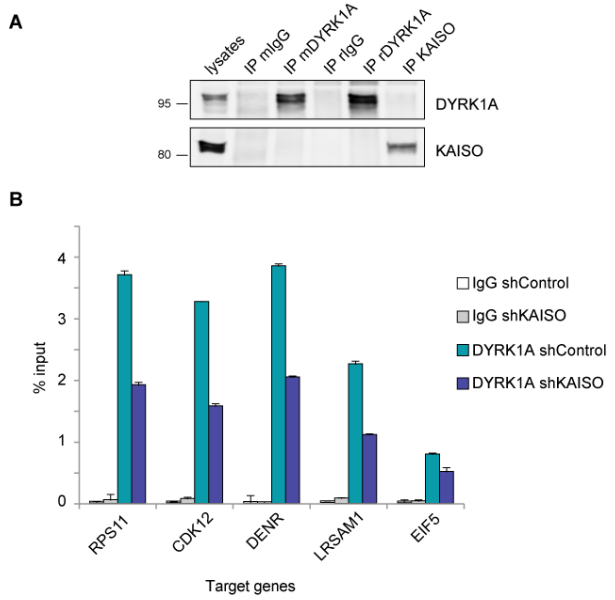


Figure R30: KAISO could act the DYRK1A recruiter to target regions. (A) HNE were immunoprecipitated with two different antibodies against DYRK1A (mouse monoclonal and rabbit polyclonal) DYRK1A or a rabbit polyclonal anti-KAISO and normal mouse IgGs (mIgG) or normal rabbit IgGs (rIgG) as controls and both the lysates (10%) and the immunocomplexes were analyzed by immunoblotting with antibodies against DYRK1A and KAISO. **(B)** ChIP-qPCR assays on samples from HeLa cells infected with a control lentivirus (shControl) or a combination of two lentiviruses against KAISO (shKAISO) were performed as described in Fig. 21A.

This data suggest that KAISO could be the responsible for the recruitment of DYRK1A to its target promoters.

DYRK1A genomic targets link DYRK1A to regulation of cell growth

Analysis of DYRK1A target genomic regions reveals a overrepresentation of genes related with cell growth

To gain insight into the biological role of DYRK1A recruitment to chromatin, we asked whether specific gene functions were overrepresented among its target genes. IPA of DYRK1A-promoter bound genes revealed a prominent enrichment for genes connected with protein synthesis and translation (Table R4, top). Indeed, eIF2 and p70S6K signaling were found significantly enriched among the signaling pathways (Table R4, bottom), pointing to a possible involvement of DYRK1A in cell growth.

Table R4: IPA analysis of DYRK1A-promoter bound genes

Molecular and Cellular Functions	Counts[1]	p-Value
RNA processing	40	1.80E-09
Translation	25	7.00E-06
mRNA processing	24	1.20E-06
Ribonucleoprotein complex biogenesis	20	1.80E-08
Ribosome biogenesis	10	3.10E-03
rRNA processing	20	2.40E-12
Canonical Pathways		
EIF2 signaling	15	5.19E-09
Regulation of eIF4 and p70S6K signaling	7	1.74E-03

[1]: number of proteins in the category

We also analyzed more in detail the type of transcripts associated to the DYRK1A-positive chromatin regions and found that the category corresponding to tRNAs was clearly enriched in the DYRK1A target regions (Fig. R31A), and notably a high statistical significant overlap was found between DYRK1A peaks and tRNAs (p-Value < 1.0E-6). This overlap was particularly apparent in the subset of peaks classified as “intergenic” in the initial analysis, where around 80% of these loci map to tRNAs and other RNA Pol III genes, including 7SK RNA, U6atac RNA, one of the four Y RNAs, 2 of the 9 U6 RNAs and the Vault RNAs. No significant preference for one specific tRNA was found, except maybe for the no representation of tRNAs for STOP codons and a lower percentage than expected of tRNAs for Asn (Fig. R31B). By looking to

the chromatin features at the UCSC browser of these DYRK1A-bound regions, we noticed that almost all of them were occupied by both RNA Pol III and RNA Pol II (data not shown).

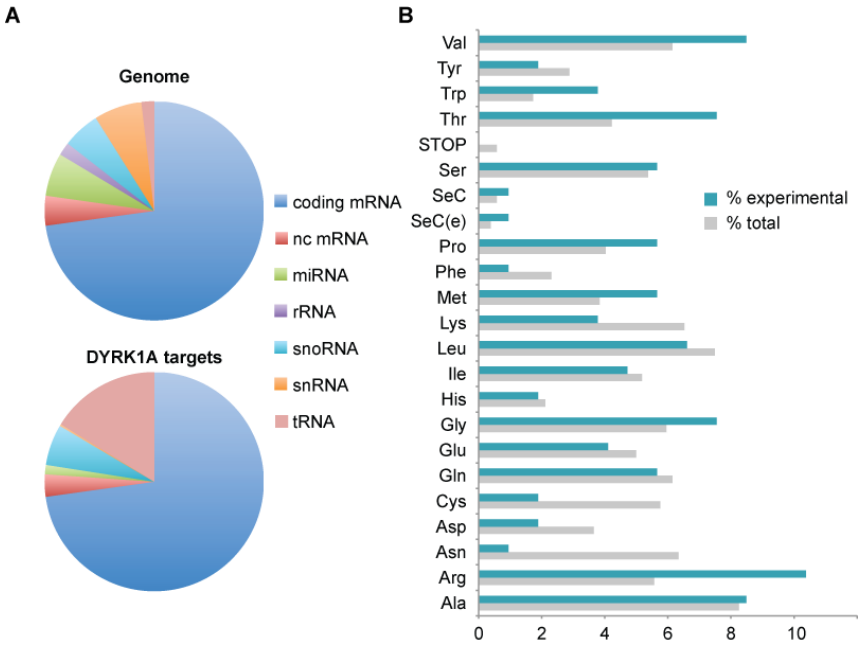


Figure R31: DYRK1A is associated to RNA Pol III genes. (A) Distribution of DYRK1A associated regions, accordingly to the type of transcripts encoded. **(B)** Distribution of ChIP regions over human tRNAs. The grey bars represent the percentage each type of tRNA in the human genome, and the light blue bars the percentage of DYRK1A-ChIP regions that correspond to each type of tRNAs. The complete list of human tRNAs was downloaded from the Genomic tRNA Database (gtrnadb.ucsc.edu/Hsapi).

We therefore wondered whether binding of DYRK1A to the RNA Pol III genes was somehow regulating their transcription and measured levels of some of the target tRNAs and one of the U6RNAs in conditions of reduced expression of DYRK1A. A decrease in the expression of chr8.tRNA10.MetCAT, chr6.tRNA2.MetCAT and chr16-RNAU6 genes was observed upon knockdown of DYRK1A to a similar extent as the decrease observed for DYRK1A-bound RNA Pol II dependent genes (Fig. R32), suggesting once more the ability of DYRK1A to act as an activator of transcription when bound to target genomic regions.

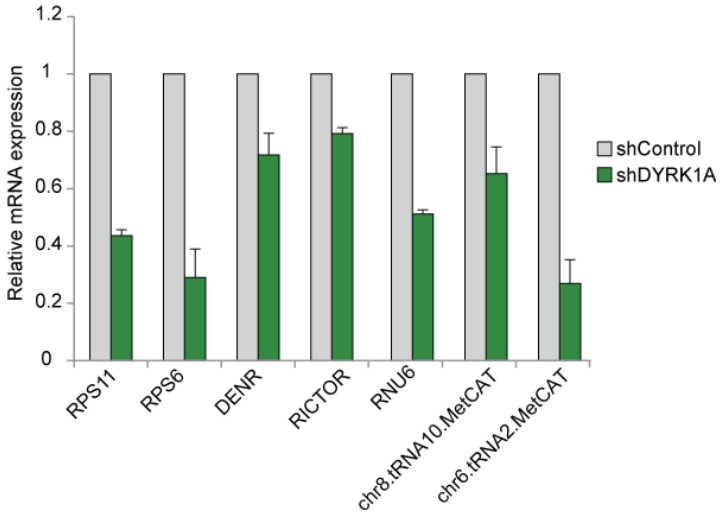


Figure R32: DYRK1A regulates expression RNA Pol III genes. T98G cells were transduced either with a control lentivirus (shControl) or a lentivirus expressing shRNAs against DYRK1A (shDYRK1A). Five days post infection RNA was prepared and the expression levels of DYRK1A target genes (indicated in the x axis) were determined by RT-qPCR using specific primers. Data represent the average and standard deviation from technical triplicates.

Loss of DYRK1A impairs cell growth responses

Cellular processes underlying growth lead to an increase in cell size (for a recent review, see *Tumaneng et al., 2012)*. Additionally, cell growth is linked to cell cycle progression because cells need to increase their size before mitosis to assure that the two daughter cells maintain their size (for a review, see *Fingar and Blenis, 2004)*. Because of the enrichment of cell growth related genes (tRNAs, snoRNAs, genes encoding ribosomal proteins, the release factor ETF1 or the putative translation initiator factor DENR) within DYRK1A targets, and because DYRK1A activity appeared to modulate their expression, we wondered whether DYRK1A could be able to affect these cellular processes. We therefore knockdown DYRK1A in T98G cells by lentiviral transduction of a shRNA and after 5 days of puromycin selection to assure that all the cell population was interfered, we analyzed the volume of the selected infected cells. As shown in Fig. 33A, shDYRK1A-expressing cells had a reduced cell volume if compared with control infected cells. The change in cell volume was not due to the cells being in different phases of the cell cycle, as T98G cells with reduced DYRK1A expression did not display an altered cell cycle profile (Fig. 33B) and have a proliferation

rate similar to control infected cells (Fig. 33C). Although at this stage we cannot rule out that DYRK1A deficiency could be inducing changes in signaling pathways that respond to extracellular osmolarity, the findings are also compatible with changes in the cell growth rate, which will certainly impact on cell size.

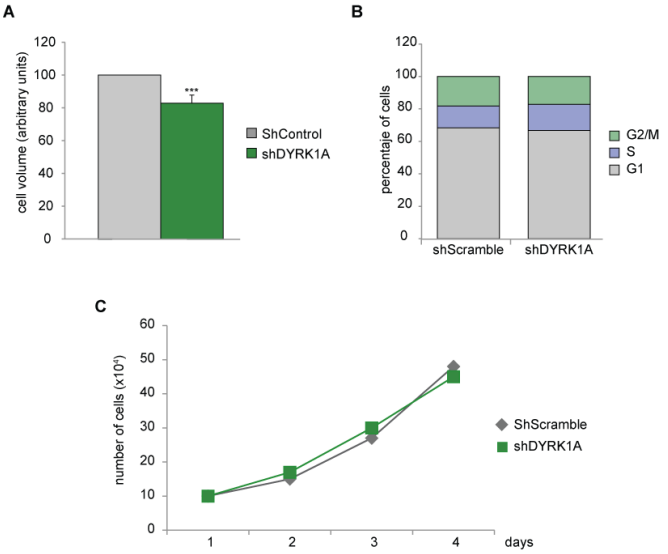


Figure R33: DYRK1A downregulation induces a reduction of cell volume (A) Cell volume of either shDYRK1A or shControl infected cells was measured five days after infection and data represented as arbitrary units, with control cells having a value of 100 (average of three independent measures; ***, p-value<5E-5). **(B)** The cell cycle profile was determined for the same cell population used in **(A)** by FACs analysis. Quantitative representation of the G1, S, and G2/M subpopulations is shown in the graph. **(C)** Cumulative cell curve of either shDYRK1A or shControl infected cells. The graph represents total number of cells.

Increase in cell mass is intimately linked to the availability of nutrients and energy *(Zoncu et al., 2011)*. Molecules such as growth factors, insulin and amino acids signal this availability to the cells, which respond by activating signaling pathways that rapidly alter patterns of gene expression. Targets of theses pathways are thus sensitive to changes in growing conditions, such as nutrient starvation. To understand if the transcriptional activity of DYRK1A on growth-related targets was modulated by changes of the extracellular environment, we first analyzed DYRK1A-dependence of its targets transcripts levels in serum starved cells. Downregulation of DYRK1A has no effect in transcript levels when compared to control cells upon serum deprivation (Fig. R34). The data suggests that the activity of DYRK1A could be mediated by nutrients.

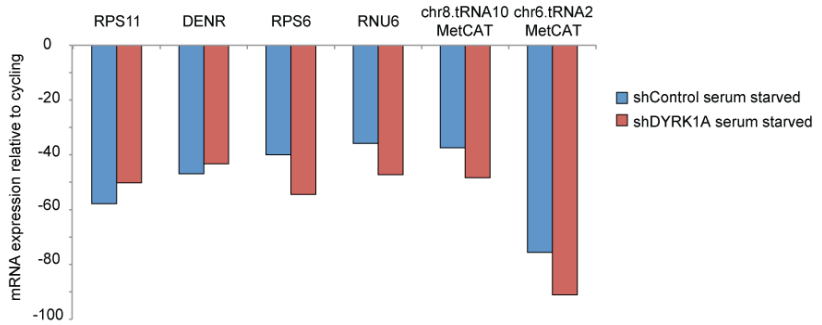


Figure R34: The transcript levels of DYRK1A targets do not respond to DYRK1A depletion in serum starved cells. T98G cells were transduced either with a control lentivirus (shControl) or a lentivirus expressing shRNAs against DYRK1A (shDYRK1A) for five days. Cells were thus collected and maintained for further 48 h in DMEM without FBS to induce serum starvation. Cells were then collected, RNA was prepared and the expression levels of DYRK1A target genes (indicated in the x axis) were determined by RT-qPCR using specific primers. Data represent the percentage of downregulation with respect to the respective cycling control.

To further confirm that DYRK1A-mediated repression is mediated by growth signaling, starved cells were stimulated with insulin and expression levels of DYRK1A-growth related targets was measured by Real Time PCR. As shown in Fig. R35, insulin stimulation led to an increase of the transcription of all DYRK1A targets in shControl cells, in a time dependent manner. However, shDYRK1A infected cells did not respond to insulin stimulation.

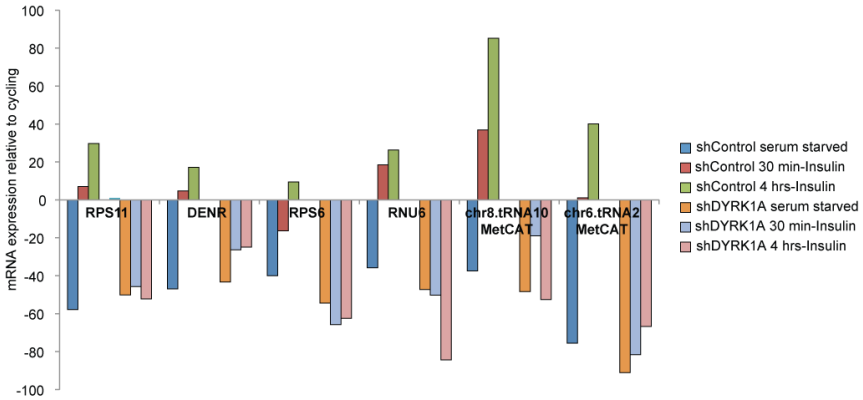


Figure R35: DYRK1A downregulation inhibits DYRK1A-target activation by insulin stimulation. T98G transduced cells with either shControl or shDYRK1A were serum starved for 48 h, and subsequently stimulated with insulin for the times indicated on the right. RNA was prepared from each population and the expression levels of DYRK1A target genes (indicated in the x axis) were determined by RT-qPCR. Data represent the percentage of down- or up- regulation with respect to the respective value in control cycling conditions.

Upon insulin stimulation, the growth signaling pathways acting through mTOR are activated, finally resulting in ribosomal protein S6 kinase beta-1 (p70S6K) phosphorylation, among other signaling events (Fig. 36A). An explanation for the defective response observed in shDYRK1A cells could rely on a defect in the activation of the mTOR signaling cascade. To explore this possibility, the activation of the pathway was monitored by following the phosphorylation status of p70S6K on the activating residue Thr389. Under serum starved conditions, p70S6K is found in a dephosphorylated state, thus inactive, in both shControl and shDYRK1A cells (Fig. R36). After insulin stimulation, mTOR is activated and able to phosphorylate p70S6K, which can be followed by an increase in the signal from the antibody against the phosphosite and by the changes in the electrophoretic mobility of the protein. No significant differences were observed comparing shControl and shDYRK1A cells, indicating that downregulation of DYRK1A does not impair the response to the growth stimuli.

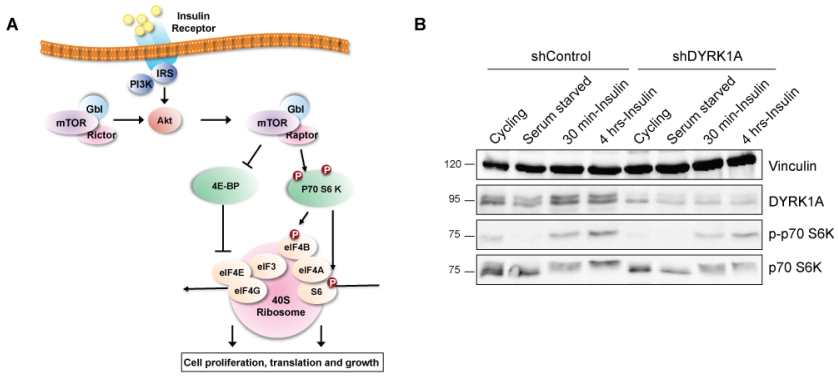


Figure R36: DYRK1A downregulation does not alter mTOR response to insulin stimulation. (A) Schematic representation of a simplified insulin signaling pathway, showing the main players in the signaling cascade. **(B)** Serum starved shControl and shDYRK1A cells stimulated with insulin for the indicated times and the expression of DYRK1A, p70 S6K and p70 S6K phosphorylated in the activating residue Thr389 (p-p70 S6K) was analyzed by Western blotting with specific antibodies. Vinculin was used as loading control.

Giving that downregulation of DYRK1A leads to a decrease in cell volume, that this is accompanied by a downregulation of the expression of growth-related genes, and given the fact that a canonical growth

signaling pathway typically responding to growth stimulus seems to be not altered, we propose that the phenotype of reduced growth could be the consequence of the transcriptional activity exerted by DYRK1A directly on its target genes.

Discussion

DYRK1A nuclear interactome: a connection with pre-mRNA processing and more

DYRK1A localizes both in the nucleus and in the cytosol of different cell types, with a prominent cytosolic staining in most cell types. In fact, most of the DYRK1A substrates described to date are cytosolic proteins, although several nuclear substrates have been described (Fig. I6). However, the activity of DYRK1A on most of those nuclear substrates (for instance, the transcription factors NFATc4 or Gli1, and the splicing factor 9G8/SRSF7) can be explained by phosphorylation events occurring in the cytoplasm, which would, positively or negatively, modulate the ability of the substrates to translocate to the nucleus. In this regard, the examination of the nuclear activities of DYRK1A has been problematic due to the fact that, in contrast with the *in vivo* situation, overexpressed DYRK1A always accumulates in the nucleus *(Alvarez et al., 2003)*. Therefore, it is an unresolved issue whether DYRK1A has specific activities within the nucleus. This Thesis work has demonstrated the presence of a specific pool of DYRK1A within the nuclear compartment, which acts as an active kinase able to interact with and phosphorylate specific substrates in physiological conditions. This is particularly relevant because many of the DYRK1A phosphorylation events described to date have only been demonstrated to occur *in vitro*, with a few exceptions *(de Graaf et al., 2006; Guo et al., 2010; Kurabayashi et al., 2010; Laguna et al., 2008)*, and our approach opens the door to the future identification of DYRK1A physiological substrates in the nuclear compartment.

Within the nucleus, DYRK1A appears to exist as two discrete pools of protein: one, representing a very small percentage of the nuclear DYRK1A, corresponds to the free, monomeric DYRK1A, and the other fractionates at high molecular weights and therefore, should be associated to different macromolecular complexes. Through an unbiased proteomic approach, this Thesis work has identified several DYRK1A novel nuclear interacting proteins, allowing to define the first “DYRK1A nuclear interactome”. An added value of this Thesis work resides on the fact that the identified interactions have been detected in endogenous conditions, without forcing them by overexpression of the putative partners. The existence of DYRK1A as part of macromolecular complexes had been already

suggested for cytosolic DYRK1A *(Aranda et al., 2008)*, and also for its closest family member, DYRK1B *(Zou et al., 2003)*; these observations together with the results of this work raise the possibility that the roles of DYRK proteins are tightly linked to the complexes they belong to.

We are well aware that the approach used does not allowed to distinguish between *bona fide* interactors, that is, those establishing a direct protein-protein interaction with DYRK1A, and those proteins appearing in the screen because of their association to multiprotein complexes. However, the strategy does enable the identification of complexes for which stoichiometry is important for both their formation and/or stability. This is likely the case for the hPrp19/CDC5L complex. The hPrp19/CDC5L complex is an integral component of the spliceosome, and it is required for stable association of U5 and U6 snRNPs with the spliceosome after U4 snRNP is dissociated *(Chan and Cheng, 2005)*. Four out of seven component of the hPrp19/CDC5L core complex were identified within the DYRK1A interactome; in addition, two accessory proteins were also detected in the screen, suggesting that DYRK1A could be recruited to the complex itself. The fact that the association does not depend on RNA further suggests that one or several subunits of the complex could directly interact with DYRK1A. Preliminary attempts to test whether hPrp19 or PLRG1 are direct interactors of DYRK1A have not been successful, so further experiments are required to solve this question. DYRK1A has been proposed as a regulator of RNA processing based on DYRK1A localization in nuclear speckles *(Alvarez et al., 2003)* and on having several alternative splicing factors as substrates *(de Graaf et al., 2004)*. Phosphorylation of some of these splicing factors alters their ability to regulate alternative splicing *(Ding et al., 2012; Qian et al., 2011; Yin et al., 2012)*. However, these Thesis results also suggest that DYRK1A could also modulate splicing events by acting directly on the spliceosome machinery itself.

Further sustaining a link with pre-mRNA processing, DYRK1A nuclear interactome was found enriched in components of the pre-mRNA 3'-end processing machinery, like CPSF or CFI, as well as many of the components of the mammalian exosome (data not shown), the protein complex involved in nuclear degradation of aberrant pre-mRNAs (for a recent review, see *Lykke-Andersen et al.,*

2011). Work from many laboratories has shown that the individual events occurring during eukaryotic gene expression, transcription, splicing, processing, RNA transport and degradation, are coupled together allowing the coordination of regulatory controls (reviewed in, for instance, *Pandit et al., 2008)*. Therefore, we can envision several explanations for the presence of the different complexes in the DYRK1A interactome. One possibility is that only one of the complexes is the DYRK1A target and because of the interconnection among the different players of the pre-mRNA processing machinery, the other complexes are mere passengers in the screen. Alternatively, DYRK1A could be recruited to distinct nuclear complexes, acting at different steps of the production of a mature mRNA, through its interaction with different targets within the complexes, to regulate the different steps in a coordinated fashion. Finally, DYRK1A could be targeting a component common to all the complexes. In this context, it is well known that the RNA Pol II CTD acts as a platform for the assembly of factors that regulate transcription and pre-mRNA processing *(Hsin and Manley, 2012; Maniatis and Reed, 2002)*. Previous data generated in our laboratory demonstrated that DYRK1A is able to interact directly with the CTD *in vitro* (E. Salichs, Doctoral Thesis, 2008), allowing us to hypothesize that the CTD might be the bridge connecting DYRK1A and the individual components of pre-mRNA processing.

We have also shown that DYRK1A is able to interact with HDAC2 and RbAp48/RBBP4, two proteins that are part of several co-repressor complexes. Transcriptional activation can occur when histone acetyltransferases acetylate the core histones of nucleosomes, resulting in chromatin relaxation. Conversely, deacetylation of histones by HDACs results in chromatin condensation and transcriptional repression. All mammalian HDACs contain potential phosphorylation sites and the majority of them have been found to be phosphorylated *in vitro* and *in vivo (Sengupta and Seto, 2004)*. However, in the case of class I HDACs (as HDAC1 and HDAC2) there are contradictory conclusions about the regulatory effects exerted by phosphorylation on the enzymatic activity, transcriptional repression potential and functional protein-protein interactions *(Cai et al., 2001; Galasinski et al., 2002; Sun et al., 2002; Tsai and Seto, 2002)*. DYRK1A phosphorylation might thus modulate the affinity of HDAC2 for key interacting proteins, regulating in turn its enzymatic activity. In the case of class II

HDACs, phosphorylation regulates their subcellular localization and therefore their biological activities. Of note, the closest homologous to DYRK1A, DYRK1B, phosphorylates HDAC5 at a conserved site within class II HDACs that lies within the NLS, and reduces HDAC5 nuclear accumulation in a dose-dependent and kinase-dependent manner *(Deng et al., 2005)*.

DYRK1A is a novel regulator of transcription

The interaction between the protein DYRK1A and components of the PICs (RNA Pol II and TAF15), as well with both subunits of p-TEFb, the catalytic CDK9 and the regulatory cyclin T1, led us initially to propose a model by which DYRK1A could be recruited to specific promoters either through transcription factors, which can be in turn phosphorylated by the kinase, or through components of the transcriptional machinery itself (as the CTD of the RNA Pol II). Indeed, a genome-wide search for DYRK1A-chromatin interactions has led to the identification of transcriptional targets of this kinase.

To examine the role of DYRK1A in transcription, we took an one-hybrid approach by targeting DYRK1A to a promoter with chimeric constructs. Our results showed that DYRK1A is able to activate gene transcription when recruited to a promoter and that the effect is dependent on DYRK1A kinase activity. DYRK1A transcriptional activity is observed both with a complex promoter, such as the HIV-1, and with a minimal promoter region containing solely the TATA-box, suggesting that no specific transcription factor binding sites are needed for activation by DYRK1A, as it is in the case of CDK9 *(Montanuy et al., 2008)*. Moreover, the fact that the DYRK1A transactivation capabilities are higher when it is tethered to the promoter region at a nearer TSS distance (100-fold for the E1b reporter versus 10-fold for the HIV-1 reporter) might be indicating that DYRK1A needs to be positioned in close proximity to the components of the basal transcription machinery to exert its effects. In this regard, DYRK1A-dependent transcription was heavily reduced, although not completely abolished, when the TATA-box was removed, suggesting that this core promoter element facilitates DYRK1A-transactivation. Transcription complexes formed on a promoter lacking TATA-box elements might be depleted of critical components, and therefore being impaired for DYRK1A-mediated

transactivation. We can thus hypothesize that PIC assembly on the TATA-box element could be the step targeted by DYRK1A, and that RNA Pol II transcription complexes that assemble on a TATA-box sequence via TBP-recruitment might mediate the transcriptional activity of DYRK1A.

Two other pieces of evidence point in this direction. First, the DYRK1A-target regions found in the wide-genome screen are within the -300 to -50 proximal promoter regions. Second, and although almost all DYRK1A-bound promoters identified were TATA-less, most of them were positive for TBP recruitment, according to the ENCODE data, with TBP-peaks almost overlapping with the DYRK1A-ChIP peaks. Nevertheless, we should not forget that promoter recognition and subsequent recruitment of TFIID and RNA Pol II holoenzyme can be mediated by others TAFs rather than only TBP in TATA-less promoters (for a review, see *Thomas and Chiang, 2006)*, which account for the 90% of promoters in the human genome *(Kim et al., 2005)*, and that the diversity in core promoter recognition complex composition is at the basis of several differentiation-specific transcriptional programs (for a review see, *D'Alessio et al., 2009) (Muller et al., 2010)*. The presence of TAF15 in the DYRK1A complexes could be an indication of this possibility.

In the same context, it is worth mentioning that cyclin T1-associated complexes containing DYRK1A have been detected, in which CDK9 is absent. Moreover, the detection of DYRK1A-Cyclin T1 complexes under overexpressing conditions (C. Di Vona, unpublished results) suggests that these two proteins could interact directly. The functional interaction between the two proteins is further sustained by results not included in this Thesis work: i) co-expression of cyclin T1 with DYRK1A enhances the nuclear accumulation of cyclin T1, and ii) cyclin T1 induces the redistribution of DYRK1A from the nuclear speckles to smaller nuclear foci that resemble those formed during active transcription (C. Di Vona and S. de la Luna, unpublished). Therefore, specific DYRK1A complexes containing cyclin T1 could exist, in which DYRK1A could act as the catalytic subunit of a specific “pTEFb-like” complex, regulating the transcription of a subset of target genes. Notably, cyclin T1 has been shown to be the recruiter element for CDK9 to the viral transactivator protein (Tat) and transactivation response element RNA (TAR) of HIV-1, and not just the obligatory activating cyclin for CDK9 *(Wei et*

al., 1998). Cyclin T1 might thus be a good candidate to direct DYRK1A to the core promoter of a subset of target genes.

Once in the promoter, the transactivation capability of DYRK1A depends, at least in part, on its catalytic activity. Therefore, one or several targets within the transcription apparatus should be direct substrates of DYRK1A. One important aspect of transcriptional regulation is the phosphorylation of components of the transcription apparatus. As discussed in the Introduction section, the phosphorylation state of the RNA Pol II CTD is controlled by a variety of protein kinases, which govern the "paused" or "elongating" phases of the transcription cycle. The ability of DYRK1A to activate transcription could thus be explained by phosphorylation of the CTD. In fact, DYRK1A is able to phosphorylate *in vitro* the CTD on both Ser5 and Ser2, but we cannot provide yet experimental evidence on these phosphorylation events occurring *in vivo*. However, we favor a model that involves a functional relationship of DYRK1A with the CTD, based on the transcriptional behavior of the human DYRK family members. Only DYRK1B, the closest member of the family, was able to induce transcription similarly to DYRK1A, an activity that correlates with the ability of interacting with the RNA Pol II CTD shared only by these two DYRK kinases.

On the other hand, the transcriptional activities of, for instance, JNK1 or Pim1 when recruited to promoter regions appear to rely on the phosphorylation of H3-S10 *(Sun et al., 2012; Tiwari et al., 2012; Zippo et al., 2007)*, and this could be also the case for DYRK1A. However, we consider this possibility unlikely because, although DYRK1A phosphorylates histone H3, it does it only on serine residue 45 *(Himpel et al., 2000)*. A role for this histone H3 modification is still not known.

At this stage, we cannot identify which of the steps of the transcription cycle is the target of DYRK1A or whether it is involved in modulating several of them. Phosphorylation of the CTD and/or other components of the basal transcriptional machinery is surely involved in the DYRK1A transcriptional activity, because of the kinase dependence observed in the one-hybrid assays. In addition, given that a smaller but consistent transcriptional activity was detected while using a kinase-death mutant of DYRK1A, a scaffold role can also be envisaged. In this case, the kinase could participate

in transcription initiation by acting as an adaptor connecting coactivators/corepressors and chromatin modifying enzymes to the general transcription machinery. Such non-catalytic functions have been, for instance, described for *S. cerevisiae* Mpk1, which it is able to activate transcription initiation and elongation of a subset of genes in response to cell wall stress independent of its kinase activity *(Kim and Levin, 2011)*.

Finally, and as an extension of this model, DYRK1A might also recruit splicing factors to the elongating polymerase. This would increase the local concentration of these factors, facilitating the assembly of a spliceosome when splice sites emerge from the polymerase. The phosphorylation of splicing factors could act as a positive feedback loop, since it has been demonstrated that splicing promotes transcriptional elongation *(Fong and Zhou, 2001; Tripathi et al., 2012)*.

DYRK1A is recruited to distinct genome-wide RNA Pol II and Pol III promoter regions

All the facts discussed in the previous sections led us to consider the idea that DYRK1A may associate with specific chromatin regions to regulate transcription. In fact, the approach used in this Thesis based on chromatin immunoprecipitation with massively parallel sequencing has allowed to map genome-wide chromatin interaction sites of DYRK1A. The data indicates that DYRK1A mainly associates to RNA Pol II promoter regions in close proximity to TSSs. The majority of DYRK1A-associated RNA Pol II promoters lie in fact in CpG islands, a common feature of human promoters. However, when the DYRK1A-associated promoters were scanned for the presence of particular core promoter elements *(Palaniswamy et al., 2005)*, no particular core promoter signature was observed. Of note, and as discussed above, although almost all DYRK1A-bound promoters were TATA-less, all of them were TBP positive, suggesting that the presence of TBP is important either for the recruitment of DYRK1A to target promoter regions or for the DYRK1A transcriptional activation capabilities. Finally, since common subsets of target regions have been found in two different cell types and in both proliferating and quiescent cells, it is reasonable to think that recruitment of DYRK1A to genomic regions

is a general feature of this kinase. Nevertheless, considering the proposed role for DYRK1A in several cell differentiation processes (reviewed in *Aranda et al., 2011)*, it is possible that DYRK1A regulates, at the chromatin level, specific subset of genes, an issue that surely deserves future attention.

Apart from promoters, a small proportion of DYRK1A-bound regions correspond to intergenic regions. Manual inspection of these genomic locations revealed that almost all of them correspond to regions positive for TBP binding and in the near proximity of a CpG island, both features indicative of the presence of a promoter and/or a transcriptional regulatory region. It could be speculated therefore that, in addition to the annotated promoter regions, a fraction of DYRK1A binding events occur at elements that might correspond to novel regulatory regions or unannotated promoters of downstream transcription units. Beside that, the DYRK1A-peaks in intergenic regions overlapped with RNA Pol III genes, both type 2 (tRNAs) and type 3 (Y, U6, 7SK, U6atac, $tRNA^{SeC}$), a result that could be either the consequence of the high degree of co-occupancy of the RNA Pol II and Pol III basal transcriptional machineries in the mammalian genome genes *(Barski et al., 2010; Moqtaderi et al., 2010; Oler et al., 2010)* or reflect a functional recruitment of DYRK1A to RNA Pol III transcriptional units (see below).

Promoter regions of active genes have reduced nucleosome occupancy and elevated histone acetylation *(Bernstein et al., 2002; Bernstein et al., 2004; Lee et al., 2004; Pokholok et al., 2005)*. DYRK1A-associated islands, in contrast to the surrounding regions, present histone acetylation marks suggestive of being depleted of nucleosomes. This chromatin environment would facilitate binding not only of DYRK1A but also of activators, with the consequent recruitment of the basal transcription machinery. Moreover, high levels of H3K4me2, and H3K4me3 are detected in DYRK1A-associated promoters surrounding TSSs, whereas H3K36me3 mark peaks after the TSSs and is present all along the gene body. This pattern of histone modifications is generally associated with promoter regions of active genes *(Bernstein et al., 2005; Roh et al., 2005)*, and in agreement with the transactivation potential showed by DYRK1A in one-hybrid assays, support the hypothesis of DYRK1A having a functional role in the positive regulation of the expression of its targets *in vivo.* Notably, depletion of DYRK1A correlates with the

downregulation of the expression of some of its targets in the cell line tested, further sustaining the idea that DYRK1A might act as an activator of transcription when recruited to promoter regions *in vivo*.

Further evidence for a direct role of DYRK1A in transcriptional regulation came from *in silico* approaches looking at expression correlations of DYRK1A and its target genes. Thus, DYRK1A target genes have higher expression levels in those tissues presenting higher DYRK1A expression and are underexpressed in those tissues with low DYRK1A levels, maybe indicating that the activity of DYRK1A on certain genes could be relevant only in some particular tissues, due to specific physiological stimuli. As an extension of this suggestion, deregulation of DYRK1A could lead to changes in the expression levels of its target genes. The idea is supported by the extremely high correlation of expression levels we have found in pathological situations such as in DS brains, condition in which is well established that DYRK1A is overexpressed *(Guimera et al., 1999)*, and in sperm cells from males with severe teratozoospermia, condition in which DYRK1A is almost absent (based on GEO database data). These findings suggest therefore that DYRK1A acts as an activator of transcription and that when DYRK1A protein levels are deregulated, the downstream physiological effects of its gain-or-loss could partly operate through the direct alteration of specific gene transcription profiles. Interestingly, the genes containing the DYRK1A-motif within their promoters, and not only those experimentally found in this work, show up-regulated expression in the cells hematopoietic lineage *(Xie et al., 2005)*, which are the ones showing the higher levels of expression for DYRK1A.

DYRK1A is recruited to RNA Pol II promoters through a highly conserved consensus sequence

The palindromic sequence -TCTCGCGAGA- has been found highly enriched in the DYRK1A-bound genomic regions. No such a similar motif was found in DYRK1A-associated regions that were not annotated as promoters or that did not have features of a putative promoter (CpG island + TBP presence). A motif very similar to this one -TMTCGCGANR- has been previously identified as one of the top ten most strongly enriched and well-conserved motifs in human promoters *(Pique-Regi et al., 2011; Xie et al., 2005)*. This

palindromic sequence has been found in TATA-less promoters *(Wyrwicz et al., 2007)*, as this is the case for the DYRK1A-target genes containing the motif, and it has been proposed to be a potential *cis*-regulatory element in 5% of human genes (including cell cycle, transcription regulators and ribosomal protein genes) *(Roepcke et al., 2006; Wyrwicz et al., 2007)*. Taking this, together with the observed high degree of conservation of the consensus sequence within placental mammals, one can speculate that the motif could be considered a novel core promoter element, which might play a role as a general transcription factor binding site facilitating the recruitment and the assembly of the polymerase and all the specific associated co-factors. Indeed, the TCTCGCGAGA motif has been characterized as the regulatory element within both the ADP-ribosylation factor 3 and HNRNPK promoters *(Haun et al., 1993; Mikula et al., 2010)*.

Based on all these observations and on the fact that the motif is positioned in the center of the DYRK1A-associated peaks, we propose that the TCTCGCGAGA-motif is the genomic sequence responsible for the recruitment of DYRK1A, directly or indirectly, to target promoters. Unfortunately, we have not been able to undoubtedly prove a direct interaction of DYRK1A to the TCTCGCGAGA-motif. A bacterially purified GST-DYRK1A fusion protein failed to bind a double-strand oligonucleotide harboring the consensus motif. Even though GST-DYRK1A expressed in bacteria behaves as an active kinase *(Alvarez et al., 2007)*, we cannot exclude the possibility that the protein is not properly folded to allow DNA binding, or even the existence of other interfering problems. Alternatively, DYRK1A does not bind directly to the palindromic DNA sequence but needs other accessories proteins to form a stable complex. In this regard, no classical DBDs are detected in the primary structure of DYRK1A, although the existence of unconventional DBDs, as in the case of ERK2, cannot be formally ruled out *(Hu et al., 2009)*. The ERK2 DBD lies within the CMGC-insert just downstream helix G within the catalytic domain, and the ability to interact with DNA has been associated to the presence of positively charged amino acids exposed in the surface of the catalytic domain (Fig. D1A); mutation of Lys 259 and Arg261 within the DBD is enough for abolishing ERK2 DNA binding activity *(Hu et al., 2009)*. In DYRK1A, this region would correspond to the first half of the complex NLS2 between helix G and helix H within the CMGC-

insert, which also is exposed in the crystal structure (Fig. D1C) *(Soundararajan et al., 2013)*. The two amino acids crucial for the ERK2 DNA binding activity are well conserved not only in DYRK1A but also in other class I DYRKs (Fig. D1B). Therefore, a direct binding of DYRK1A to DNA through a dedicated motif cannot completely rule out. Notably, evidences for binding of the CTD of the RNA Pol II to this region have been already generated, supporting a possible role as a scaffold module for the CMGC-insert of DYRK1A.

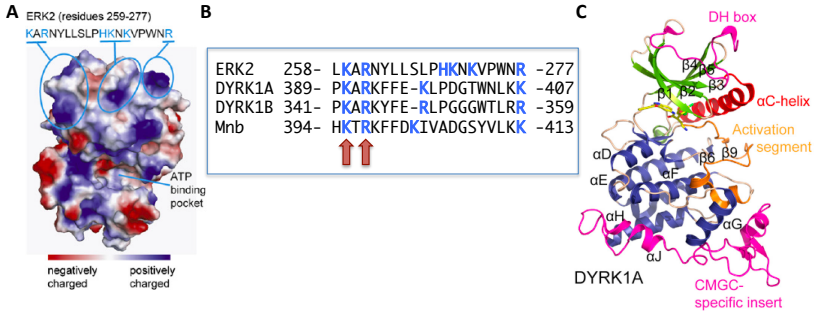


Figure D1. DNA binding domain in ERK2. (A) Structural analysis for DNA-binding domain in ERK2. A surface patch (residues 259–277) comprised of three positively charged clusters are indicated with the amino acid sequence showing above (adapted from *(Hu et al., 2009)*. **(B)** Amino acid sequence comparison of the indicated protein fragment of human ERK2, human DYRK1A, human DYRK1B and Drosophila class I DYRK, minibrain (Mnb). The positively charged clusters in **(A)** are indicated in blue. The two residues shown important for DNA binding activity in ERK2 are indicated with red arrows *(Hu et al., 2009)*. **(C)** Structure of DYRK1A kinase domain. The DH-box and CMGC-specific inset are shown in magenta and the activation segment in orange (adapted from *(Soundararajan et al., 2013)*.

Although binding of purified DYRK1A to the motif was not detected, several other reasons support the proposal that DYRK1A participates in transcriptional complexes that are bound to the TCTCGCGAGA-motif. First, the striking enrichment of DYRK1A-associated promoters that contain the consensus (83%) suggests specificity of DYRK1A binding to such regions. No other particular genomic feature, apart from CpG-islands, has been detected within these promoters. Second, the formation of protein complexes on a double strand oligonucleotide harboring the consensus is completely abolished by incubation with a specific antibody to DYRK1A, suggesting that DYRK1A might be a part of the complexes built on it. In addition, a weak supershift was observed in EMSA experiments with another DYRK1A specific antibody that indeed was successfully used in ChIP experiments.

Very recently, the transcription factor KAISO/ZBTB33 was found bound to TCTCGCGAGA-containing pre-adypocite specific promoters *(Raghav et al., 2012)*. Moreover, within the framework of the ENCODE project, KAISO and two other transcriptional modulators, CHD2 and BRCA1, were identified as putative TCTCGCGAGA interacting factors *(Wang et al., 2012)*. As in the case of DYRK1A, the analysis of the peaks revealed exact motif positioning around the peak center for all three factors (Fig. R29A, C. Di Vona and D. Bezdan, unpublished results). However, only KAISO, a vertebrate specific transcriptional repressor of the BTB (POZ)- and zinc finger-domain-containing protein family, has DNA binding activity *(Daniel et al., 2002)*. Moreover, only KAISO displays a genomic pattern of association similar to the one found for DYRK1A (Fig. D2). In fact, while in the case of CHD2 and BRCA1 less than the 30% of associated promoters contain the motif, more than the 60% of KAISO-associated promoters are motif-positive, suggesting that this transcription factor could be the one directly and specifically interacting with the TCTCGCGAGA motif, bridging DYRK1A to the consensus TCTCGCGAGA-motif at specific genomic locations.

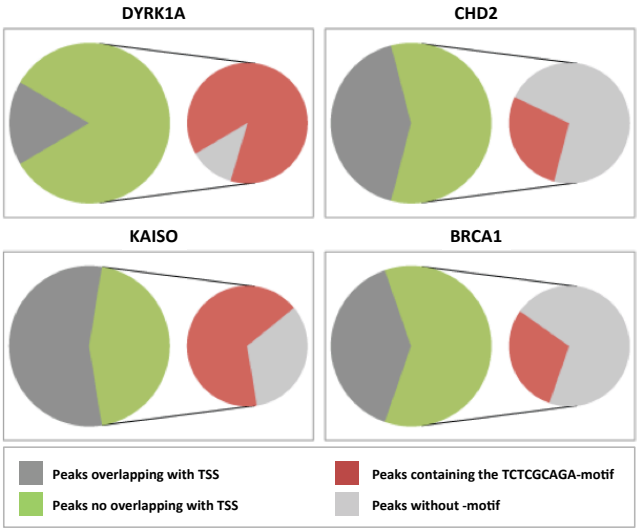


Figure D2: Intersection of BRCA1, CHD2, and ZBTB33/KAISO peaks with DYRK1A peaks and associated motif. For each protein ChIP data, the large pie chart shows the fraction of peaks not overlapping (in dark gray) or overlapping (in green) with promoter regions from SW-TSS-1.1k. The small pie chart shows the fraction of promoter-peaks harboring a TCTCGCGAGA motif (in red) or without the motif (in light gray).

Concurring with this proposal, down-regulation of KAISO leads to a partial decrease in DYRK1A binding to target promoters, while KAISO and DYRK1A do not interact in soluble extracts. However, an intrinsic contradiction exists though with the proposal of KAISO acting as a bridge between DNA and DYRK1A. KAISO is a bi-modal transcriptional repressor and a p120 catenin-interacting protein (reviewed in *van Roy and McCrea, 2005)*, which binds two types of DNA sequences located within gene-regulatory regions, methylated CpG islands, and sequence-specific sites: TCCTGCNA (reviewed in *Daniel et al., 2002)* and the DYRK1A-motif TCTCGCGAGA *(Raghav et al., 2012)*. Kaiso recruits co-repressor components in both contexts (reviewed in *van Roy and McCrea, 2005)*, and in agreement it has been found associated to DNA regions enriched in the repressive mark H3K9me3 *(Yoon et al., 2003)*. Considering all the evidence that point to a role for DYRK1A as a transcriptional activator, it is challenging to think on a transcriptional repressor being the recruiter. However, in contrast to previously published data, we observed that KAISO associated regions within TCTCGCGAGA-specific promoters are accessible and enriched in activating chromatin marks and in RNA Pol II, suggesting that this subset of KAISO-target genes are actively transcribed. A scenario can be envisioned in which, in basal conditions, KAISO recruits DYRK1A to its target genes, which are primed but not active. Upon stimuli, the interaction between DYRK1A and components of the basal machinery could be stabilized, KAISO might dissociate form the promoters and hence DYRK1A-mediated transcriptional activation occurs. In an alternative model, KAISO does not leave the DNA and DYRK1A is actively involved in releasing the repressor state by acting on KAISO or on those co-repressors recruited by KAISO such as NCOR1 (nuclear receptor co-repressor 1) or SIN3A.

DYRK1A binds to RNA Pol III dependent genes

Aside from the presence of DYRK1A within RNA Pol II proximal promoter regions, endogenous DYRK1A also associates with RNA Pol III transcribed genes. With the exception of RNA Pol III type I transcripts (5S and 5S related genes), both type II (tRNAs) and type III (Y, U6, 7SK, U6atac, $tRNA^{SeC}$) RNA Pol III-dependent genes were identified as targets of DYRK1A, in both HeLa and T98G cells. As discussed above, the co-occurrence could just be the consequence

of the high degree of co-occupancy of the RNA Pol II and Pol III basal transcriptional machineries *(Barski et al., 2010; Moqtaderi et al., 2010; Oler et al., 2010)*: RNA Pol II binding peaks are usually found at around 200 bp upstream of the RNA Pol III initiation site and, in most cases, they do not correspond to a known RNA Pol II transcription unit. Therefore, DYRK1A could be present at these genomic loci because of its association with RNA Pol II. However, down-regulation of DYRK1A induces a decrease in the expression of several of the RNA Pol III target-genes, such as $tRNA10^{Met}$, $tRNA2^{Met}$ and RNAU6, strongly suggesting that DYRK1A is an activator of RNA Pol III gene expression. Sustaining this proposal, it has to be mentioned that green tea component EGCG, and inhibitor of DYRK kinases, has been shown to inhibit transcription *in vitro* from the U6 and the VAI promoter, two RNA Pol III dependent promoters *(Jacob et al., 2007)*.

How is DYRK1A recruited to the RNA Pol III promoters? The results collected within this Thesis support the idea that DYRK1A is recruited to genomic regions through the binding to the palindromic consensus sequence TCTCGCGAGA. However, RNA Pol III genes lack such motif. Furthermore, no enrichment in any DNA motif, apart from the characteristic promoter elements of the RNA Pol III transcribed genes, were found in this DYRK1A-gene subset. These data raise the possibility that DYRK1A is recruited to these genes through the interaction with the RNA Pol III transcription machinery. The RNA Pol III apparatus consists of three complexes: the RNA Pol III enzyme (which comprise 17 subunits, some of them common to either RNA Pol I and Pol II), and the general factors TFIIIB and TFIIIC required for transcription initiation and for promoter recognition, respectively (for a recent review, see *White, 2011)*. Of note, three out of six subunits of the transcription factor TFIIIC (IIIC-220, IIIC-110 and IIIC-90) were found in the DYRK1A nuclear interactome. Although we have not been able to validate the interaction between DYRK1A and TFIIIC-220 (data not shown), it cannot be excluded that the interaction could exist between DYRK1A and the other members of TFIIIC, and/or that the interaction could be restricted to the DNA-bound proteins, and therefore not easily detectable. In addition, TBP is a subunit of TFIIIB and therefore an essential component of the basal RNA Pol III transcription machinery, and it may represent a recruitment element for DYRK1A common to both RNA Pol II and RNA Pol III promoters.

Very little is known yet about gene specific regulators of RNA Pol III transcription, except maybe for the tumor suppressor p53 and the regulator MAF1, although it has recently become clear that variations in usage of individual tRNA genes between different cell types clearly exist and that therefore, RNA Pol III transcription has to be tightly regulated (reviewed in *White, 2011)*. In this regard recent reports have linked phosphorylation of components of the RNA Pol III transcriptional machinery to cell responses to changes in environmental conditions. This is the case of the phosphorylation of Rpc53p, a TFIIF-like subunit of yeast RNA Pol III, by Kns1p (a CLK ortholog) and Mck1p (a GSK3 ortholog) as downstream effectors of TOR signaling *(Lee et al., 2012)*. In fact, the presence of mTOR has been detected at different RNA Pol III-transcribed genes and phosphorylation of the transcriptional inhibitor MAF1 at those loci has been associated to mTOR-dependent activation of RNA Pol III transcription *(Kantidakis et al., 2010)*. Our results point to DYRK1A being one of such kinases able to directly regulate RNA Pol III transcription.

Transcriptional regulation of RNA Pol II and RNA Pol III translation-related genes links DYRK1A to cell growth control

DYRK1A-bound RNA Pol II genes show significant enrichment in GO terms associated with protein translation, including ribosomal proteins (RPS6 or RPS11) and factors directly involved in the translation process (EIF4A3 or DENR). In addition, tRNAs are also targets of DYRK1A and, at least for the ones tested, depletion of DYRK1A correlates with a reduction in their expression. Given that protein translation is intimately linked to the increase in mass associated to cell growth, the reduction in cell volume resulting from DYRK1A depletion might be directly related to the reduced transcription of the group of target genes related with the control of protein synthesis.

The contribution of DYRK1A to normal growth could be particularly relevant in cell types that depend on reaching a certain size to perform correctly their functions, as in the case of, for instance, neurons or skeletal muscle cells. In this line, the central nervous system defects shown by *Dyrk1a* heterozygous mice have been

associated to both cell number and cell size defects *(Fotaki et al., 2002)*.

Cell growth and the cell cycle are coordinated but separable processes in eukaryotes (reviewed in for instance, *Goranov and Amon, 2010; Jorgensen and Tyers, 2004)*. In yeast, growth regulates cell division through mechanisms that sense cell size: the "size checkpoint" regulates cell cycle progression until cells reach a minimal o critical size. Noteworthy, a key element in the sensing mechanism in fission yeast is the class II DYRK kinase Pom1p *(Moseley et al., 2009)*. In contrast, in animal cells the existence of a size-checkpoint is still a controversial issue, and genetic evidence from animal models (fly and mice) mutated in genes belonging to growth factor signaling pathways suggest that, in certain contexts, cell growth and division are not coupled (reviewed in *Cook and Tyers, 2007)*. In the cell system we have used to manipulate the levels of DYRK1A, the reduction in cell size is not accompanied by defects in cell proliferation. The cell size reduction in DYRK1A-depleted cells could still fulfill the requirement of the critical mass, and thus allow cells to progress normally throughout the cell cycle. However, the reduction of cell volume in DYRK1A knocked down cells seem to increase with cell passages (data not shown), and it remains to be established whether this progressive reduction could finally results in the alteration of proliferation rates. Again, DYRK1A-growth related activities could be more important in those cell types is which growth can occur in absence of proliferation as is the case of neurons.

The size of an organism is the result of the contribution of both cell size and cell number. Genetic manipulation in flies and mice has shown that defects in signaling pathways that control translation or ribosomal biogenesis result in body size defects. In this line, *minibrain* (*DYRK1A* ortholog) mutant flies are smaller in size if compared to wild type flies *(Tejedor et al., 1995)*, and *Dyrk1a* heterozygous mice present a significant body size reduction *(Fotaki et al., 2002)*.

Among the different signaling pathways that regulate growth in mammalian cells, mTOR (mechanistic target of rapamycin, previously referred to as mammalian target of rapamycin) plays a key role in coupling cell growth with the nutritional status of the cell

(see a simplified scheme for this signaling pathway in Fig. R36A; for a recent review, see *Russell et al., 2011)*. Upstream effectors of this pathway are insulin and insulin-like growth factors, important regulators of growth and metabolism (recent review in *Zoncu et al., 2011)*. Therefore, we have explored the DYRK1A-transcriptional activity in response to insulin as a growth inducer. In the absence of nutrients, the mRNA levels of DYRK1A-targets showed no differences between control and DYRK1A down regulated cells, suggesting that DYRK1A is not transcriptionally active. On the contrary, DYRK1A knockdown strongly impaired insulin response, measured as increased transcript levels of DYRK1A-targets. The intracellular pathway that responds to insulin via mTOR activation is not altered in DYRK1A knocked down cells, indicating that DYRK1A acts downstream the activation of, at least p70 S6K. ChIP results showed that DYRK1A interaction with a subset of promoters, including those assayed for insulin response, is independent of the growing status of the cells, suggesting that the transcriptional activity of DYRK1A might be regulated by the growth stimulus rather than by its recruitment to DNA. DYRK1A might sit on paused promoters in an inactive state, waiting for the proper signal to exert its role in promoting transcription of target genes. Further studies are thus needed to explore the possibility of insulin being the upstream activator that mediated transcriptional activity of DYRK1A.

Final remarks

All the data collected within this Thesis work allowed us to build a model for DYRK1A-mediate transcription (Fig. D3). In this model, DYRK1A is recruited to RNA Pol II-dependent promoter regions of specific target genes through the DNA consensus sequence TCTCGCGAGA. The recruitment of DYRK1A to these promoters could be mediated by the transcription factor KAISO. The experimental data do not allow, however, to exclude that the binding could be mediated by one or more other transcription factors or directly by factors of the basal machinery itself, as the CTD. Once on target promoters, DYRK1A might stimulate transcription of its target genes by different mechanism: i) it could act on chromatin remodeling factors to facilitate accessibility for the binding of co-activators needed for transcriptional activation; ii) it could act on others components of the basal transcriptional machinery; iii) or it

could directly phosphorylates the CTD of the RNA Pol II, both on Ser5 thus stimulating initiation of transcription and/or on Ser2 to stimulate promoter clearance and subsequent transcriptional elongation.

In addition, DYRK1A is recruited to RNA Pol III-transcribed regions by a still unknown mechanism, which could probably involve RNA Pol III associated factors, as TFIIIC. Given the close proximity of RNA Pol III islands to RNA Pol II bound regions, a possibility exists that this recruitment could be mediated also by factors from the RNA Pol II-associated machinery itself. The target for DYRK1A-dependent transactivation on these promoters needs still to be found.

Upon a growth stimuli, DYRK1A regulates transcription of several RNA Pol II and Pol III specific target genes, activity that have an impact on cell growth.

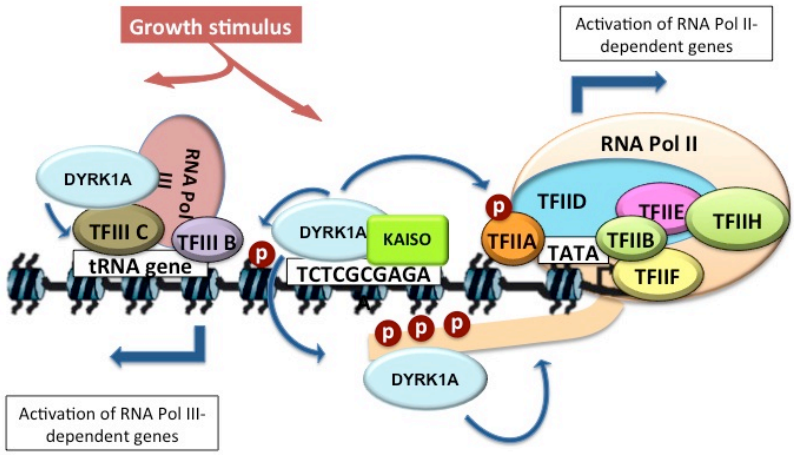


Figure D3: Schematic representation of the model for DYRK1A activity on genomic regions. See text for details

Conclusions

1. Nuclear DYRK1A behaves as an active kinase able to interact and phosphorylates specific substrates within the nucleus.

2. DYRK1A participates in distinct macromolecular complexes within the nucleus, including complexes involved in the pre-mRNA processing, as well as with chromatin remodeling factors *in vivo*.

3. DYRK1A associates with RNA polymerase II containing complexes, together with factors of both the pre-initiation and elongation steps of transcription, including the heterodimer cyclin T1-CDK9 (the transcription elongation factor, P-TEFb).

4. DYRK1A interacts with cyclin T1 in CDK9-free complexes, suggesting that it could act as a catalytic subunit in P-TEFb like complexes.

5. DYRK1A is able to activate transcription in a kinase dependent manner from TATA-box containing promoters, when artificially tethered to DNA.

6. DYRK1A-transcriptional activity is shared by the closest family member DYRK1B, but not by other human DYRK family members.

7. DYRK1A is recruited predominantly to proximal promoter regions, as shown by genome-wide kinase-chromatin interactions, independently of the proliferative status of the cells.

8. The recruitment of DYRK1A to promoter regions depends on the presence of the highly conserved DNA motif TCTCGCGAGA within target promoters. This recruitment is, partly mediated by the transcription factor KAISO.

9. DYRK1A is associated to intergenic regions that correspond to RNA polymerase III transcripts.

10. DYRK1A associated regions are free of nucleosomes and enriched in chromatin marks associated to actively transcribed genes.

11. DYRK1A is able to modulate the expression of its RNA polymerase II- and III-target genes *in vivo*.

12. DYRK1A depletion impairs cell growth.

13. The transcriptional activity of DYRK1A is silenced in conditions of nutrient deprivation.

14. DYRK1A mediates the transcriptional response of its target genes to insulin stimulation in a mTOR-independent manner.

Abbreviations

AD: Activation domain
Aβ: β-amyloid protein
ATP: adenosine triphosphate
AP4: activating enhancer binding protein 4
Brd4: bromodomain protein 4
CDK: cyclin-cependent kinase
ChIP: Chromatin Immunoprecipitation
CKI: casein kinase I
CLK: CDK-like kinase
CMV: cytomegalovirus
CNS: central nervous system
CPSF: cleavage/polyadenylation specificity factor
CREB: CRE-binding protein
Cstf: cleavage stimulation factor
CTD: C-terminal domain
C-terminal: carboxy-terminal
DBD: DNA binding domain
DMEM: Dulbecco's modified Eagle's medium
DNA: desoxiribonucleic acid
DS: Down syndrome
DTT: dithiothreitol
DYRK: dual-specificity tyrosine-phosphorylated and regulated kinase
DH-box: DYRK-homology box
EDTA: ethylendiamine tetracetic acid
EGCG: (-)-epigallocatechin-3-gallate
EMSA: electrophoretic mobility shift assay
ER: estrogen receptor
ERK: extracellular signal-regulated kinase
FBS: foetal bovine serum
G4: Gal4
GSK: glycogen synthase kinase
GST: glutathione-S-transferase
GTF-II: general transcription factor - II
HDAC: histone deacetylase
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HNE: HeLa nuclear extracts
HIPK: homeodomain-interacting protein kinase
Hog1p: high-osmolarity glycerol 1p
IP: immunoprecipitation
IPA: Ingenuity Pathway Analysis
IVK: *in vitro* kinase assay
JNK: c-Jun amino (N)-terminal kinase
kDa: kilodalton
Luc: luciferase
MAP1B: microtubule-associated protein 1B
MAPK: mitogen activated protein kinase
miRNA: micro-RNA
mRNA: messenger RNA

MS: mass spectrometry
NFAT: nuclear factor of activated T-cells
NLS: nuclear localization signal
NP-40: nonidet P-40
N-terminal: amino-terminal
PAGE: polyacrylamide gel electrophoresis
PBS: phosphate-buffered saline
PCR: polymerase chain reaction
PEST: region rich in Pro, Glu, Ser and Thr residues
PIC: pre-initiation complex
PRPF4: pre-mRNA processing protein 4 kinase
P-TEFb: positive transcription elongation factor b
RANKL: receptor activator for nuclear factor κ B ligand
REST: RE1 silencing transcription factor
RNA PoI: RNA polymerase
RNL: *Renilla* luciferase
rRNA: ribosomal RNA
RT: room temperature
SDS: sodium dodecil sulfate
shRNA: short hairpin RNA
snRNA: small Nuclear RNA
snRNP: small nuclear ribonucleoprotein particles
SPRED: Sprouty-related Ena/vasodilator-stimulated phosphoprotein homology-1 (EVH1) domain-containing protein
SR: serine/arginine rich protein
TAF: TBP-associated factor
TBP: TATA-Binding Protein
TFII: transcription factor II
TSS: transcription start site
tRNA: transfer RNA
WB: Western blot
wt: wild-type

References

Adayev, T., Chen-Hwang, M.C., Murakami, N., Lee, E., Bolton, D.C., and Hwang, Y.W. (2007). Dual-specificity tyrosine phosphorylation-regulated kinase 1A does not require tyrosine phosphorylation for activity in vitro. *Biochemistry 46*, 7614-7624.

Ahn, K.J., Jeong, H.K., Choi, H.S., Ryoo, S.R., Kim, Y.J., Goo, J.S., Choi, S.Y., Han, J.S., Ha, I., and Song, W.J. (2006). DYRK1A BAC transgenic mice show altered synaptic plasticity with learning and memory defects. *Neurobiol Dis 22*, 463-472.

Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J.C., Mann, M., and Lamond, A.I. (2000). Functional analysis of the human CDC5L complex and identification of its components by mass spectrometry. *EMBO J 19*, 6569-6581.

Akoulitchev, S., Makela, T.P., Weinberg, R.A., and Reinberg, D. (1995). Requirement for TFIIH kinase activity in transcription by RNA polymerase II. *Nature 377*, 557-560.

Alepuz, P.M., de Nadal, E., Zapater, M., Ammerer, G., and Posas, F. (2003). Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. *EMBO J 22*, 2433-2442.

Allocco, D.J., Kohane, I.S., and Butte, A.J. (2004). Quantifying the relationship between co-expression, co-regulation and gene function. *BMC Bioinformatics 5*, 18.

Altafaj, X., Dierssen, M., Baamonde, C., Marti, E., Visa, J., Guimera, J., Oset, M., Gonzalez, J.R., Florez, J., Fillat, C.*, et al.* (2001). Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome. *Hum Mol Genet 10*, 1915-1923.

Alvarez, M., Altafaj, X., Aranda, S., and de la Luna, S. (2007). DYRK1A autophosphorylation on serine residue 520 modulates its kinase activity via 14-3-3 binding. *Mol Biol Cell 18*, 1167-1178.

Alvarez, M., Estivill, X., and de la Luna, S. (2003). DYRK1A accumulates in splicing speckles through a novel targeting signal and induces speckle disassembly. *J Cell Sci 116*, 3099-3107.

Aranda, S., Alvarez, M., Turro, S., Laguna, A., and de la Luna, S. (2008). Sprouty2-mediated inhibition of fibroblast growth factor signaling is modulated by the protein kinase DYRK1A. *Mol Cell Biol 28*, 5899-5911.

Aranda, S., Laguna, A., and de la Luna, S. (2011). DYRK family of protein kinases: evolutionary relationships, biochemical properties, and functional roles. *Faseb J 25*, 449-462.

Arron, J.R., Winslow, M.M., Polleri, A., Chang, C.P., Wu, H., Gao, X., Neilson, J.R., Chen, L., Heit, J.J., Kim, S.K.*, et al.* (2006). NFAT dysregulation by increased dosage of DSCR1 and DYRK1A on chromosome 21. *Nature 441*, 595-600.

Baek, K.H., Zaslavsky, A., Lynch, R.C., Britt, C., Okada, Y., Siarey, R.J., Lensch, M.W., Park, I.H., Yoon, S.S., Minami, T.*, et al.* (2009). Down's syndrome suppression of tumour growth and the role of the calcineurin inhibitor DSCR1. *Nature 459*, 1126-1130.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res 37*, W202-208.

Bain, J., Plater, L., Elliott, M., Shpiro, N., Hastie, C.J., McLauchlan, H., Klevernic, I., Arthur, J.S., Alessi, D.R., and Cohen, P. (2007). The selectivity of protein kinase inhibitors: a further update. *Biochem J 408*, 297-315.

Barak, O., Lazzaro, M.A., Lane, W.S., Speicher, D.W., Picketts, D.J., and Shiekhattar, R. (2003). Isolation of human NURF: a regulator of Engrailed gene expression. *EMBO J 22*, 6089-6100.

Barski, A., Chepelev, I., Liko, D., Cuddapah, S., Fleming, A.B., Birch, J., Cui, K., White, R.J., and Zhao, K. (2010). Pol II and its associated epigenetic marks are present at Pol III-transcribed noncoding RNA genes. *Nat Struct Mol Biol 17*, 629-634.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell 129*, 823-837.

Becker, W., and Joost, H.G. (1999). Structural and functional characteristics of Dyrk, a novel subfamily of protein kinases with dual specificity. *Prog Nucleic Acid Res Mol Biol 62*, 1-17.

Becker, W., Weber, Y., Wetzel, K., Eirmbter, K., Tejedor, F.J., and Joost, H.G. (1998). Sequence characteristics, subcellular localization, and substrate specificity of DYRK-related kinases, a novel family of dual specificity protein kinases. *J Biol Chem 273*, 25893-25902.

Benkoussa, M., Brand, C., Delmotte, M.H., Formstecher, P., and Lefebvre, P. (2002). Retinoic acid receptors inhibit AP1 activation by regulating extracellular signal-regulated kinase and CBP recruitment to an AP1-responsive promoter. *Mol Cell Biol 22*, 4522-4534.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci U S A 99*, 8695-8700.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R*., et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell 120*, 169-181.

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. *Genome Biol 5*, R62.

Branchi, I., Bichler, Z., Minghetti, L., Delabar, J.M., Malchiodi-Albedi, F., Gonzalez, M.C., Chettouh, Z., Nicolini, A., Chabert, C., Smith, D.J*., et al.* (2004). Transgenic mouse in vivo library of human Down syndrome critical region 1: association between DYRK1A overexpression, brain development abnormalities, and cell cycle protein alteration. *J Neuropathol Exp Neurol 63*, 429-440.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annu Rev Biochem 50*, 349-383.

Bruna, A., Nicolas, M., Munoz, A., Kyriakis, J.M., and Caelles, C. (2003). Glucocorticoid receptor-JNK interaction mediates inhibition of the JNK pathway by glucocorticoids. *EMBO J 22*, 6035-6044.

Buratti, E., De Conti, L., Stuani, C., Romano, M., Baralle, M., and Baralle, F. (2010). Nuclear factor TDP-43 can affect selected microRNA levels. *FEBS J 277*, 2268-2281.

Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev 10*, 711-724.

Cai, R., Kwon, P., Yan-Neale, Y., Sambuccetti, L., Fischer, D., and Cohen, D. (2001). Mammalian histone deacetylase 1 protein is posttranslationally modified by phosphorylation. *Biochem Biophys Res Commun 283*, 445-453.

Campos, E.I., and Reinberg, D. (2009). Histones: annotating chromatin. *Annu Rev Genet 43*, 559-599.

Chambon, P. (1975). Eukaryotic nuclear RNA polymerases. *Annu Rev Biochem 44*, 613-638.

Chan, S.P., and Cheng, S.C. (2005). The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J Biol Chem 280*, 31190-31199.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E., and Eick, D. (2007). Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science 318*, 1780-1782.

Chen, Z., and Manley, J.L. (2003). Core promoter elements and TAFs contribute to the diversity of transcriptional activation in vertebrates. *Mol Cell Biol 23*, 7350-7362.

Christmann, J.L., and Dahmus, M.E. (1981). Monoclonal antibody specific for calf thymus RNA polymerases IIO and IIA. *J Biol Chem 256*, 11798-11803.

Cole, A.R., Causeret, F., Yadirgi, G., Hastie, C.J., McLauchlan, H., McManus, E.J., Hernandez, F., Eickholt, B.J., Nikolic, M., and Sutherland, C. (2006). Distinct priming kinases contribute to differential regulation of collapsin response mediator proteins by glycogen synthase kinase-3 in vivo. *J Biol Chem 281*, 16591-16598.

Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J.L., Bell, J.C., and Duncan, P.I. (1996). The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J 15*, 265-275.

Cook, M., and Tyers, M. (2007). Size control goes global. *Curr Opin Biotechnol 18*, 341-350.

Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., and Chambon, P. (1980). Promoter sequences of eukaryotic protein-coding genes. *Science 209*, 1406-1414.

Courcet, J.B., Faivre, L., Malzac, P., Masurel-Paulet, A., Lopez, E., Callier, P., Lambert, L., Lemesle, M., Thevenon, J., Gigot, N*., et al.* (2012). The DYRK1A gene is a cause of syndromic intellectual disability with severe microcephaly and epilepsy. *J Med Genet 49*, 731-736.

D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting players and paradigms in cell-specific transcription. *Mol Cell 36*, 924-931.

da Costa Martins, P.A., Salic, K., Gladka, M.M., Armand, A.S., Leptidis, S., el Azzouzi, H., Hansen, A., Coenen-de Roo, C.J., Bierhuizen, M.F., van der Nagel, R*., et al.* (2010). MicroRNA-199b targets the nuclear kinase Dyrk1a in an auto-amplification loop promoting calcineurin/NFAT signalling. *Nat Cell Biol 12*, 1220-1227.

Daniel, J.M., Spring, C.M., Crawford, H.C., Reynolds, A.B., and Baig, A. (2002). The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res 30*, 2911-2919.

de Graaf, K., Czajkowska, H., Rottmann, S., Packman, L.C., Lilischkis, R., Luscher, B., and Becker, W. (2006). The protein kinase DYRK1A phosphorylates the splicing factor SF3b1/SAP155 at Thr434, a novel in vivo phosphorylation site. *BMC Biochem 7*, 7.

de Graaf, K., Hekerman, P., Spelten, O., Herrmann, A., Packman, L.C., Bussow, K., Muller-Newen, G., and Becker, W. (2004). Characterization of cyclin L2, a novel cyclin with an arginine/serine-rich domain: phosphorylation by DYRK1A and colocalization with splicing factors. *J Biol Chem 279*, 4612-4624.

de la Luna, S., Allen, K.E., Mason, S.L., and La Thangue, N.B. (1999). Integration of a growth-suppressing BTB/POZ domain protein with the DP component of the E2F transcription factor. *EMBO J 18*, 212-228.

de la Serna, I.L., Ohkawa, Y., Berkes, C.A., Bergstrom, D.A., Dacwag, C.S., Tapscott, S.J., and Imbalzano, A.N. (2005). MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex. *Mol Cell Biol 25*, 3997-4009.

De Nadal, E., Zapater, M., Alepuz, P.M., Sumoy, L., Mas, G., and Posas, F. (2004). The MAPK Hog1 recruits Rpd3 histone deacetylase to activate osmoresponsive genes. *Nature 427*, 370-374.

Demange, L., Abdellah, F.N., Lozach, O., Ferandin, Y., Gresh, N., Meijer, L., and Galons, H. (2013). Potent inhibitors of CDK5 derived from roscovitine: synthesis, biological evaluation and molecular modelling. *Bioorg Med Chem Lett 23*, 125-131.

Deng, X., Ewton, D.Z., Mercer, S.E., and Friedman, E. (2005). Mirk/dyrk1B decreases the nuclear accumulation of class II histone deacetylases during skeletal muscle differentiation. *J Biol Chem 280*, 4894-4905.

Ding, S., Shi, J., Qian, W., Iqbal, K., Grundke-Iqbal, I., Gong, C.X., and Liu, F. (2012). Regulation of alternative splicing of tau exon 10 by 9G8 and Dyrk1A. *Neurobiol Aging 33*, 1389-1399.

Dowjat, W.K., Adayev, T., Kuchna, I., Nowicki, K., Palminiello, S., Hwang, Y.W., and Wegiel, J. (2007). Trisomy-driven overexpression of DYRK1A kinase in the brain of subjects with Down syndrome. *Neurosci Lett 413*, 77-81.

Drobic, B., Perez-Cadahia, B., Yu, J., Kung, S.K., and Davie, J.R. (2010). Promoter chromatin remodeling of immediate-early genes is mediated through H3 phosphorylation at either serine 28 or 10 by the MSK1 multi-protein complex. *Nucleic Acids Res 38*, 3196-3208.

Dyson, M.H., Thomson, S., Inagaki, M., Goto, H., Arthur, S.J., Nightingale, K., Iborra, F.J., and Mahadevan, L.C. (2005). MAP kinase-mediated phosphorylation of distinct pools of histone H3 at S10 or S28 via mitogen- and stress-activated kinase 1/2. *J Cell Sci 118*, 2247-2259.

Edmunds, J.W., and Mahadevan, L.C. (2006). Cell signaling. Protein kinases seek close encounters with active genes. *Science 313*, 449-451.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D., and Murphy, S. (2007). Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science 318*, 1777-1779.

Epstein, C.J. (2006). Down's syndrome: critical genes in a critical region. *Nature 441*, 582-583.

Fernandez-Martinez, J., Vela, E.M., Tora-Ponsioen, M., Ocana, O.H., Nieto, M.A., and Galceran, J. (2009). Attenuation of Notch signalling by the Down-syndrome-associated kinase DYRK1A. *J Cell Sci 122*, 1574-1583.

Ferreiro, I., Barragan, M., Gubern, A., Ballestar, E., Joaquin, M., and Posas, F. (2010). The p38 SAPK is recruited to chromatin via its interaction with transcription factors. *J Biol Chem 285*, 31819-31828.

Fingar, D.C., and Blenis, J. (2004). Target of rapamycin (TOR): an integrator of nutrient and growth factor signals and coordinator of cell growth and cell cycle progression. *Oncogene 23*, 3151-3171.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. *Nature 414*, 929-933.

Fotaki, V., Dierssen, M., Alcantara, S., Martinez, S., Marti, E., Casas, C., Visa, J., Soriano, E., Estivill, X., and Arbones, M.L. (2002). Dyrk1A Haploinsufficiency Affects Viability and Causes Developmental Delay and Abnormal Brain Morphology in Mice. *Mol Cell Biol 22*, 6636-6647.

Fujita, H., Torii, C., Kosaki, R., Yamaguchi, S., Kudoh, J., Hayashi, K., Takahashi, T., and Kosaki, K. (2010). Microdeletion of the Down syndrome critical region at 21q22. *Am J Med Genet A 152A*, 950-953.

Galasinski, S.C., Resing, K.A., Goodrich, J.A., and Ahn, N.G. (2002). Phosphatase inhibition leads to histone deacetylases 1 and 2 phosphorylation and disruption of corepressor interactions. *J Biol Chem 277*, 19618-19626.

Galceran, J., de Graaf, K., Tejedor, F.J., and Becker, W. (2003). The MNB/DYRK1A protein kinase: genetic and biochemical properties. *J Neural Transm Suppl*, 139-148.

Gershenzon, N.I., and Ioshikhes, I.P. (2005). Synergy of human Pol II core promoter elements revealed by statistical sequence analysis. *Bioinformatics 21*, 1295-1300.

Gockler, N., Jofre, G., Papadopoulos, C., Soppa, U., Tejedor, F.J., and Becker, W. (2009). Harmine specifically inhibits protein kinase DYRK1A and interferes with neurite formation. *FEBS J 276*, 6324-6337.

Goke, J., Chan, Y.S., Yan, J., Vingron, M., and Ng, H.H. (2013). Genome-wide Kinase-Chromatin Interactions Reveal the Regulatory Network of ERK Signaling in Human Embryonic Stem Cells. *Mol Cell*.

Goranov, A.I., and Amon, A. (2010). Growth and division--not a one-way road. *Curr Opin Cell Biol 22*, 795-800.

Graham, F.L., and van der Eb, A.J. (1973). A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology 52*, 456-467.

Grillari, J., Ajuh, P., Stadler, G., Loscher, M., Voglauer, R., Ernst, W., Chusainow, J., Eisenhaber, F., Pokar, M., Fortschegger, K.*, et al.* (2005). SNEV is an evolutionarily conserved splicing factor whose oligomerization is necessary for spliceosome assembly. *Nucleic Acids Res 33*, 6868-6883.

Grote, M., Wolf, E., Will, C.L., Lemm, I., Agafonov, D.E., Schomburg, A., Fischle, W., Urlaub, H., and Luhrmann, R. (2010). Molecular architecture of the human Prp19/CDC5L complex. *Mol Cell Biol 30*, 2105-2119.

Guedj, F., Sebrie, C., Rivals, I., Ledru, A., Paly, E., Bizot, J.C., Smith, D., Rubin, E., Gillet, B., Arbones, M.*, et al.* (2009). Green tea polyphenols rescue of brain defects induced by overexpression of DYRK1A. *PLoS One 4*, e4606.

Gui, J.F., Lane, W.S., and Fu, X.D. (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature 369*, 678-682.

Guimera, J., Casas, C., Estivill, X., and Pritchard, M. (1999). Human minibrain homologue (MNBH/DYRK1): characterization, alternative splicing, differential tissue expression, and overexpression in Down syndrome. *Genomics 57*, 407-418.

Guimera, J., Casas, C., Pucharcos, C., Solans, A., Domenech, A., Planas, A.M., Ashley, J., Lovett, M., Estivill, X., and Pritchard, M.A. (1996). A human homologue of Drosophila minibrain (MNB) is expressed in the neuronal regions affected in Down syndrome and maps to the critical region. *Hum Mol Genet 5*, 1305-1310.

Guimera, J., Pucharcos, C., Domenech, A., Casas, C., Solans, A., Gallardo, T., Ashley, J., Lovett, M., Estivill, X., and Pritchard, M. (1997). Cosmid contig and transcriptional map of three regions of human chromosome 21q22: identification of 37 novel transcripts by direct selection. *Genomics 45*, 59-67.

Guo, X., Williams, J.G., Schug, T.T., and Li, X. (2010). DYRK1A and DYRK3 promote cell survival through phosphorylation and activation of SIRT1. *J Biol Chem 285*, 13223-13232.

Gwack, Y., Sharma, S., Nardone, J., Tanasa, B., Iuga, A., Srikanth, S., Okamura, H., Bolton, D., Feske, S., Hogan, P.G.*, et al.* (2006). A genome-wide Drosophila RNAi screen identifies DYRK-family kinases as regulators of NFAT. *Nature 441*, 646-650.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol 11*, 394-403.

Hall, D.B., and Struhl, K. (2002). The VP16 activation domain interacts with multiple transcriptional components as determined by protein-protein cross-linking in vivo. *J Biol Chem 277*, 46043-46050.

Hammerle, B., Carnicero, A., Elizalde, C., Ceron, J., Martinez, S., and Tejedor, F.J. (2003). Expression patterns and subcellular localization of the Down syndrome candidate protein MNB/DYRK1A suggest a role in late neuronal differentiation. *Eur J Neurosci 17*, 2277-2286.

Hammerle, B., Vera-Samper, E., Speicher, S., Arencibia, R., Martinez, S., and Tejedor, F.J. (2002). Mnb/Dyrk1A is transiently expressed and asymmetrically segregated in neural progenitor cells at the transition to neurogenic divisions. *Dev Biol 246*, 259-273.

Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev 62*, 465-503.

Han, J., Xiong, J., Wang, D., and Fu, X.D. (2011). Pre-mRNA splicing: where and when in the nucleus. *Trends Cell Biol 21*, 336-343.

Haun, R.S., Moss, J., and Vaughan, M. (1993). Characterization of the human ADP-ribosylation factor 3 promoter. Transcriptional regulation of a TATA-less promoter. *J Biol Chem 268*, 8793-8800.

Havugimana, P.C., Hart, G.T., Nepusz, T., Yang, H., Turinsky, A.L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S.*, et al.* (2012). A census of human soluble protein complexes. *Cell 150*, 1068-1081.

Hazzalin, C.A., and Mahadevan, L.C. (2002). MAPK-regulated transcription: a continuously variable gene switch? *Nat Rev Mol Cell Biol 3*, 30-40.

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

Himpel, S., Panzer, P., Eirmbter, K., Czajkowska, H., Sayed, M., Packman, L.C., Blundell, T., Kentrup, H., Grotzinger, J., Joost, H.G.*, et al.* (2001). Identification of the autophosphorylation sites and characterization of their effects in the protein kinase DYRK1A. *Biochem J 359*, 497-505.

Himpel, S., Tegge, W., Frank, R., Leder, S., Joost, H.G., and Becker, W. (2000). Specificity determinants of substrate recognition by the protein kinase DYRK1A. *J Biol Chem 275*, 2431-2438.

Hirose, Y., and Ohkuma, Y. (2007). Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *J Biochem 141*, 601-608.

Hoskins, A.A., and Moore, M.J. (2012). The spliceosome: a flexible, reversible macromolecular machine. *Trends Biochem Sci 37*, 179-188.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev 26*, 2119-2137.

Hu, S., Xie, Z., Onishi, A., Yu, X., Jiang, L., Lin, J., Rho, H.S., Woodard, C., Wang, H., Jeong, J.S.*, et al.* (2009). Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell 139*, 610-622.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc 4*, 44-57.

Impey, S., McCorkle, S.R., Cha-Molstad, H., Dwyer, J.M., Yochum, G.S., Boss, J.M., McWeeney, S., Dunn, J.J., Mandel, G., and Goodman, R.H. (2004). Defining the

CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell 119*, 1041-1054.

Ioshikhes, I.P., and Zhang, M.Q. (2000). Large-scale human promoter mapping using CpG islands. *Nat Genet 26*, 61-63.

Irimia, M., and Blencowe, B.J. (2012). Alternative splicing: decoding an expansive regulatory layer. *Curr Opin Cell Biol 24*, 323-332.

Jacob, J., Cabarcas, S., Veras, I., Zaveri, N., and Schramm, L. (2007). The green tea component EGCG inhibits RNA polymerase III transcription. *Biochem Biophys Res Commun 360*, 778-783.

Jorgensen, P., and Tyers, M. (2004). How cells coordinate growth and division. *Curr Biol 14*, R1014-1027.

Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev Biol 339*, 225-229.

Kantidakis, T., Ramsbottom, B.A., Birch, J.L., Dowding, S.N., and White, R.J. (2010). mTOR associates with TFIIIC, is found at tRNA and 5S rRNA genes, and targets their repressor Maf1. *Proc Natl Acad Sci U S A 107*, 11823-11828.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res 32*, D493-496.

Kasler, H.G., Victoria, J., Duramad, O., and Winoto, A. (2000). ERK5 is a novel type of mitogen-activated protein kinase containing a transcriptional activation domain. *Mol Cell Biol 20*, 8382-8389.

Kim, D., Won, J., Shin, D.W., Kang, J., Kim, Y.J., Choi, S.Y., Hwang, M.K., Jeong, B.W., Kim, G.S., Joe, C.O., *et al.* (2004). Regulation of Dyrk1A kinase activity by 14-3-3. *Biochem Biophys Res Commun 323*, 499-504.

Kim, E., Du, L., Bregman, D.B., and Warren, S.L. (1997). Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. *J Cell Biol 136*, 19-28.

Kim, K.Y., and Levin, D.E. (2011). Mpk1 MAPK association with the Paf1 complex blocks Sen1-mediated premature transcription termination. *Cell 144*, 745-756.

Kim, M.Y., Jeong, B.C., Lee, J.H., Kee, H.J., Kook, H., Kim, N.S., Kim, Y.H., Kim, J.K., Ahn, K.Y., and Kim, K.K. (2006). A repressor complex, AP4 transcription factor and geminin, negatively regulates expression of target genes in nonneuronal cells. *Proc Natl Acad Sci U S A 103*, 13074-13079.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. *Nature 436*, 876-880.

Kimura, R., Kamino, K., Yamamoto, M., Nuripa, A., Kida, T., Kazui, H., Hashimoto, R., Tanaka, T., Kudo, T., Yamagata, H., *et al.* (2007). The DYRK1A gene, encoded in chromosome 21 Down syndrome critical region, bridges between beta-amyloid production and tau phosphorylation in Alzheimer disease. *Hum Mol Genet 16*, 15-23.

Kinstrie, R., Luebbering, N., Miranda-Saavedra, D., Sibbet, G., Han, J., Lochhead, P.A., and Cleghon, V. (2010). Characterization of a domain that transiently converts class 2 DYRKs into intramolecular tyrosine kinases. *Sci Signal 3*, ra16.

Knuppel, R., Dietze, P., Lehnberg, W., Frech, K., and Wingender, E. (1994). TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. *J Comput Biol 1*, 191-198.

Kojima, T., Zama, T., Wada, K., Onogi, H., and Hagiwara, M. (2001). Cloning of human PRP4 reveals interaction with Clk1. *J Biol Chem 276*, 32247-32256.

Kuhn, C., Frank, D., Will, R., Jaschinski, C., Frauen, R., Katus, H.A., and Frey, N. (2009). DYRK1A is a novel negative regulator of cardiomyocyte hypertrophy. *J Biol Chem 284*, 17320-17327.

Kurabayashi, N., Hirota, T., Sakai, M., Sanada, K., and Fukada, Y. (2010). DYRK1A and glycogen synthase kinase 3beta, a dual-kinase mechanism directing proteasomal degradation of CRY2 for circadian timekeeping. *Mol Cell Biol 30*, 1757-1768.

Kuroyanagi, N., Onogi, H., Wakabayashi, T., and Hagiwara, M. (1998). Novel SR-protein-specific kinase, SRPK2, disassembles nuclear speckles. *Biochem Biophys Res Commun 242*, 357-364.

Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev 16*, 2893-2905.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., and Ebright, R.H. (1998). New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev 12*, 34-44.

Laguna, A., Aranda, S., Barallobre, M.J., Barhoum, R., Fernandez, E., Fotaki, V., Delabar, J.M., de la Luna, S., de la Villa, P., and Arbones, M.L. (2008). The protein kinase DYRK1A regulates caspase-9-mediated apoptosis during retina development. *Dev Cell 15*, 841-853.

Lamond, A.I., and Spector, D.L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol 4*, 605-612.

Landolin, J.M., Johnson, D.S., Trinklein, N.D., Aldred, S.F., Medina, C., Shulha, H., Weng, Z., and Myers, R.M. (2010). Sequence features that drive human promoter function and tissue specificity. *Genome Res 20*, 890-898.

Lawrence, M.C., McGlynn, K., Shao, C., Duan, L., Naziruddin, B., Levy, M.F., and Cobb, M.H. (2008). Chromatin-bound mitogen-activated protein kinases transmit dynamic signals in transcription complexes in beta-cells. *Proc Natl Acad Sci U S A 105*, 13315-13320.

Laybourn, P.J., and Dahmus, M.E. (1989). Transcription-dependent structural changes in the C-terminal domain of mammalian RNA polymerase subunit IIa/o. *J Biol Chem 264*, 6693-6698.

Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet 36*, 900-905.

Lee, D.H., Gershenzon, N., Gupta, M., Ioshikhes, I.P., Reinberg, D., and Lewis, B.A. (2005). Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol Cell Biol 25*, 9674-9686.

Lee, J., Moir, R.D., McIntosh, K.B., and Willis, I.M. (2012). TOR signaling regulates ribosome and tRNA synthesis via LAMMER/Clk and GSK-3 family kinases. *Mol Cell 45*, 836-843.

Lee, K., Deng, X., and Friedman, E. (2000). Mirk protein kinase is a mitogen-activated protein kinase substrate that mediates survival of colon cancer cells. *Cancer Res 60*, 3631-3637.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annu Rev Genet 34*, 77-137.

Lee, Y., Ha, J., Kim, H.J., Kim, Y.S., Chang, E.J., Song, W.J., and Kim, H.H. (2009). Negative feedback Inhibition of NFATc1 by DYRK1A regulates bone homeostasis. *J Biol Chem 284*, 33343-33351.

Lefstin, J.A., and Yamamoto, K.R. (1998). Allosteric effects of DNA on transcriptional regulators. *Nature 392*, 885-888.

Lemon, B., and Tjian, R. (2000). Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev 14*, 2551-2569.

Lewis, B.A., Kim, T.K., and Orkin, S.H. (2000). A downstream element in the human beta-globin promoter: evidence of extended sequence-specific transcription factor IID contacts. *Proc Natl Acad Sci U S A 97*, 7172-7177.

Li, D., Jackson, R.A., Yusoff, P., and Guy, G.R. (2010). Direct association of Sprouty-related protein with an EVH1 domain (SPRED) 1 or SPRED2 with DYRK1A modifies substrate/kinase interactions. *J Biol Chem 285*, 35374-35385.

Li, K., Zhao, S., Karur, V., and Wojchowski, D.M. (2002). DYRK3 activation, engagement of protein kinase A/cAMP response element-binding protein, and modulation of progenitor cell survival. *J Biol Chem 277*, 47052-47060.

Lim, C.Y., Santoso, B., Boulay, T., Dong, E., Ohler, U., and Kadonaga, J.T. (2004). The MTE, a new core promoter element for transcription by RNA polymerase II. *Genes Dev 18*, 1606-1617.

Lin, Y.S., Ha, I., Maldonado, E., Reinberg, D., and Green, M.R. (1991). Binding of general transcription factor TFIIB to an acidic activating region. *Nature 353*, 569-571.

Litovchick, L., Florens, L.A., Swanson, S.K., Washburn, M.P., and DeCaprio, J.A. (2011). DYRK1A protein kinase promotes quiescence and senescence through DREAM complex assembly. *Genes Dev 25*, 801-813.

Lochhead, P.A., Sibbet, G., Morrice, N., and Cleghon, V. (2005). Activation-loop autophosphorylation is mediated by a novel transitional intermediate form of DYRKs. *Cell 121*, 925-936.

Lockstone, H.E., Harris, L.W., Swatton, J.E., Wayland, M.T., Holland, A.J., and Bahn, S. (2007). Gene expression profiling in the adult Down syndrome brain. *Genomics 90*, 647-660.

Lu, M., Zheng, L., Han, B., Wang, L., Wang, P., Liu, H., and Sun, X. (2011). REST regulates DYRK1A transcription in a negative feedback loop. *J Biol Chem 286*, 10755-10763.

Lykke-Andersen, S., Tomecki, R., Jensen, T.H., and Dziembowski, A. (2011). The eukaryotic RNA exosome: same scaffold but variable catalytic subunits. *RNA Biol 8*, 61-66.

Madak-Erdogan, Z., Lupien, M., Stossi, F., Brown, M., and Katzenellenbogen, B.S. (2011). Genomic collaboration of estrogen receptor alpha and extracellular signal-regulated kinase 2 in regulating gene and proliferation programs. *Mol Cell Biol 31*, 226-236.

Maenz, B., Hekerman, P., Vela, E.M., Galceran, J., and Becker, W. (2008). Characterization of the human DYRK1A promoter and its regulation by the transcription factor E2F1. *BMC Mol Biol 9*, 30.

Majello, B., Napolitano, G., Giordano, A., and Lania, L. (1999). Transcriptional regulation by targeted recruitment of cyclin-dependent CDK9 kinase in vivo. *Oncogene 18*, 4598-4605.

Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J 23*, 2381-2391.

Maldonado, E., Shiekhattar, R., Sheldon, M., Cho, H., Drapkin, R., Rickert, P., Lees, E., Anderson, C.W., Linn, S., and Reinberg, D. (1996). A human RNA polymerase II complex associated with SRB and DNA-repair proteins. *Nature 381*, 86-89.

Malik, S., and Roeder, R.G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci 30*, 256-263.

Malinge, S., Bliss-Moreau, M., Kirsammer, G., Diebold, L., Chlon, T., Gurbuxani, S., and Crispino, J.D. (2012). Increased dosage of the chromosome 21 ortholog Dyrk1a promotes megakaryoblastic leukemia in a murine model of Down syndrome. *J Clin Invest 122*, 948-962.

Mandel, C.R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci 65*, 1099-1122.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. *Nature 416*, 499-506.

Manning, G., Whyte, D.B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002). The protein kinase complement of the human genome. *Science 298*, 1912-1934.

Mao, J., Maye, P., Kogerman, P., Tejedor, F.J., Toftgard, R., Xie, W., Wu, G., and Wu, D. (2002). Regulation of Gli1 transcriptional activity in the nucleus by Dyrk1. *J Biol Chem 277*, 35156-35161.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. *J Biol Chem 271*, 27176-27183.

Marshall, N.F., and Price, D.H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol 12*, 2078-2090.

Marti, E., Altafaj, X., Dierssen, M., de la Luna, S., Fotaki, V., Alvarez, M., Perez-Riba, M., Ferrer, I., and Estivill, X. (2003). Dyrk1A expression pattern supports specific roles of this kinase in the adult central nervous system. *Brain Res 964*, 250-263.

Maston, G.A., Evans, S.K., and Green, M.R. (2006). Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet 7*, 29-59.

McDonel, P., Costello, I., and Hendrich, B. (2009). Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. *Int J Biochem Cell Biol 41*, 108-116.

Meyer, L.R., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Kuhn, R.M., Wong, M., Sloan, C.A., Rosenbloom, K.R., Roe, G., Rhead, B.*, et al.* (2013). The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res 41*, D64-69.

Mikula, M., Gaj, P., Dzwonek, K., Rubel, T., Karczmarski, J., Paziewska, A., Dzwonek, A., Bragoszewski, P., Dadlez, M., and Ostrowski, J. (2010). Comprehensive analysis of the palindromic motif TCTCGCGAGA: a regulatory element of the HNRNPK promoter. *DNA Res 17*, 245-260.

Mochizuki, K., Nishiyama, A., Jang, M.K., Dey, A., Ghosh, A., Tamura, T., Natsume, H., Yao, H., and Ozato, K. (2008). The bromodomain protein Brd4 stimulates G1 gene transcription and promotes progression to S phase. *J Biol Chem 283*, 9040-9048.

Moller, R.S., Kubart, S., Hoeltzenbein, M., Heye, B., Vogel, I., Hansen, C.P., Menzel, C., Ullmann, R., Tommerup, N., Ropers, H.H.*, et al.* (2008). Truncation of the Down syndrome candidate gene DYRK1A in two unrelated patients with microcephaly. *Am J Hum Genet 82*, 1165-1170.

Montanuy, I., Torremocha, R., Hernandez-Munain, C., and Sune, C. (2008). Promoter influences transcription elongation: TATA-box element mediates the assembly of processive transcription complexes responsive to cyclin-dependent kinase 9. *J Biol Chem 283*, 7368-7378.

Moqtaderi, Z., Wang, J., Raha, D., White, R.J., Snyder, M., Weng, Z., and Struhl, K. (2010). Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nat Struct Mol Biol 17*, 635-640.

Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A., and Berezney, R. (1996). A hyperphosphorylated form of the large subunit of

RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci U S A 93*, 8253-8257.

Moseley, J.B., Mayeux, A., Paoletti, A., and Nurse, P. (2009). A spatial gradient coordinates cell size and mitotic entry in fission yeast. *Nature 459*, 857-860.

Muller, F., and Tora, L. (2004). The multicoloured world of promoter recognition complexes. *EMBO J 23*, 2-8.

Muller, F., Zaucker, A., and Tora, L. (2010). Developmental regulation of transcription initiation: more than just changing the actors. *Curr Opin Genet Dev 20*, 533-540.

Murakami, N., Bolton, D., and Hwang, Y.W. (2009). Dyrk1A binds to multiple endocytic proteins required for formation of clathrin-coated vesicles. *Biochemistry 48*, 9297-9305.

Myrianthopoulos, V., Kritsanida, M., Gaboriaud-Kolar, N., Magiatis, P., Ferandin, Y., Durieu, E., Lozach, O., Cappel, D., Soundararajan, M., Filippakopoulos, P.*, et al.* (2013). Novel Inverse Binding Mode of Indirubin Derivatives Yields Improved Selectivity for DYRK Kinases. *ACS Med Chem Lett 4*, 22-26.

Nguyen, V.T., Kiss, T., Michels, A.A., and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature 414*, 322-325.

Ni, Z., Schwartz, B.E., Werner, J., Suarez, J.R., and Lis, J.T. (2004). Coordination of transcription, RNA processing, and surveillance by P-TEFb kinase on heat shock genes. *Mol Cell 13*, 55-65.

O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K.*, et al.* (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science 338*, 1619-1622.

Ogawa, Y., Nonaka, Y., Goto, T., Ohnishi, E., Hiramatsu, T., Kii, I., Yoshida, M., Ikura, T., Onogi, H., Shibuya, H.*, et al.* (2010). Development of a novel selective inhibitor of the Down syndrome-related kinase Dyrk1A. *Nat Commun 1*, 86.

Oler, A.J., Alla, R.K., Roberts, D.N., Wong, A., Hollenhorst, P.C., Chandler, K.J., Cassiday, P.A., Nelson, C.A., Hagedorn, C.H., Graves, B.J.*, et al.* (2010). Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nat Struct Mol Biol 17*, 620-628.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev 10*, 2657-2683.

Palaniswamy, S.K., Jin, V.X., Sun, H., and Davuluri, R.V. (2005). OMGProm: a database of orthologous mammalian gene promoters. *Bioinformatics 21*, 835-836.

Pandit, S., Wang, D., and Fu, X.D. (2008). Functional integration of transcriptional and RNA processing machineries. *Curr Opin Cell Biol 20*, 260-265.

Papadopoulos, C., Arato, K., Lilienthal, E., Zerweck, J., Schutkowski, M., Chatain, N., Muller-Newen, G., Becker, W., and de la Luna, S. (2011). Splice variants of the dual specificity tyrosine phosphorylation-regulated kinase 4 (DYRK4) differ in their subcellular localization and catalytic activity. *J Biol Chem 286*, 5494-5505.

Park, J., Oh, Y., Yoo, L., Jung, M.S., Song, W.J., Lee, S.H., Seo, H., and Chung, K.C. (2010). Dyrk1A phosphorylates p53 and inhibits proliferation of embryonic neuronal cells. *J Biol Chem 285*, 31895-31906.

Peng, J., Zhu, Y., Milton, J.T., and Price, D.H. (1998). Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev 12*, 755-762.

Phatnani, H.P., and Greenleaf, A.L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev 20*, 2922-2936.

Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res 21*, 447-455.

Platts, A.E., Dix, D.J., Chemes, H.E., Thompson, K.E., Goodrich, R., Rockett, J.C., Rawe, V.Y., Quintana, S., Diamond, M.P., Strader, L.F.*, et al.* (2007). Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Hum Mol Genet 16*, 763-773.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E.*, et al.* (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell 122*, 517-527.

Pokholok, D.K., Zeitlinger, J., Hannett, N.M., Reynolds, D.B., and Young, R.A. (2006). Activated signal transduction kinases frequently occupy target genes. *Science 313*, 533-536.

Pozo, N., Zahonero, C., Fernandez, P., Linares, J.M., Ayuso, A., Hagiwara, M., Perez, A., Ricoy, J.R., Hernandez-Lain, A., Sepulveda, J.M.*, et al.* (2013). Inhibition of DYRK1A destabilizes EGFR and reduces EGFR-dependent glioblastoma growth. *J Clin Invest 123*, 2475-2487.

Prelich, G. (2002). RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. *Eukaryot Cell 1*, 153-162.

Price, D.H. (2000). P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol Cell Biol 20*, 2629-2634.

Proft, M., and Struhl, K. (2002). Hog1 kinase converts the Sko1-Cyc8-Tup1 repressor complex into an activator that recruits SAGA and SWI/SNF in response to osmotic stress. *Mol Cell 9*, 1307-1317.

Proudfoot, N.J. (2011). Ending the message: poly(A) signals then and now. *Genes Dev 25*, 1770-1782.

Proudfoot, N.J., Furger, A., and Dye, M.J. (2002). Integrating mRNA processing with transcription. *Cell 108*, 501-512.

Ptashne, M., and Gann, A. (1997). Transcriptional activation by recruitment. *Nature 386*, 569-577.

Qian, W., Liang, H., Shi, J., Jin, N., Grundke-Iqbal, I., Iqbal, K., Gong, C.X., and Liu, F. (2011). Regulation of the alternative splicing of tau exon 10 by SC35 and Dyrk1A. *Nucleic Acids Res 39*, 6161-6171.

Raaf, L., Noll, C., Cherifi, M., Benazzoug, Y., Delabar, J.M., and Janel, N. (2010). Hyperhomocysteinemia-induced Dyrk1a downregulation results in cardiomyocyte hypertrophy in rats. *Int J Cardiol 145*, 306-307.

Raghav, S.K., Waszak, S.M., Krier, I., Gubelmann, C., Isakova, A., Mikkelsen, T.S., and Deplancke, B. (2012). Integrative genomics identifies the corepressor SMRT as a gatekeeper of adipogenesis through the transcription factors C/EBPbeta and KAISO. *Mol Cell 46*, 335-350.

Remenyi, A., Scholer, H.R., and Wilmanns, M. (2004). Combinatorial control of gene expression. *Nat Struct Mol Biol 11*, 812-815.

Roeder, R.G., and Rutter, W.J. (1969). Multiple forms of DNA-dependent RNA polymerase in eukaryotic organisms. *Nature 224*, 234-237.

Roeder, R.G., and Rutter, W.J. (1970). Specific nucleolar and nucleoplasmic RNA polymerases. *Proc Natl Acad Sci U S A 65*, 675-682.

Roepcke, S., Zhi, D., Vingron, M., and Arndt, P.F. (2006). Identification of highly specific localized sequence motifs in human ribosomal protein gene promoters. *Gene 365*, 48-56.

Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. *Proc Natl Acad Sci U S A 103*, 15782-15787.

Roh, T.Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev 19*, 542-552.

Russell, J., and Zomerdijk, J.C. (2005). RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem Sci 30*, 87-96.

Russell, R.C., Fang, C., and Guan, K.L. (2011). An emerging role for TOR signaling in mammalian tissue and stem cell physiology. *Development 138*, 3343-3356.

Ryoo, S.R., Cho, H.J., Lee, H.W., Jeong, H.K., Radnaabazar, C., Kim, Y.S., Kim, M.J., Son, M.Y., Seo, H., Chung, S.H., *et al.* (2008). Dual-specificity tyrosine(Y)-phosphorylation regulated kinase 1A-mediated phosphorylation of amyloid precursor protein: evidence for a functional link between Down syndrome and Alzheimer's disease. *J Neurochem 104*, 1333-1344.

Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. *Nature 335*, 563-564.

Saitoh, N., Spahr, C.S., Patterson, S.D., Bubulya, P., Neuwald, A.F., and Spector, D.L. (2004). Proteomic analysis of interchromatin granule clusters. *Mol Biol Cell 15*, 3876-3890.

Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res 32*, D91-94.

Saul, V.V., de la Vega, L., Milanovic, M., Kruger, M., Braun, T., Fritz-Wolf, K., Becker, K., and Schmitz, M.L. (2013). HIPK2 kinase activity depends on cis-autophosphorylation of its activation loop. *J Mol Cell Biol 5*, 27-38.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol 7*, 557-567.

Scales, T.M., Lin, S., Kraus, M., Goold, R.G., and Gordon-Weeks, P.R. (2009). Nonprimed and DYRK1A-primed GSK3 beta-phosphorylation sites on MAP1B regulate microtubule dynamics in growing axons. *J Cell Sci 122*, 2424-2435.

Schwartz, L.B., and Roeder, R.G. (1975). Purification and subunit structure of deoxyribonucleic acid-dependent ribonucleic acid polymerase II from the mouse plasmacytoma, MOPC 315. *J Biol Chem 250*, 3221-3228.

Seifert, A., Allan, L.A., and Clarke, P.R. (2008). DYRK1A phosphorylates caspase 9 at an inhibitory site and is potently inhibited in human cells by harmine. *Febs J 275*, 6268-6280.

Seifert, A., and Clarke, P.R. (2009). p38alpha- and DYRK1A-dependent phosphorylation of caspase-9 at an inhibitory site in response to hyperosmotic stress. *Cell Signal 21*, 1626-1633.

Sengupta, N., and Seto, E. (2004). Regulation of histone deacetylase activities. *J Cell Biochem 93*, 57-67.

Serra, C., Palacios, D., Mozzetta, C., Forcales, S.V., Morantte, I., Ripani, M., Jones, D.R., Du, K., Jhala, U.S., Simone, C., *et al.* (2007). Functional interdependence at the chromatin level between the MKK6/p38 and IGF1/PI3K/AKT pathways during muscle differentiation. *Mol Cell 28*, 200-213.

Shi, J., Zhang, T., Zhou, C., Chohan, M.O., Gu, X., Wegiel, J., Zhou, J., Hwang, Y.W., Iqbal, K., Grundke-Iqbal, I., *et al.* (2008). Increased dosage of Dyrk1A alters alternative splicing factor (ASF)-regulated alternative splicing of tau in Down syndrome. *J Biol Chem 283*, 28660-28669.

Shilatifard, A., Conaway, R.C., and Conaway, J.W. (2003). The RNA polymerase II elongation complex. *Annu Rev Biochem 72*, 693-715.

Shimada, M., Nakadai, T., Fukuda, A., and Hisatake, K. (2010). cAMP-response element-binding protein (CREB) controls MSK1-mediated phosphorylation of histone H3 at the c-fos promoter in vitro. *J Biol Chem 285*, 9390-9401.

Shimokawa, T., Tostar, U., Lauth, M., Palaniswamy, R., Kasper, M., Toftgard, R., and Zaphiropoulos, P.G. (2008). Novel human glioma-associated oncogene 1 (GLI1) splice variants reveal distinct mechanisms in the terminal transduction of the hedgehog signal. *J Biol Chem 283*, 14345-14354.

Shore, S.M., Byers, S.A., Maury, W., and Price, D.H. (2003). Identification of a novel isoform of Cdk9. *Gene 307*, 175-182.

Siepi, F., Gatti, V., Camerini, S., Crescenzi, M., and Soddu, S. (2013). HIPK2 catalytic activity and subcellular localization are regulated by activation-loop Y354 autophosphorylation. *Biochim Biophys Acta 1833*, 1443-1453.

Simone, C., Forcales, S.V., Hill, D.A., Imbalzano, A.N., Latella, L., and Puri, P.L. (2004). p38 pathway targets SWI-SNF chromatin-remodeling complex to muscle-specific loci. *Nat Genet 36*, 738-743.

Sitz, J.H., Tigges, M., Baumgartel, K., Khaspekov, L.G., and Lutz, B. (2004). Dyrk1A potentiates steroid hormone-induced transcription via the chromatin remodeling factor Arip4. *Mol Cell Biol 24*, 5821-5834.

Skurat, A.V., and Dietrich, A.D. (2004). Phosphorylation of Ser640 in muscle glycogen synthase by DYRK family protein kinases. *J Biol Chem 279*, 2490-2498.

Smale, S.T., and Kadonaga, J.T. (2003). The RNA polymerase II core promoter. *Annu Rev Biochem 72*, 449-479.

Song, W.J., Sternberg, L.R., Kasten-Sportes, C., Keuren, M.L., Chung, S.H., Slack, A.C., Miller, D.E., Glover, T.W., Chiang, P.W., Lou, L.*, et al.* (1996). Isolation of human and murine homologues of the Drosophila minibrain gene: human homologue maps to 21q22.2 in the Down syndrome "critical region". *Genomics 38*, 331-339.

Soundararajan, M., Roos, A.K., Savitsky, P., Filippakopoulos, P., Kettenbach, A.N., Olsen, J.V., Gerber, S.A., Eswaran, J., Knapp, S., and Elkins, J.M. (2013). Structures of Down Syndrome Kinases, DYRKs, Reveal Mechanisms of Kinase Activation and Substrate Recognition. *Structure*.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. *Cold Spring Harb Perspect Biol 3*.

Stewart, S.A., Dykxhoorn, D.M., Palliser, D., Mizuno, H., Yu, E.Y., An, D.S., Sabatini, D.M., Chen, I.S., Hahn, W.C., Sharp, P.A.*, et al.* (2003). Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA 9*, 493-501.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G.*, et al.* (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A 101*, 6062-6067.

Suganuma, T., and Workman, J.L. (2011). Signals and combinatorial functions of histone modifications. *Annu Rev Biochem 80*, 473-499.

Sun, J.M., Chen, H.Y., Moniwa, M., Litchfield, D.W., Seto, E., and Davie, J.R. (2002). The transcriptional repressor Sp3 is associated with CK2-phosphorylated histone deacetylase 2. *J Biol Chem 277*, 35783-35786.

Sun, M., Isaacs, G.D., Hah, N., Heldring, N., Fogarty, E.A., and Kraus, W.L. (2012). Estrogen regulates JNK1 genomic localization to control gene expression and cell growth in breast cancer cells. *Mol Endocrinol 26*, 736-747.

Sutcliffe, E.L., Bunting, K.L., He, Y.Q., Li, J., Phetsouphanh, C., Seddiki, N., Zafar, A., Hindmarsh, E.J., Parish, C.R., Kelleher, A.D.*, et al.* (2011). Chromatin-associated protein kinase C-theta regulates an inducible gene expression program and microRNAs in human T lymphocytes. *Mol Cell 41*, 704-719.

Taira, N., Mimoto, R., Kurata, M., Yamaguchi, T., Kitagawa, M., Miki, Y., and Yoshida, K. (2012). DYRK2 priming phosphorylation of c-Jun and c-Myc modulates cell cycle progression in human cancer cells. *J Clin Invest 122*, 859-872.

Tejedor, F., Zhu, X.R., Kaltenbach, E., Ackermann, A., Baumann, A., Canal, I., Heisenberg, M., Fischbach, K.F., and Pongs, O. (1995). minibrain: a new protein kinase family involved in postembryonic neurogenesis in Drosophila. *Neuron 14*, 287-301.

Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. *Crit Rev Biochem Mol Biol 41*, 105-178.

Tiwari, V.K., Stadler, M.B., Wirbelauer, C., Paro, R., Schubeler, D., and Beisel, C. (2012). A chromatin-modifying function of JNK during stem cell differentiation. *Nat Genet 44*, 94-100.

Tokusumi, Y., Ma, Y., Song, X., Jacobson, R.H., and Takada, S. (2007). The new core promoter element XCPE1 (X Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-independent RNA polymerase II transcription from TATA-less promoters. *Mol Cell Biol 27*, 1844-1858.

Trinkle-Mulcahy, L., Boulon, S., Lam, Y.W., Urcia, R., Boisvert, F.M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H.*, et al.* (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J Cell Biol 183*, 223-239.

Trinklein, N.D., Aldred, S.J., Saldanha, A.J., and Myers, R.M. (2003). Identification and functional analysis of human transcriptional promoters. *Genome Res 13*, 308-312.

Tripathi, V., Song, D.Y., Zong, X., Shevtsov, S.P., Hearn, S., Fu, X.D., Dundr, M., and Prasanth, K.V. (2012). SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol Biol Cell 23*, 3694-3706.

Tsai, S.C., and Seto, E. (2002). Regulation of histone deacetylase 2 by protein kinase CK2. *J Biol Chem 277*, 31826-31833.

Tumaneng, K., Russell, R.C., and Guan, K.L. (2012). Organ size control by Hippo and TOR pathways. *Curr Biol 22*, R368-379.

van Roy, F.M., and McCrea, P.D. (2005). A role for Kaiso-p120ctn complexes in cancer? *Nat Rev Cancer 5*, 956-964.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A.*, et al.* (2001). The sequence of the human genome. *Science 291*, 1304-1351.

Vicent, G.P., Ballare, C., Nacht, A.S., Clausell, J., Subtil-Rodriguez, A., Quiles, I., Jordan, A., and Beato, M. (2006). Induction of progesterone target genes requires activation of Erk and Msk kinases and phosphorylation of histone H3. *Mol Cell 24*, 367-381.

Vicent, G.P., Nacht, A.S., Zaurin, R., Font-Mateu, J., Soronellas, D., Le Dily, F., Reyes, D., and Beato, M. (2013). Unliganded progesterone receptor-mediated targeting of an RNA-containing repressive complex silences a subset of hormone-inducible genes. *Genes Dev 27*, 1179-1197.

Wang, H.Y., Lin, W., Dyck, J.A., Yeakley, J.M., Songyang, Z., Cantley, L.C., and Fu, X.D. (1998). SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol 140*, 737-750.

Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y.*, et al.* (2012). Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res 22*, 1798-1812.

Wegiel, J., Kuchna, I., Nowicki, K., Frackowiak, J., Dowjat, K., Silverman, W.P., Reisberg, B., DeLeon, M., Wisniewski, T., Adayev, T.*, et al.* (2004). Cell type- and brain structure-specific patterns of distribution of minibrain kinase in human brain. *Brain Res 1010*, 69-80.

Wei, P., Garber, M.E., Fang, S.M., Fischer, W.H., and Jones, K.A. (1998). A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell 92*, 451-462.

White, R.J. (2011). Transcription by RNA polymerase III: more complex than we thought. *Nat Rev Genet 12*, 459-463.

Wingender, E. (1994). Recognition of regulatory regions in genomic sequences. *J Biotechnol 35*, 273-280.

Woods, Y.L., Cohen, P., Becker, W., Jakes, R., Goedert, M., Wang, X., and Proud, C.G. (2001a). The kinase DYRK phosphorylates protein-synthesis initiation factor eIF2Bepsilon at Ser539 and the microtubule-associated protein tau at Thr212: potential role for DYRK as a glycogen synthase kinase 3-priming kinase. *Biochem J 355*, 609-615.

Woods, Y.L., Rena, G., Morrice, N., Barthel, A., Becker, W., Guo, S., Unterman, T.G., and Cohen, P. (2001b). The kinase DYRK1A phosphorylates the transcription factor FKHR at Ser329 in vitro, a novel in vivo phosphorylation site. *Biochem J 355*, 597-607.

Wyrwicz, L.S., Gaj, P., Hoffmann, M., Rychlewski, L., and Ostrowski, J. (2007). A common cis-element in promoters of protein synthesis and cell cycle genes. *Acta Biochim Pol 54*, 89-98.

Xie, X., Lu, J., Kulbokas, E.J., Golub, T.R., Mootha, V., Lindblad-Toh, K., Lander, E.S., and Kellis, M. (2005). Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature 434*, 338-345.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell 2*, 851-861.

Yabut, O., Domogauer, J., and D'Arcangelo, G. (2010). Dyrk1A overexpression inhibits proliferation and induces premature neuronal differentiation of neural progenitor cells. *J Neurosci 30*, 4004-4014.

Yamamoto, T., Shimojima, K., Nishizawa, T., Matsuo, M., Ito, M., and Imai, K. (2011). Clinical manifestations of the deletion of Down syndrome critical region including DYRK1A and KCNJ6. *Am J Med Genet A 155A*, 113-119.

Yang, Z., Zhu, Q., Luo, K., and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature 414*, 317-322.

Yik, J.H., Chen, R., Nishimura, R., Jennings, J.L., Link, A.J., and Zhou, Q. (2003). Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol Cell 12*, 971-982.

Yin, X., Jin, N., Gu, J., Shi, J., Zhou, J., Gong, C.X., Iqbal, K., Grundke-Iqbal, I., and Liu, F. (2012). Dual-specificity tyrosine phosphorylation-regulated kinase 1A (Dyrk1A) modulates serine/arginine-rich protein 55 (SRp55)-promoted Tau exon 10 inclusion. *J Biol Chem 287*, 30497-30506.

Yoon, H.G., Chan, D.W., Reynolds, A.B., Qin, J., and Wong, J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol Cell 12*, 723-734.

Yu, H., Luscombe, N.M., Qian, J., and Gerstein, M. (2003). Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet 19*, 422-427.

Yudkovsky, N., Ranish, J.A., and Hahn, S. (2000). A transcription reinitiation intermediate that is stabilized by activator. *Nature 408*, 225-229.

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M., and Corden, J.L. (1996). The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A 93*, 6975-6980.

Zhang, Y., Liao, J.M., Zeng, S.X., and Lu, H. (2011). p53 downregulates Down syndrome-associated DYRK1A through miR-1246. *EMBO Rep 12*, 811-817.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev 13*, 1924-1935.

Zhu, Y., Pe'ery, T., Peng, J., Ramanathan, Y., Marshall, N., Marshall, T., Amendt, B., Mathews, M.B., and Price, D.H. (1997). Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro. *Genes Dev 11*, 2622-2632.

Zippo, A., De Robertis, A., Serafini, R., and Oliviero, S. (2007). PIM1-dependent phosphorylation of histone H3 at serine 10 is required for MYC-dependent transcriptional activation and oncogenic transformation. *Nat Cell Biol 9*, 932-944.

Zoncu, R., Efeyan, A., and Sabatini, D.M. (2011). mTOR: from growth signal integration to cancer, diabetes and ageing. *Nat Rev Mol Cell Biol 12*, 21-35.

Zou, Y., Lim, S., Lee, K., Deng, X., and Friedman, E. (2003). Serine/threonine kinase Mirk/Dyrk1B is an inhibitor of epithelial cell migration and is negatively regulated by the Met adaptor Ran-binding protein M. *J Biol Chem 278*, 49573-49581.

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L., and Trono, D. (1997). Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo. *Nat Biotechnol 15*, 871-875.