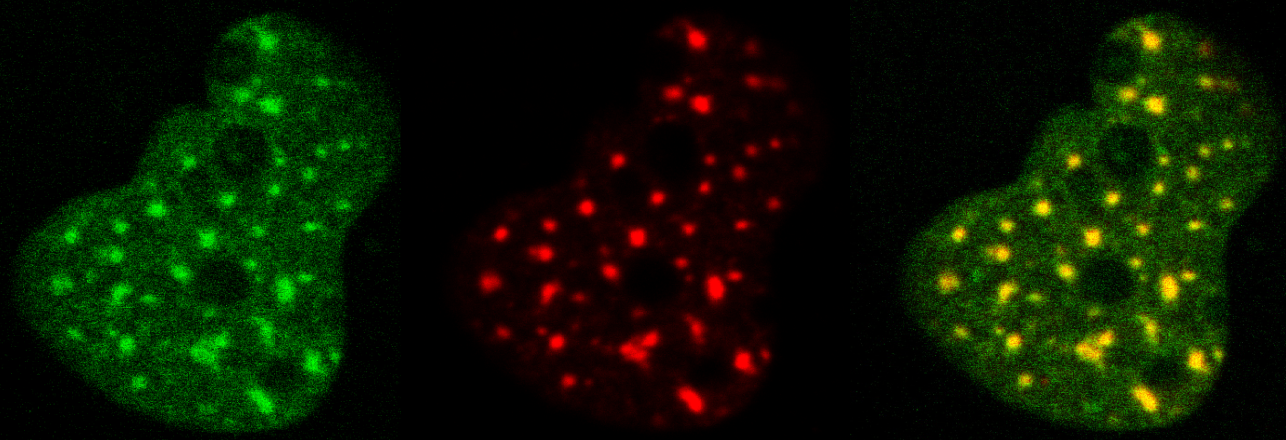
Doctoral Thesis 2008

PolyHistidine repeats and DYRK1A: from the localization to the function



Doctoral Thesis

Eulàlia Salichs Fradera
December 2008

Eulàlia Salichs Fradera

PolyHistidine repeats and DYRK1A: from the localization to the function

Doctoral Thesis Report

Eulàlia Salichs Fradera

Barcelona, December 2008

PolyHistidine repeats and DYRK1A: from the localization to the function

Memoria presentada per **Eulàlia Salichs Fradera** per optar al Grau de Doctor

Treball realitzat sota la direcció de la Dra. Susana de la Luna Gargantilla
al Centre de Regulació Genòmica (CRG)

Tesi adscrita al Departament de Ciències Experimentals i de la Salut, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra
Programa de Biomedicina, bienni 2003-2005

Susana de la Luna Gargantilla
(Directora de la tesi)

Eulàlia Salichs Fradera
(Doctoranda)

Barcelona, desembre de 2008

Eulàlia Salichs està finançada dins el Programa de Becas Predoctorales de Formación de Personal Investigador del Ministerio de Educación y Ciencia (BFU2004-01768).
Susana de la Luna és Profesora d'Investigació ICREA.

Si m'ho permeteu…

Al Joan,
poeta de la vida,
etern far en la distància.

Aquest és el raconet més personal de la tesi des d'on, amb el cor a la mà, dono les gràcies a tots aquells i aquelles que m'han ajudat a créixer i fer camí. Ahir, avui i demà.

Vull començar donant-li les gràcies a la Susana. Per la seva gran capacitat didàctica, per la seva enorme paciència amb mi i els meus cent mil dubtes, per la seva comprensió i la força per tirar (i fer-me tirar) endavant en moments fluixos. Per acceptar-me, per encomanar-me les ganes i fer-me creure que jo també puc. Per la seva il.limitada capacitat de treballar, pel seu bon criteri i, sobetot, per la seva humanitat. És un autèntic plaer (i una sort!) treballar amb tu, Susana.

El laboratori, la nostra segona casa. És tan important trobar-s'hi bé...! Per això la imprescindibilitat de tota la gent que m'ha envoltat dia rere dia durant aquests anys. El p35A i del p35B (hi ha coses que no canvien, encara que a la nostra porta ara hi posi "528"). Crec que hem acabat parlant en un dialecte propi, oi? Tantes hores dins i fora de lab amb la imprescindible Alicia, sempre allí. La Kriszti-chocokrispies, diana de tantes bromes (i dels tants problemes de mútua comprensió!). La dolça Chiara-garbancito, sempre amb una paraula bonica als llavis (aquí tienes tu Biblia, jiji!). La Sílvia o com riure amb algú taaan despistat. La Maria José (batejada MJ des del primer dia), la mami echá-pa-lante i sempre amb voluntat d'ajudar. La Uli-panxeta, pleníssima de força i caràcter. L'Elisa i el seu grandiós cor, pura passió. La Erika i el seus jerseis impossibles, provocant sempre mil i una carcajades. El "musculoco" Valentín. La Sònia (mig de Canet!) i les seves ganes. La Mariona, sempre tan dolça amb mi. I, és clar, l'Ari (Lake!) i el Sergi. Dos puntals imprescindibles al llarg d'aquests quasi sis anys, des d'aquell 7 de gener del 2003 en què a l'Ari i a mi ens van portar els reis. Sempre heu estat allí, entre plors i tant riure que fa mal tot. Complicitat. Sortides. Recolzament incondicional. I és que mira que ens han passat coses durant aquest temps…! També penso en la gent grandiosa que ha passat pel lab: la Moniqueta, la Lali, la Lauriña… En fi, us dec molt a tots, gent!

I és clar, de laboratori en laboratori… La SílviaPe i les mil conyes (minipreeeep!), la Martí sempre amb un piropo a punt, el Dani (la bona companyia, els viatges…), la Laura, la Núria, la dolça Meri, l'Anabel, el Jon de veu suau i amor per la fotografia, l'Altafaj (quanta energia de la bona!), la Glòria (quin autèntic i sincer plaer descobrir-te, rrrubia), l'Alex i les fogueres de Sant Joan, les animalades del Gari, els Xavi Gallego-Maria-Carla sempre amb el somriure a punt en creuar-nos pel passadís. I l'Ignasi, plaer comú per la fotografia… i per la vida. Ha estat i segueix essent genial compartir moments passats, presents i futurs amb tu, ninyo… i el que encara falta!

I evidentment, d'una combinació de gent així (i algun membre afegit!) només en pot sortir un equip de volei com els Face Pelotazo's. De blau o a ratlles taronges i negres. Quantes tardes

“mordiendo arena”! He rigut tant! I quins nervissss que hem passat! Però ha valgut la pena, perquè finalment tenim copa! Ooooooeeeeeeee!!

I pugem de pis i allí hi trobem el Marcus (tantes tardes de volei i nits de birra!) i els “meus” músics. El Holger, el Lluís, l’Alexis, el John. Tot i els intents mai vam tenir nom, però van ser moltes tardes compartides creant, rient, aprenent alhora música i anglès… I amb el Holger (“hijaaa!”) i el Lluís (“hombre, Sala-x!”) , des del primer dia en què em va adoptar quan el baix era un complert extrany per a mi. I la complicitat amb el Morey, per les mil vivències compartides entre birres i algun sopar mentre has estat per aquí (bona sort pel fred nord!).

Fora de l’edifici, però sense deixar encara la ciència, ens trobem tota la colla de UPFs. De quan fa deu anys érem els reis de la fila del mig. L’Alberto (qué grande eres!), el Joan (Aaamparo!), el Sergio, el Dabit. I les meves nenes de sempre! La Clara, la Laureta, la Putxi (meiosi!!), la Carme i la Bea. “Eh, que os quiero!”. I quin perill de dones, ara, entre calçotades i converses irreverents! Amb la Bea fa anys que compartim l’aventura de viatjar (el rollo-bollo i las correrías de Boladelana y el Rábano Gigante)… i també l’aventura de viure. Ets molt gran, rèptil.

I deixant la ciència de banda… he dit banda? Els meus germans de Puri López! Leri, Ferran, Enric i Efren. Uns grandíssims còmplices. Pels grans moments (i els que encara falten!)… i per la paciència infinita i la incondicionalitat en moments baixos. Ma-que-taaaaa! Quan toquem ens ho passem teta, i això no té preu. I és clar, darrere de tota banda hi ha sempre unes *groupies*: las ovejas. Les incondicionals Lola, Lore i Bea (otra vez??)…

També hi ha la gent de Premià, és clar. Després de tants anys i seguim caminant junts, quina sort i quin orgull…! Els dofinencs! L’Anna, la Nora, la Joanna, l’Anguera, l’Albert, el Furi, l’Isaac, l’Arnau, el Llamas, el Roger. I les incorporacions posteriors del Jaime, el Xavi, la Maria, la Núria, l’Alba… Per les seves preguntes de “com van les teves cèl.lules?”, “per què treballes un diumenge al matí?”, entre d’altres. I és que aquesta feina no és fàcil d’entendre, però sempre s’han preocupat de com m’anava, de com estava. I mireu, gent, per fi comença la nova etapa. I el millor és que hi seguireu sent per contar-nos com va tot. De Premià també és el Miquel Àngel, amb qui he compartit tantes neures científiques i vitals…

I penso també en la gent del CAP, per descobrir tantes coses noves fora de la bombolla, per la injecció d’energia i de ganes que vam fabricar durant els mesos de classe… I també penso en l’Eli, amant del chaï, dels llibres i de les converses a mitja veu.

També vull fer un calaix de sastre amb tota la gent que m’ha deixat emprempta durant aquests anys. Gent que et moldeja amb un somriure, poc a poc. El meu estimat Jordi i les nostres vides paral.leles. La meva estimadíssim sister-de-mentida Iolanda i el Carles, gent bona (i viatgera!) per tenir ben a propet. El Waiman, l’emperador dels cors grans i lliures.

Som tots família.

La mare, la incansable i incondicional Tresa sempre fent costat. És com aquell raconet preuat on resguardar-se de la tempesta del món. I el Joanet, noble i ferm. Tot cor que ets tu, moco! I família, és clar, són també la iaia Antònia, l'Antònio, la Susana, la Elo, la Loli, l'Elio, el Gabriel, la Elo "petita", la Montse… una petita-gran món de gent plena de personalitat. El meu si. Per fi es completa un cicle que va començar ja fa força temps. Tinc ganes de seguir endavant també de la vostra mà, de saber-nos a tots cada dia una mica més grans, entre moments íntims i gran vetllades amb palmes, vi i carcajades del cor…
… i *tocant* al cor hi és el Marc. Algú excepcional amb qui comparteixo tant… pura força i sensibilitat. Cada dia al teu costat és un regal, una intensa aventura. L'aventura de créixer, de créixer moltíssim. D'aprendre a escoltar el propi ritme i seguir-lo amb pas ferm. I saps que des d'ara vull que sigui així, poc a poc, sense línies ni fronteres, sense càrrega, sense mapa.

…començo a comptar de nou…

Amb els que hi sou, us abraçaré, saltaré i cridaré contenta. Amb els que falten, m'imaginaré un brindis per celebrar-ho entre somriures còmplices.

This thesis work has been carried out under the direction of Doctor Susana de la Luna in the group of Gene Function in the Genes and Disease Program, which belongs to the Centre for Genomic Regulation-CRG (Barcelona).

This work is part of a general objective that try to understand the role of the protein kinase DYRK1A in the cell nucleus. Most of the hypotheses tested here come from experimental evidences previously obtained by the group: they vertebrate the project and, at the same time, served as a starting point to formulate new questions.

The report has the classical structure of Introduction, Objectives, Materials and Methods, Results, Discussion, Conclusions and References.

In the Introduction Section, three main blocks are reviewed. First, the features of the DYRK family of protein kinases are described, with specific attention to the member studied in this work, DYRK1A. The other parts are dedicated to briefly present two different and relevant issues to understand this thesis work. One of them describes the characteristics of subnuclear organelles found in the eukaryotic nucleus, and the other summarizes the current knowledge on single amino acid repeats present in eukaryotic proteins.

In the Objectives part, the main aim of this work is exposed, followed by the subobjectives emerged during the course of this thesis. After that, an accurate description of the Materials and Methods used is given as a reference tool.

The Results Section begins with the analysis of the His repeat of DYRK1A as a *bona fide* signal for the kinase targeting to nuclear speckles. After that, a more general question about the role of His runs in human proteins is addressed, finishing up with the definition of His repeats as novel directing signals to nuclear speckles. The next three parts of the Section are dedicated to the characterization of the molecule/s responsible for DYRK1A retention in nuclear speckles. Finally, part 6 investigates whether polyHis-containing proteins that accumulate in the SFC can form macromolecular complexes and, in part 7, the characterization of the interaction between two of these proteins, DYRK1A and Brn-3b, is addressed.

In the Discussion, the most relevant contributions of this work are explained and commented. The results presented are contrasted with evidences found in the literature.

In addition, the list of the main conclusions obtained during this work is given, as well as a Bibliography section including all the references mentioned along the report.

It is worth to mention that all the results regarding the characterization of polyHis repeats as novel nuclear speckle-targeting signals have generated the manuscript entitled “Genome-wide analysis of histidine repeats reveals their role in the localization of human proteins to the nuclear speckles compartment”, done in collaboration with the group of M.

Albà in the Universitat Pompeu Fabra. The manuscript has been sent to PLoS Genetics and at the moment of this writing is under revision.

ABBREVIATIONS

AP-4: activating enhancer binding protein 4
ATP: adenosine triphosphate
Arip4: androgen receptor-interacting protein 4
ASK1: apoptosis signal-regulating kinase 1
BCA: bicinchoninic acid
BLAST: basic local alignment search tool
BSA: bovine serum albumine
BrEt: ethidium bromide
C-terminal: carboxy-terminal
CBP: calmodulin binding peptide
CDK: cyclin-dependent kinase
CDKL: cdk-like kinase
cDNA: coding DNA
CK1: casein kinase 1
CLK: cdc2-like kinase
CNS: central nervous system
COMP: cartilage oligomeric matrix protein
CRE: cyclic AMP response element
CREB: CRE-binding protein
CRF1: corepressor with FHL1
CRMP4: collapsin response mediator protein 4
CtBP2: C-terminal binding protein 2
CTD: carboxy-terminal domain (RNA polymerase II)
DAPI: 4, 6-diamidine-2-phenylindol
DDX1: DEAD (Asp-Glu-Ala-Asp) box polypeptide 1
DH-box: DYRK homology box
DMEM: Dulbecco's modified Eagle Medium
DNA: desoxiribonucleic acid
DNAse: desoxiribonuclease
dNTPs: desoxiribonucleotide triphosphate
DS: Down syndrome
dsDNA: double stranded DNA
DTT: dithiotreithol
DYRK: dual specificity tyrosine-regulated kinase/dual specificity yak-regulated kinase
EDTA: ethylendiamine tetracetic acid
EGTA: ethylene glycol tetraacetic acid
eIF2B: eukaryotic initiation factor 2B
ERK: extracellular signal-regulated kinase
EST: expressed sequence tag
FADD: Fas-associated death domain
FBS: foetal bovine serum
FGF: fibroblast growth factor
FITC: fluorescein isothiocyanate
FKHR: forkhead in rabdomyosarcoma
FL: full-length
GFP: green fluorescent protein
GO: Gene Ontology
GSK3: glycogen synthase kinase 3
GST: gluthatione-S-transferase
HA: haemaglutinin
HDAC: histone deacetylase
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIC: human l-mfa domain-containing protein
HIPK: homeodomain-interacting protein kinase
His: histidine
HRP: horseradish peroxidase
HSA21: *Homo sapiens* autosome 21
IGC: interchromatin granule cluster

IgG: immunoglobulin G
IPTG: isopropyl-β-D-thiogalactopiranoside
IVK: *in vitro* kinase assay
kb: kilobase
kDa: kilodalton
JNK: jun N-teminal kinase
LB: Luria Broth medium
mnb: minibrain
mRNA: messenger RNA
MAPK: mitogen-activated protein kinase
MBK: minibrain kinase
MIG: mitotic interchromatin granule
MEF2: myocyte enhancer factor-2
NaF: sodium fluoride
NaPPi: sodium pyrophosphate
$NaVO_4$: sodium orhovanadate
NCBI: National Center of Biotechnology Information
NES: nuclear export sequence
NFAT: nuclear factor of activated T cells
NGF: nerve growth factor
NLS: nuclear localization signal
NOR: nucleolar organizer region
N-terminal: amino-terminal
OD: optic density
OMA-1: oocyte maturation protein 1
ORF: open reading frame
PAGE: polyacrilamide gel electrophoresis
PBS: phosphate-buffered saline
PcG: polycomb group
PCR: polymerase chain reaction
PKA: c-AMP-dependent protein kinase
PML: promyelocytic leukemia protein
PMSF: metylphenylsulfonyl fluoride
POD: PML oncogenic domain
POUhd: POU homeodomain
POUsd: POU specific domain
PSA: alkaline phosphatase
PSP: paraspeckle protein
PTB: poly-pyrimidine-tract-binding protein
P-TEFb: positive transcription elongation factor
RBM: RNA-binding motif
RGC: retinal ganglion cell
RMS: rabdomyosarcoma
RNA: ribonucleic acid
RNAi: RNA interference
RNP: riboucleoprotein
RRM: RNA recognition motif
rRNA: ribosomal RNA
RT: room temperature
SAR: single amino acid repeat
SARP: single aminoacid repeat-containing protein
SCA: spinocerebellar ataxia
SFC: splicing factor compartment
SDS: sodium dodecil sulfate
SIP1: survival of motor neuron protein interacting protein 1
SMA: spinal muscular atrophy
SMART: Simple Modular Architecture Research Tool
SMN: survival of motor neuron
SNR1: Snf-5 related 1
snRNA: small nuclear RNA
snRNP: small nuclear ribonucleoprotein
snoRNP: small nucleolar ribonucleoprotein
SP-3: serine/proline repeat 3
SRPK: SR-protein kinase
SRR-1: serine-rich region 1
ssDNA: single stranded DNA
TAP: tandem affinity purification
TBS: tris-buffered saline
TGFβ: transforming growth factor beta

TNF-α: tumor necrosis factor alpha

TRX: trithorax

UHD: upstream homology domain

UV: ultraviolet

vol: volume

INTRODUCTION

1. THE FAMILY OF DYRK PROTEIN KINASES

The family of DYRK (dual-specificity tyrosine-regulated kinases or dual-specificity Yak-related kinases) proteins belongs to the group of CMGC kinases, within the superfamily of eukaryotic protein kinases (Becker and Joost, 1999). The CMGC group also includes cyclin-dependent kinases (CDKs), glycogen synthase kinases (GSKs), CDK-like kinases (CDKLs), cdc2-like kinases (CLKs), serine/arginine-rich domain (SR) protein kinases (SRPKs), and mitogen-activated protein kinases (MAPKs) (Manning et al., 2002) (Figure I1).

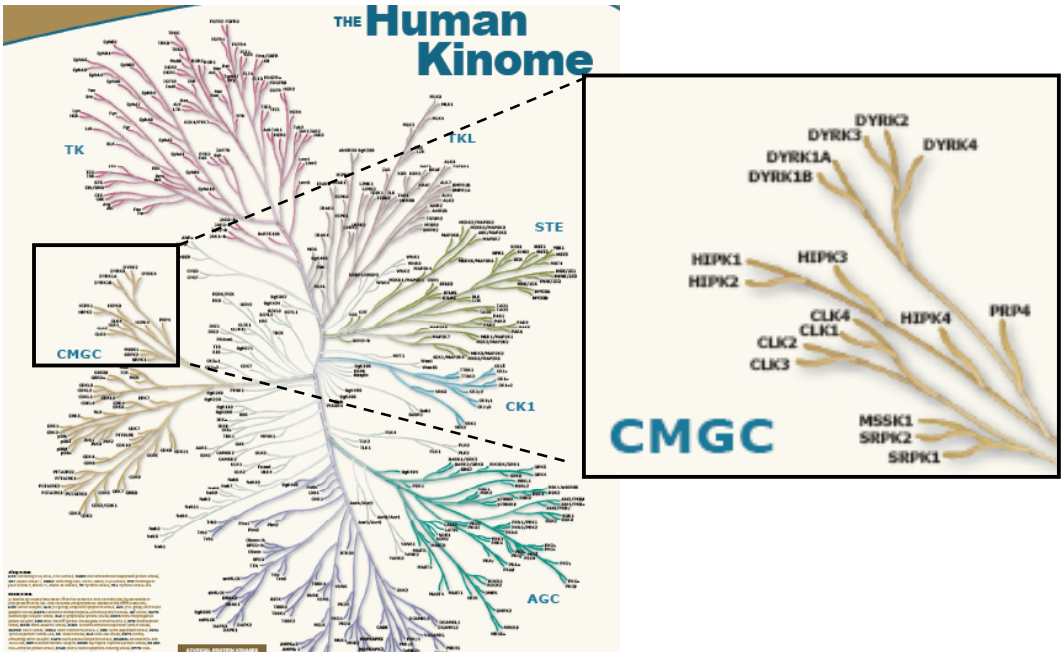


Figure I1: Phylogenetic tree of the superfamily of human protein kinases. Human kinases are classified primarily by sequence comparison of their catalytic domains, giving rise to the seven groups of protein kinases found in other eukaryotic kinomes. On the right, a magnification of the CMGC group which DYRK kinases belong to (adapted from Manning et al., 2002).

The members of the DYRK family of protein kinases show a high degree of conservation and are present all along the eukaryotic evolution (Becker and Joost, 1999). Regarding the conservation and structure of the proteins, the DYRK family is divided in two subfamilies: the DYRK subfamily and the homeodomain-interacting protein kinase (HIPK) subfamily. The DYRK subfamily is divided, in turn, in two phylogenetically related groups: i) the group of the kinases related to the *Saccharomyces cerevisiae* Yak1p protein, which includes YakA (*Dictyostelium discoideum*), MBK1 (*Caenorhabditis elegans*), minibrain (*Drosophila melanogaster*) and DYRK1A and DYRK1B (vertebrates), and ii) the group of kinases evolutionary related to *Schizosaccharomyces pombe* Pom1p, which includes MBK2

(*C. elegans*), dDYRK2 and dDYRK3 (*D. melanogaster*) and DYRK2, DYRK3 and DYRK4 (vertebrates) (Figure I2).

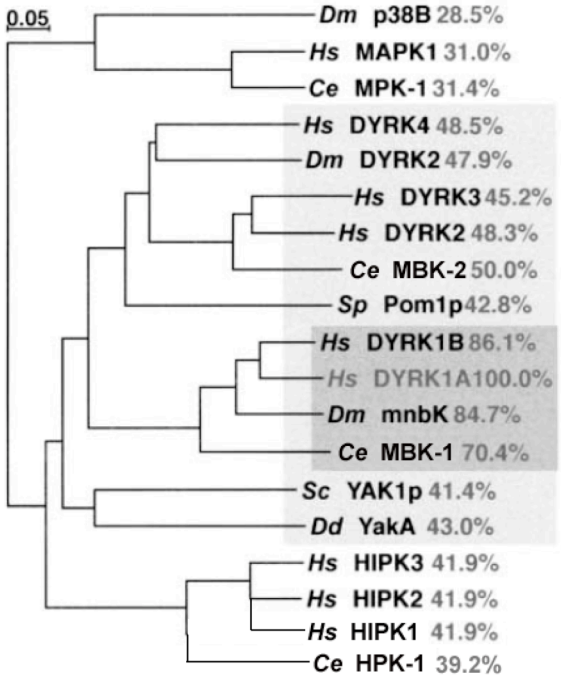


Figure I2: Classification of the DYRK family of protein kinases. The phylogenetic tree was generated by the alignment of the kinase domains of all the members. The percentage identity between each kinase domain and that of human DYRK1A is listed on the right. *Ce, C. elegans; Dd, D. discoideum; Dm, D. melanogaster; Hs, Homo sapiens; Sc, S. cerevisiae; Sp, S. pombe* (adapted from Raich et al., 2003).

It has been suggested that the two DYRK phylogenetic subgroups are also linked to distinct subcellular localization: nuclear for the Yak1p-related group and cytosolic for the Pom1p-related group (Becker et al., 1998). However, this proposal has to be reviewed since at least three members, Yak1p, DYRK1A and DYRK2, have the ability to shuttle between the cytoplasm and the nucleus (Alvarez, 2004; Moriya et al., 2001; Taira et al., 2007).

DYRK-related kinases present unique structural and enzymatic features that distinguish them with regard to other CMGC protein kinase subfamilies. These include three specific sequence motifs in the vicinity of the catalytic cleft: i) the YxY motif between subdomains VII and VIII, which corresponds to the autophosphorylation loop; ii) a SSC motif following the conserved DFG motif of subdomain VII, and iii) the conserved sequences HCDLKPEN and YXYSRFYR(S/A)PE in the subdomains VI and VII, respectively (Becker and Joost, 1999). Outside the kinase region, the homology is restricted to a short stretch of acidic residues (DDDNxDY) immediately preceding the catalytic domain and named as the DYRK homology box or DH-box (Becker and Joost, 1999). Molecular modelling in DYRK1A suggests that the D- box and the catalytic domain form a common tertiary structure (Himpel et al., 2001). Moreover, the DH-box is required for the kinase activity of the *D. melanogaster* dDYRK2 protein (Kinstrie et al., 2006). Finally, sequence similarity in the N- and C-terminal regions is apparent only in closely related members of the family (Becker and Joost, 1999).

All the DYRK family kinases are considered as dual-specificity kinases due to their ability to phosphorylate Tyr and Ser/Thr residues, although only Ser/Thr residues are phosphorylated in exogenous substrates (Garrett et al., 1991; Kentrup et al., 1996; Kinstrie et al., 2006; Nishi and Lin, 2005; Souza et al., 1998). To be catalytically active, DYRK proteins have to be phosphorylated in the second Tyr residue of the YxY motif within the activation loop (Himpel et al., 2001; Kentrup et al., 1996). The position of this motif corresponds to the TxY motif present in other protein kinases of the CMGC group that are activated by phosphorylation of these residues (Miyata and Nishida, 1999). However, the activation mechanism of DYRK proteins differs from that of the MAPKs, in which their TxY residues are targeted by protein kinases that operate upstream in the signalling pathway (for a review, see Raman et al., 2007). This DYRK activation mechanism was proposed by Lochhead and co-workers using the *D. melanogaster* minibrain and the paralog dDYRK2 as examples. The model suggests that the autophosphorylation of the critical Tyr in the activation loop is intramolecular and mediated by a transitional intermediate "Tyr kinase" during the translation of the protein. Once being mature, the kinase suffers a conformational change that induces a variation in the specificity of the substrate, such as the mature species functions then as a Ser/Thr kinase (Lochhead et al., 2005) (Figure I3).

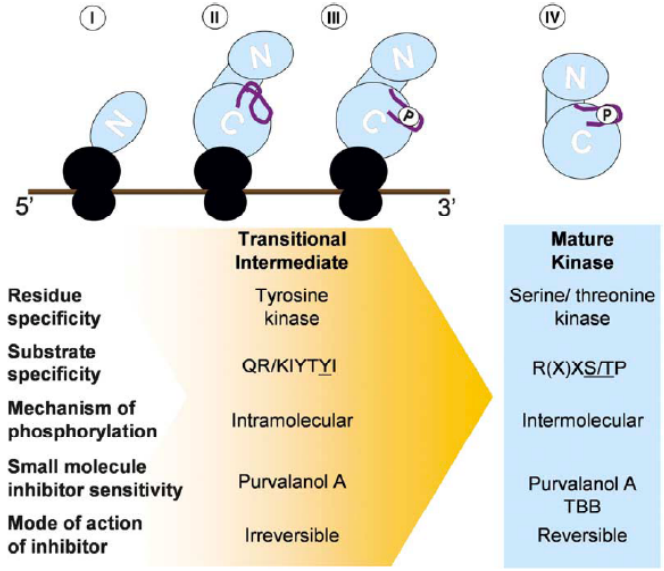


Figure I3. Model for DYRK activation mechanism. (I) During translation, the polypeptide begins to form a tertiary structure. (II) During folding of the polypeptide, an active intermediate is formed, and the activation loop is the first substrate seen by the nascent kinase domain. (III) The intermediate possesses Tyr kinase activity that phosphorylates the critical Tyr residue of the activation loop. (IV) The fully translated and activation loop- phosphorylated DYRK is now in its mature conformation. The Tyr-kinase activity is lost, and the kinase functions as a Ser/Thr kinase. The residue and substrate specificities and sensitivities to small-molecule inhibitors are outlined below (adapted from Lochhead et al., 2005).

Analysis of phosphate incorporation into synthetic peptides allowed to define the consensus sequence of phosphorylation of several DYRK members, such as DYRK1A (RPX(S/T)P) (Himpel et al., 2000) and DYRK2 and DYRK3 (RX(S/T)P) (Campbell and Proud, 2002). These optimal substrate sequences are similar to that of ERK1/2 (PX(S/T)P) (Songyang et al., 1996), and indicate that DYRK proteins are proline-directed kinases due to the presence of a Pro residue in the +1 position.

An important characteristic of DYRK proteins is their ability to act as priming kinases, in the sense that phosphorylation of a given residue allows a more efficient phosphorylation by a subsequent protein kinase. Considering the type and the relative position of the phosphorylated residue, priming mediated by different DYRK proteins can be classified as: i) GSK3 classical priming; ii) GSK3 non-classical priming, and iii) polo-like kinase (PLK) priming. Figure I4 exemplifies these three different scenarios.

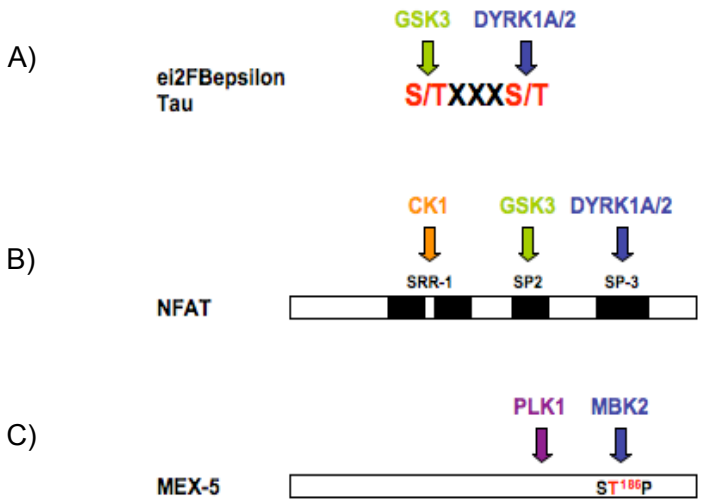


Figure I4: DYRK proteins act as priming kinases. A) GSK-3 classical priming: GSK3 requires a priming phosphate at +4 position in order to phosphorylate many of its substrates (Fiol et al., 1988). DYRK1A and DYRK2 phosphorylate the +4 residue in the GSK3 substrates eI2Fβ and tau (Woods et al., 2001). B) GSK3 non-classical priming: DYRK1A and DYRK2 phosphorylate the conserved Ser/Pro repeat 3 (SP-3) motif of the NFAT regulatory domain, thus priming further phosphorylation of the SP-2 and Ser-rich region 1 (SRR-1) motifs by GSK3 and casein kinase 1 (CK1), respectively (Arron et al., 2006; Gwack et al., 2006). This priming mechanism has been also described for *C. elegans* oocyte maturation protein 1 (OMA-1), in which phosphorylation of Thr^{239} by the DYRK-family kinase MBK2 potentiates GSK3-mediated phosphorylation of Thr^{339} (Nishi and Lin, 2005). C) PLK priming: MBK2 primes Thr^{186} in the polarity protein MEX5 for subsequent and enhanced PLK1-dependent phosphorylation (Nishi et al., 2008).

As previously mentioned, DYRK proteins are present along all the eukaryotic evolution. Once the common features present in all the members of the family have been introduced, specific characteristics of each kinase from the subfamilies HIPK and DYRK will be further discussed.

1.1 The subfamily of HIPK protein kinases

The mammalian HIPK family includes four highly conserved nuclear Ser/Thr kinases, which were originally identified through their ability to interact with NK homeoproteins (Kim et al., 1998). HIPKs are involved in the regulation of gene transcription during two fundamental biological events: embryonic development and cell response to DNA damage (reviewed in Rinaldo et al., 2007).

HIPK2 is the best-studied HIPK protein. Its primary sequence includes an N-terminal region containing a sumoylation site and a kinase domain, followed by a homeobox-interacting domain and a PEST sequence. The C-terminal region comprises a speckle-retention signal, a putative autoinhibitory domain and an ubiquitination site. Overexpression studies indicate that HIPK2 localizes mainly in the so-called HIPK2 bodies, and upon genotoxic stimuli, it can be recruited to promielocytic leukemia (PML) bodies (Hofmann et al., 2002). The work of several groups indicates that HIPK2 is a pleiotropic regulator of gene expression through the interaction and/or phosphorylation of a variety of transcription factors and cofactors from crucial signalling pathways, such as the axis transforming growth factor β (TGFβ-Smad3 (Zhang et al., 2007) or Wnt-1 (Kanei-Ishii et al., 2004) (reviewed in Rinaldo et al., 2007) (Figure I5). In general, HIPK2 is activated in response to morphogenic signals and DNA-damaging agents and, accordingly, HIPK2-guided gene expression programs trigger differentiation and development or apoptosis. In the former, HIPK2 is known to play a co-repressor role for several homeobox transcription factors (Kim et al., 1998; Wiggins et al., 2004). Regarding the response to DNA damage, the main role of HIPK2 is the phosphorylation of the tumor suppressor p53 at residue Ser[46], which enhances the p53-mediated transcriptional activation of proapoptotic factors and the repression of anti-apoptotic molecules (D'Orazi et al., 2002; Hofmann et al., 2002). Studies in the central nervous system (CNS) have defined a dual role for the kinase: it displays pro-apoptotic effects in sensory neurons (Wiggins et al., 2004), whereas it is required for TGFβmediated survival during programmed cell death in dopaminergic neurons (Zhang et al., 2007).

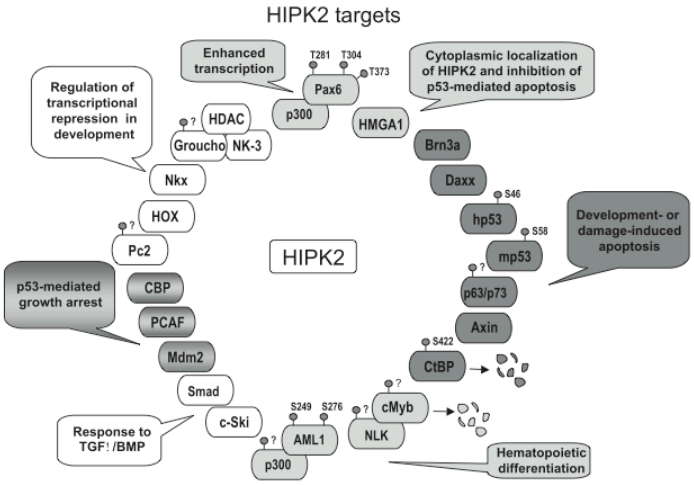


Figure I5: Schematic representation of HIPK2 targets. HIPK2 functions by acting on chromatin-remodelling complexes, influencing the stability or localization of transcription factors and co-regulators, and by modifying protein/protein and/or protein-DNA interactions. The proteins that were shown to interact with HIPK2 are reported together with their respective biological activities. The known sites of HIPK2-phosphorylation are indicated. The proteolitic degradation induced by HIPK2-mediated phosphorylation of CtBP and c-Myb is also reported (adapted from Rinaldo et al., 2007).

HIPK1 is a modulator of p53 activity that can promote oncogenesis (Kondo et al., 2003). Moreover, HIPK1 has been also implicated in apoptosis regulation through the participation in the tumor necrosis factor alpha (TNFα-mediated activation of the apoptosis signal-regulating kinase 1 (ASK1)-JNK/p38 axis (Li et al., 2005). In fact, analysis of the double *Hipk1/Hipk2* knock-out mice show some functional redundancy between the two kinases, with overlapping roles in mediating cell proliferation and apoptosis in response to morphogenetic and genotoxic signals during development (Isono et al., 2006).

HIPK3 regulates the Fas signalling pathway in two different ways: i) by phosphorylation of the Fas-associated death domain (FADD), and ii) by phosphorylation of an unknown target protein that negatively regulated signals leading to Fas ligand-induced JNK activation (Rochat-Steiner et al., 2000).

HIPK4 is the less conserved member and was identified by homology with the other members of the family during the human genome-sequencing project (Gerhard et al., 2004). HIPK4 phosphorylates p53 at Ser^9, increasing its repression activity at p53 repressive promoters (Arai et al., 2007).

1.2 The subfamily of DYRK protein kinases

During the last years, several members of the DYRK subfamily of protein kinases have been identified, showing that the DYRK subfamily has expanded during evolution, arising from a sole member in lower eukaryotes to several paralogs in humans.

1.2.1 DYRK proteins in lower eukaryotes

The yeast *S. cerevisiae* and *S. pombe* and the fungus *D. discoideum* present a unique DYRK kinase (Yak1p, Pom1p and YakA, respectively). In general terms, the activity of these proteins is linked to the control of the cell cycle, cytokinesis and cell differentiation.

Yak1p was identified as a negative regulator of cell proliferation under stress conditions, functionally antagonizing the RAS/protein kinase A (PKA) pathway (Garrett et al., 1991). Yak1p is part of the sensor system that detects the presence of environmental glucose. Upon nutrient deprivation, Yak1p is activated by PKA depletion and the subsequent Msnp2/Msnp4 activation, and promotes cell cycle arrest through phosphorylation of the transcription factor Pop2p (Moriya et al., 2001). Yak1p also represses the transcription of ribosomal protein genes by antagonizing the TOR/RAS/PKA pathway through phosphorylation of the corepressor CRF1 (Martin et al., 2004). Yak1p has the ability to shuttle from the cytoplasm to the nucleus in response to glucose (Moriya et al., 2001). Interestingly, the molecular action of the kinase is related with the control of the subcellular

localization of the transcriptional regulators it modulates (Martin et al., 2004; Moriya et al., 2001).

Pom1p is involved in providing positional information both for polarized growth and for cell division in the fission yeast *S. pombe* (Bahler and Pringle, 1998). Pom1p kinase activity is cell cycle regulated: it is high during symmetrical growth and division, but lower in newly born cells that growth asymmetrically (Bahler and Nurse, 2001). The localization of Pom1p depends on its kinase activity, on microtubules and on the polarity factor Tea1p (Bahler and Pringle, 1998), and changes during cell cycle: it associates with the cell cortex at the tips in interphase and with the cell middle during mitosis (Bahler and Nurse, 2001). During cell division, Pom1p is necessary for the formation of the contractile actomyosin ring (Celton-Morizur et al., 2006; La Carbona and Le Goff, 2006). On the other hand, the complex Tea1p, Tea4p and Pom1p prevents division-septum assembly at the cell ends (Huang et al., 2007). Finally, Pom1p plays a crucial role in the translation of microtubule-dependent polarity cues into F-actin formation, through a cdc42 GTPase-activating protein (Tatebe et al., 2008).

YakA is the orholog of the DYRK family in the fungus *D. discoideum*, and was initially described as a regulator of the transition from growth to development (Souza et al., 1998). YakA mRNA accumulates during growth and peaks in abundance at the time of starvation, when it promotes growth arrest by decreasing vegetative mRNA expression and increasing PKA activity. This ultimately favours the initiation of development by induction of the cAMP expression (Souza et al., 1998; van Es et al., 2001). In a similar way, YakA/PKA/cAMP also participate in cell growth arrest under heat and oxidative stress conditions (Taminato et al., 2002).

1.2.2 DYRK proteins in *C. elegans*

Two DYRK genes were identified in the genome of the worm *C. elegans*, termed *mbk1* and *mbk2*.

MBK1 is expressed in the nucleus of all somatic cells. During development, its expression is firstly observed when morphogenesis begins and increases during later embryonic stages, remaining at comparable levels throughout larval and adult stages. The loss of the gene has no obvious defects, whereas its overexpression results in olfactory defects that correlate with *mbk1* gene dosage (Raich et al., 2003).

MBK2 is expressed in the cytoplasm of the cells. The null *mbk2* mutant is lethal due to defects in spindle positioning and cytokinesis in the early embryo (Raich et al., 2003). Functional studies have defined several roles for the protein, such as i) degradation of maternal proteins during the oocyte-to-embryo transition (Nishi and Lin, 2005; Pellettieri et al., 2003; Quintin et al., 2003; Shirayama et al., 2006; Stitzel et al., 2006); ii) appropriate

localization of the mitotic spindle (Pang et al., 2004), and iii) regulation of the proper asymmetric distribution of cytoplasmic factors in the fertilized embryo, including the P-granules and the germline-specific factor PIE-1 (Pang et al., 2004). Very recently, MBK2 has reported to be a priming kinase for PLK1 to enhance the function of the maternal polarity factor MEX-5 (Nishi et al., 2008).

1.2.3 DYRK proteins in *D. melanogaster*

Three members of the DYRK family exist in *Drosophila*: minibrain (mnb), the ortholog of DYRK1A and DYRK1B; dDYRK2, ortholog of DYRK4 and a third member, dDYRK3, only present in the databases, which could be the ortholog of vertebrate DYRK2 and DYRK3.

Mutant *mnb* flies owe their name to a specific reduction in the volume of the adult optic lobes and central brain hemispheres, which correlates with specific abnormalities in visual and olfactory behaviour. The reduction in the adult brain size is caused by an abnormal spacing of neuroblasts in the outer proliferation centre of the larval brain, rather than by neuronal architecture defects (Tejedor et al., 1995).

dDYRK2 is the putative protein product of the gene *smi35A*. It is expressed in the cytoplasm and its kinase activity is maintained at all stages of fly development, with elevated levels observed during embryogenesis and pupation (Lochhead et al., 2003). Both dDYRK2 and Mnb interact with the chromatin-remodelling factor SNR1 and the essential chromatin component TRX and phosphorylate at least SNR1, although the structural requirements are not the same for each kinase. In the case of dDYRK2, the interaction depends on the C-terminal region of the kinase and the DH-box is necessary for the phosphorylation to occur, features that have not been observed for Mnb. This suggests the existence of fundamental structural/functional differences between the two kinases (Kinstrie et al., 2006).

1.2.4 DYRK proteins in mammals

Five members of the DYRK family are found in mammals: DYRK1A, DYRK1B, DYRK2, DYRK3 and DYRK4. Their homology is restricted to the kinase domain, with a 70% of identity among them. However, in the case of the closest relatives DYRK1A and DYRK1B, this homology reaches the 95% and is extensive to the nuclear localization signals (NLS) in the N-terminal region and in the C-terminal region downstream the kinase domain (Alvarez, 2004; Becker et al., 1998) (Figure I6).

Figure I6: Comparison of the sequence features in mammalian DYRK kinases. NLS: nuclear localization signal; DH-box: DYRK-homology box; His: histidine tract; Ser/Thr: region enriched in Ser and Thr residues

The analysis of the genomic structure reveals that the five members arose by gene duplication events, which allow their classification into two subgroups: DYRK1A/DYRK1B and DYRK2/DYRK3. The evolutionary process of DYRK4 is still unknown. When overexpressed, DYRK1A and DYRKB accumulate in the nucleus, while DYRK2, DYRK3 and DYRK4 tend to be found in the cytoplasm. However, as in the case of the ortholog Yak1p, DYRK1A, DYRK1B and DYRK2 are shuttling kinases (Alvarez, 2004; Taira et al., 2007) whose subcellular localization changes upon certain stimuli. Regarding tissue expression, DYRK1A is expressed ubiquitously, whereas transcripts for the other members are more restricted to specific tissues, as detailed in Table I1.

Table I1: Characteristics of human DYRK protein kinases

Protein	Size (amino acids)[a]	Chromosomic localization[b]	Tissue expression
DYRK1A	763/754	21q22.2	Ubiquitous
DYRK1B/MIRK	629/601/589	19q12-13	Testes, heart, muscle, carcinoma
DYRK2	528/601	12q14.3	Testes, brain
DYRK3/REDK	568/588	1q32	Testes, fetal liver, erithroid cells
DYRK4	524	12p13.3	Testes

a: number of amino acids of the described isoforms resulting from alternative splicing events

DYRK1B has been implicated in skeletal muscle differentiation and survival, by blocking the proliferation of cycling myoblasts in G0 and limiting the apoptosis in fusing myoblasts (reviewed in Mercer and Friedman, 2006). These effects are very likely the consequence of DYRK1B action at different levels: i) phosphorylation of class II histone deacetylases (HDACs), which reduces their nuclear accumulation and allows the myoblast differentiation transcription factor MEF2 to activate myogenin transcription (Deng et al., 2005); ii) phosphorylation of cyclin D1 and D3, which destabilizes them (Ewton et al., 2003; Takahashi-Yanaga et al., 2006; Zou et al., 2004), and of the CDK inhibitor p27kip1, which blocks its degradation (Deng et al., 2004); and iii) phosphorylation of p21cip1, which maintains its cytoplasmic localization in postmitotic myotubes and favours its anti-apoptotic

role (Mercer et al., 2005). In close connection with this role in G0 arrest, the transcription of DYRK1B is strongly upregulated under mitogen deprivation conditions (Deng et al., 2003; Lee et al., 2000). Unexpectedly, amplification of the DYRK1B gene, upregulation of DYRK1B expression and/or constitutive activation of the kinase have been observed in several different types of cancer, such as colon, pancreatic and ovarian tumours (reviewed in Friedman, 2007). This apparent contradiction could be explained by a pro-survival role of DYRK1B in tumors. Loss of the G0 checkpoint by genetic alterations could abrogate the post-translational G0 growth arrest effects of DYRK1B without affecting its survival function. Morever, in pancreatic cancer DYRK1B is activated by the oncogene K-ras and mediates some of the survival signals mediated by Ras signalling (Jin et al., 2007). Additionally, it has been observed that DYRK1B becomes restricted from the nucleus in most tumors (Deng et al., 2006; Mercer et al., 2006), in agreement with the loss of pro-differentiation and growth arrest functions, all of which most likely require DYRK1B nuclear activity.

During the last years, several DYRK2 *in vivo* and *in vitro* substrates have been identified: histone H3 (Himpel et al., 2000), histone H2B (Becker et al., 1998), glycogen syntase (Skurat and Dietrich, 2004), the neurite outgrowth regulator CRMP4 (Cole et al., 2006), the translation initiation factor eIF2B (Woods et al., 2001), the microtubule-associated protein tau (Woods et al., 2001) and the transcription factor NFAT1 (Gwack et al., 2006). Very recently, DYRK2 was shown to directly phosphorylate and induce the proteasome-dependent degradation of the key Hedgehog pathway-regulated transcription factor GLI2 (Varjosalo et al., 2008). Furthermore, recent work indicates that upon exposure to genotoxic stress, DYRK2 phosphorylates p53 at Ser^{46}, inducing cell apoptosis in a DNA damage-dependent manner (Taira et al., 2007). Interestingly, the DYRK kinase HIPK2 phosphorylates the same residue upon UV irradiation (D'Orazi et al., 2002), suggesting a functional conservation between both kinases. As in the case of DYRK1B, DYRK2 growth control activities are in contradiction with the fact that DYRK2 mRNA is highly overexpressed in esophageal and lung adenocarcinomas (Miller et al., 2003).

DYRK3 is expressed primarily in erythroid progenitor cells and modulates late erythropoiesis, by acting as a developmental suppressor (Geiger et al., 2001), and inducing hematopoietic progenitor cells apoptosis due to cytokine withdrawal (Li et al., 2002). The kinase also regulates CREB response pathways through CREB phosphorylation and transcriptional activation. This occurs via PKA, since PKA specific inhibitors block DYRK3-mediated phosphorylation of CREB (Li et al., 2002). DYRK3 has been identified in a wide variety of screenings related to cell survival and death. First, and together with DYRK1B, DYRK3 was identified as a pro-survival kinase in a RNAi screening of human kinases and

phosphatases that regulate chemoresistance and apoptosis (MacKeigan et al., 2005). Secondly, DYRK3 was also identified in the search of expressed genes that characterize long-term survival in malignant glioma patients. In this case, DYRK3 expression levels correlate with survival expectations (Yamanaka et al., 2006). Moreover, a very recent study reports increased levels of DYRK3 mRNA in high-invasive mouse adenocarcinoma cell strains (Sargent et al., 2008).

DYRK4 has been characterized as a testis-specific kinase, with a very restricted expression to stage VIII postmeiotic spermatids. However, the knock-out mice are fertile and do not present any functional defect in spermatogenesis, probably due to the functional redundance of other DYRK kinases expressed during spermiogenesis (Sacher et al., 2007). DYRK4 is upregulated in a paradigm of retinoic-acid-induced differentiation of a neuronal precursor cell line (Leypoldt et al., 2001). In addition, DYRK4 is found in the group of genes identified as promoters of tumour resistance to radiation in the treatment of cervical cancer (Tewari et al., 2005).

What follows from this summary is that DYRK kinases seem to be acting in signalling pathways controlling differentiation and survival. Moreover, DYRK kinases are over-expressed in a great variety of tumours, a fact that can be linked to their pro-survival roles.

1.3 The protein kinase DYRK1A

DYRK1A maps in the human chromosome 21 (HSA21), specifically in the region 21q22.2 (Guimera et al., 1996; Shindoh et al., 1996; Song et al., 1996). *DYRK1A* is organized in 17 exons and several transcripts have been described, due to the use of different promoters and/or the alternative splicing of the mRNA, which do not affect the predicted coding region (Guimera et al., 1999; Maenz et al., 2008; Wang el tal., 1998) (Figure I7).

Another alternative splicing event has been described in exon 6, which gives rise to two protein isoforms of 763 and 754 amino acids (Guimera et al., 1997). No differences in the tissue expression or the function of these protein isoforms have been reported so far.

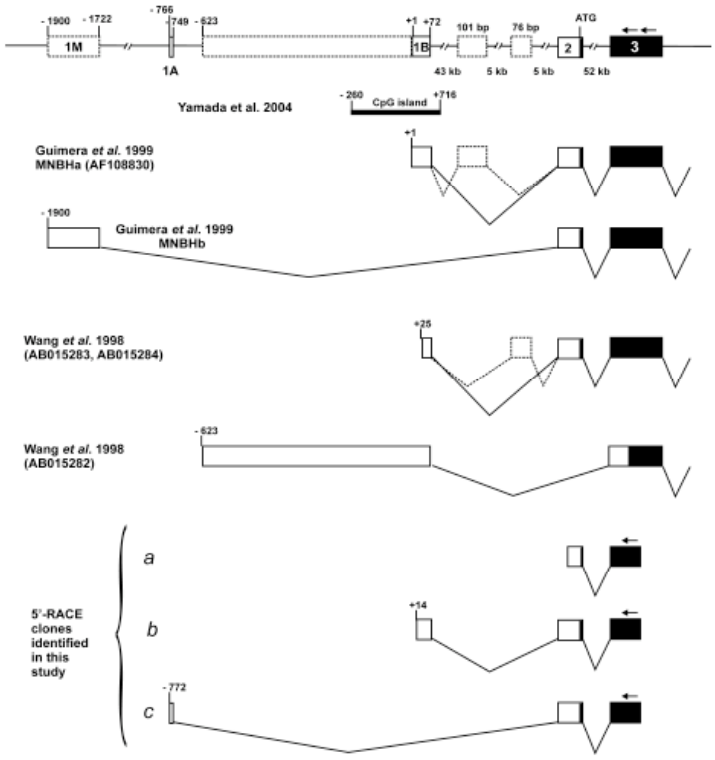


Figure I7: Mapping of transcripts to the 5'-region of the human *DYRK1A* gene. A schematic representation of the exon-intron organization is shown on top. Exons are numbered (1M, 1A, 1B, 2 and 3) and drawn to scale. Open boxes are noncoding regions. Filled boxes represent the coding sequence. Open boxes with dotted lines are alternative exonic sequences. Numbering of nucleotides in the promoter region refers to the 5'-end transcript variant 3 (NM_101395) in the NCBI database. Different DYRK1A transcript variants (accession numbers indicated) have been defined in several studies (references indicated) (adapted from Maenz et al., 2008).

DYRK1A has been proposed as a candidate gene responsible for some of the phenotypes of Down syndrome (DS) individuals (Epstein, 1990). The gene *DYRK1A* is overexpressed in DS embryonic brain at the RNA level (Guimera et al., 1999), and in adult brain at the protein level at approximately 1.5 fold (Dowjat et al., 2007; Wegiel et al., 2004), confirming that the protein is over-expressed in a gene dosage-dependent manner as postulated in the “gene dosage effect” hypothesis, a prevailing concept in DS research (Epstein, 1990). Remarkably, in the group of DS infants 1 to 3 years of age, the levels of DYRK1A did not significantly differ from the control cases, suggesting the existence of developmentally regulated mechanism of DYRK1A expression (Dowjat et al., 2007). On the other hand, a recent study reports the truncation of the *DYRK1A* gene in two patients with microcephaly (Moller et al., 2008), with phenotypes resembling that of individuals with partial monosomy 21 (Matsumoto et al., 1997), suggesting that DYRK1A decreased dosage is directly responsible of the disorder.

The analysis of DYRK1A over-expression mouse models has allowed to evaluate the contribution of *DYRK1A* to the neuropathological traits associated to DS (Ahn et al., 2006; Altafaj et al., 2001; Branchi et al., 2004; Rachidi et al., 2007; Smith et al., 1997). Both phenotypic and behavioural studies indicate that mutant mice display impairment in spatial learning and memory, hyperactivity during development and increased brain weight and neuronal size. To date, there is only one DYRK1A null mice created by gene targeting (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 (Dyrk1A$^{+/-}$) show decreased neonatal viability and a significant body size reduction from birth to adulthood, more important in specific regions as the brain or the cerebellum. The neurobehavioral analysis revealed preweaning developmental delay and specific motor (Fotaki et al., 2004) and learning deficits (Fotaki et al., 2002) in adult Dyrk1A$^{+/-}$ mice. The brains of these animals are decreased in size in a region-specific manner, although the cytoarchitecture and neuronal components in most areas are not altered. The reduction in the brain size and behavioural effects observed provide evidences about the non-redundant vital role of DYRK1A and suggest a conserved mode of action that determines normal growth and brain size in both mice and flies (Fotaki et al., 2002; Tejedor et al., 1995).

Altogether, gain- and loss of function studies confirm that normal amounts of the DYRK1A gene are necessary for normal brain development, so either a reduction or excess dosage interfere with it.

1.3.1 Gene and protein expression

The gene *DYRK1A* is broadly expressed in both developing and adult tissues in mammals, with higher prevalence in heart, lung, brain and skeletal muscle (Guimera et al., 1999; Guimera et al., 1996; Okui et al., 1999; Shindoh et al., 1996; Song et al., 1996). Studies performed in chick (Hammerle et al., 2002) and mouse (Hammerle et al., 2008) early embryonic, mid/late and postnatal tissues indicate that *Dyrk1A* displays a rather dynamic pattern of expression during vertebrate development. Moreover, the close similarity observed in the expression patterns of both organisms probably reflects the evolutionary conservation in the regulatory mechanisms of *Dyrk1A* gene expression. Regarding transcriptional control, it is known that human and mouse *DYRK1A* promoters are positively regulated by the transcription factor E2F1 (Maenz et al., 2008), and that the complex formed by the transcription factor AP-4 and the co-repressor geminin negatively regulates the expression of *DYRK1A* in non-neuronal cells (Kim et al., 2006). Moreover, several microarray studies have demonstrated striking changes in the abundance of *DYRK1A* mRNA in various systems of cellular differentiation and proliferation (Chang et al., 2007; Choong et al., 2004; de Wit et al., 2002; Subrahmanyam et al., 2001; Teague et al., 1999). An additional level of gene regulation is given by the localization of the mRNA. In the chick embryo, the mRNA segregates asymmetrically in neural progenitors in the transition from proliferative to neurogenic divisions during development (Hammerle et al., 2002).

In agreement with its RNA expression profile, Dyrk1A shows an ubiquitous pattern of expression during rodent development, with strong reactivity in CNS regions (Okui et al.,

1999; Rahmani et al., 1998). The levels of Dyrk1A protein gradually decreased at the adult stage in all tissues studied, although they remain high in the heart and CNS (Okui et al., 1999), where it is clearly apparent in the olfactory bulb, cerebellum, spinal cord and most motor nuclei of midbrain and brain stem (Marti et al., 2003). A systematic analysis of the cellular pattern of Dyrk1A expression in the mouse developing CNS indicated that the protein is specifically expressed in four sequential steps, from preneurogenic progenitors to late differentiating neurones (Hammerle et al., 2008). The expression pattern of DYRK1A in human brains differs from that reported in the adult mice, and reveals quantitative differences both among brain regions and cell types within the same region (Wegiel et al., 2004). Interestingly, DYRK1A distribution changes with aging, with an accumulation in corpora amylacea in the memory and motor system subdivisions of adult brains, and therefore a contribution to neurofibrillary degeneration in DS and Alzheimer disease has been suggested (Liu et al., 2007; Ryoo et al., 2008; Wegiel et al., 2008; Wegiel et al., 2004).

1.3.2 Protein features

The analysis of amino acid sequence reveals the presence of several characteristic features, from N- to C-terminal (Figure I8): a functional bipartite nuclear localization signal (NLS_1) (Alvarez et al., 2003); a DYRK homology box (DH-box) with unknown function; a kinase domain; a second nuclear localization signal (NLS_2) between kinase subdmains X and XI, which overlaps with a nuclear export signal (NES) (Alvarez, 2004; Alvarez et al., 2003); a PEST region rich in Pro, Gln, Ser and Thr commonly associated to the rapid degradation of proteins, but with unknown activity in DYRK1A; a stretch of 13 consecutive His residues, which targets the protein to the splicing factor compartment (SFC) (Alvarez et al., 2003); and a region enriched in Ser and Thr residues at the C-terminus of the protein.

Figure I8: Schematic representation of the structural features of the protein kinase DYRK1A. NLS, N-terminal nuclear localization signal (105-125); DH-box, DYRK homology box (141-147); Kinase, kinase domain (159-479); PEST, region rich in Pro, Gln, Ser and Thr residues (482-525); His, poly-His stretch (599-616); Ser/Thr, sequence rich in Ser and Thr residues (659-672). The asterisk indicates the spliced site and the number of amino acids in the two alternative spliced isoforms is given.

Since DYRK1A is synthesized as a constitutively active kinase, it is thought to be regulated through other control mechanisms such as tyrosine phosphatases, subcellular localization, protein stability or interaction with other proteins. For instance, Tyr

phosphorylation and DYRK1A kinase activity were reported to be induced by FGF in hippocampal progenitor cells (Yang et al., 2001); autophosphorylation in Ser^{520} residue within the PEST region allows the binding of 14-3-3β, which in turn stimulates the catalytic activity of DYRK1A (Alvarez et al., 2007); and DYRK1A translocation to the cytosol ha been described during Purkinje cell differentiation (Hammerle et al., 2003).

Substrates for DYRK1A have been reported over recent years, including cytosolic proteins and several transcription and splicing factors. The distinct nature of these substrates likely responds to the ability of DYRK1A to shuttle between the nucleus and the cytoplasm, exerting its role in different cellular compartments (see next section). Table I2 presents a list of the proteins phosphorylated by DYRK1A, although it has to be mentioned that for most of them evidences for phosphorylation only exist *in vitro*. In some cases, DYRK1A plays its role through the interaction with other proteins, with a complete independence on the kinase activity. As examples, DYRK1A induces the transcriptional activation of FOXO1a (a member of the forkhead family of transcription factors) (von Groote-Bidlingmaier et al., 2003), potentiates transcription mediated by the chromatin remodelling factor Arip4 (Sitz et al., 2004) and potentiates nerve-growth factor (NGF)-mediated neuronal differentiation by up-regulating the Ras/MAP kinase signalling pathway (Kelly and Rahmani, 2005).

Table I2: Substrates and/or interactors of DYRK1A

Protein	Interaction	Phosphorylation	Localization	Reference
14-3-3β	+	-	cytoplasm	Alvarez et al., 2007
α-synuclein	+	GAG**S**^{87}I	cytoplasm	Kim et al., 2006
Amphiphysin 1	n.d.	RPR**S**^{293}P	cytoplasm	Murakami et al., 2006
APP	n.d.	AAV**T**^{668}P	cytoplasm	Ryoo et al., 2007
B-Raf	+	n.d.	cytoplasm	Kelly and Rahmani, 2005
Caspase 9	+	RPE**T**^{125}P	cytoplasm	Laguna, 2008
Dynamin 1	-	RPE**S**^{857}P	cytoplasm	Chen-Wang et al., 2002; Huang et al., 2004
eIF2ε	n.d.	RAG**S**^{539}P	cytoplasm	Woods et al., 2001
Glycogen synthase	n.d.	RPA**S**^{640}V	cytoplasm	Skurat et al., 2004
Han11	+	n.d.	n.d.	Skurat et al., 2004
HIP1	+	+	cytoplasm	Kang et al., 2005
MEK1	+	n.d.	cytoplasm	Kelly and Rahmani, 2005
PAHX-AP1	+	n.d.	cytoplasm	Bescond and Rahmani, 2005
Ras	+	n.d.	cytoplasm	Kelly and Rahmani, 2005
Sprouty-2	+	RPS**T**$'^{5}$Q	cytoplasm	Aranda et al., 2008
Synaptojanin 1	n.d.	+	cytoplasm	Adayev et al., 2006
Tau	n.d.	RSR**T**^{212}P	cytoplasm	Woods et al., 2001
Arip4	+	n.d.	nucleus	Sitz et al., 2004

c-jun	n.d.	PPL**S**^{243}P	nucleus	Morton et al., 2003
CREB	+	RRP**S**^{133}Y	nucleus	Yang et al., 2001
Cyclin L2	+	SGF**S**^{330}P LVE**S**^{338}P KAD**S**^{369}P	nucleus	De Graaf et al., 2004
FOXO1a	+	n.d.	nucleus	Von Groote-Bidlingmaier et al., 2003
FKHR	+	GRL**S**^{329}P	nucleus	Woods et al., 2001
Gli1	n.d.	+	nucleus	Mao et al., 2002
Histone H3	n.d.	RPG**T**^{45}V	nucleus	Himpel et al., 2000
HPV16E7	+	GD**T**^{5}P**T**^{7}L	nucleus	Liang et al., 2008
NF-AT1	n.d.	RSR**S**^{270}P RSP**S**^{272}P	nucleus/cytoplasm	Gwack et al., 2006
NF-AT3	n.d.	PRA**S**^{215}P	nucleus/cytoplasm	Arron et al., 2006
SF2/ASF	+	RSY**S**^{227}P RG**S**^{234}P RY**S**^{238}P	nucleus	Shi et al., 2008
SF3b1	n.d.	LTA**T**^{434}P	nucleus	De Graaf et al., 2006
SFRS4 (SRp40)	n.d.	+	nucleus	De Graaf et al., 2004
SFRS5	n.d.	+	nucleus	De Graaf et al., 2004
Stat3	n.d.	LPM**S**^{727}P	nucleus	Matsuo et al., 2001
SWAP2	n.d.	+	nucleus	De Graaf et al., 2004
TRA2β1	n.d.	+.	nucleus	De Graaf et al., 2004

n.d.: not determined. Cytosolic proteins are shown in green; nuclear proteins are shown in blue

1.3.3 The subcellular localization

Several studies describe the intracellular distribution of DYRK1A *in vivo*. In the mouse brain, DYRK1A is localized both in the nucleus and the cytoplasm of neurons and glial cells (Hammerle et al., 2008; Marti et al., 2003). In the chicken, a switch from the nucleus to the cytoplasm has been described during Purkinje cells differentiation (Hammerle et al., 2003). In human brains, DYRK1A is distributed in the nucleus and the cytoplasm of neurons, including neuronal processes and synapses, but only in the cytoplasm of astrocytes, ependymal and endothelial cells (Wegiel et al., 2004). Within the nucleus, DYRK1A accumulates in nuclear spots in chicken, mouse and human neurons (Hammerle et al., 2003; Hammerle et al., 2008; Marti et al., 2003; Wegiel et al., 2004). The low incidence of cells with nuclear DYRK1A *in vivo* supports a model of transient nuclear translocation in particular cell contexts and in response to certain signals.

In vitro studies have shown that the exogenous expression of DYRK1A mainly results in its nuclear accumulation in several cell types (Alvarez et al., 2003; Becker et al., 1998; Mao et al., 2002; Sitz et al., 2004; Song et al., 1997), indicating that it may be the default localization when distribution and targeting mechanisms are saturated. Moreover, an

accurate analysis has revealed that nuclear DYRK1A has a dot-like pattern of distribution, compatible with a subnuclear compartment known as speckles or splicing factor compartment (SFC). In fact, co-localization with SFC markers has been shown by our group (Alvarez et al., 2003). The group has also shown that the subcellular localization of DYRK1A is mediated by two predicted NLSs (Alvarez et al., 2003; Becker et al., 1998), which are responsible of the nuclear targeting, and a NES (Alvarez, 2004) which may regulate the cytosolic accumulation. Moreover, the localization in nuclear speckles is mediated by the repeat of thirteen histidines localized in the C-terminal region of the protein (Alvarez et al., 2003).

2. COMPARTMENTALIZATION IN THE EUKARYOTIC NUCLEUS

The mammalian cell nucleus is a complex but highly organized organelle, which contains numerous morphological subcompartments that form and maintain themselves in the absence of membranes. When observed at either light or electron microscopes, the interior of the nucleus is clearly non homogeneous. In fact, most macromolecular constituents of the cell nucleus are compartmentalized. For example, each chromosome occupies a nuclear space or territory in the interphase nucleus, and specific DNA sequences within each chromosome are organized as either euchromatin or heterochromatin (reviewed in Cremer and Cremer, 2001). Proteins and ribonucleoprotein complexes involved in mRNA metabolism occupy the interchromatin space of the nucleoplasm and are generally exclude from the nucleolus, while components of the ribosome biogenesis pathways are predominantly confined to nucleoli (reviewed in Boisvert et al., 2007). Additional compartments identified in the interchromatin nucleoplasmic space include nuclear speckles (Misteli et al., 1997), Cajal bodies (Gall, 2000), gems (Liu and Dreyfuss, 1996), PML nuclear bodies (Zhong et al., 2000) and a rapidly growing family of small dot-like nuclear bodies (Matera, 1999), composed by Gomafu granules (Sone et al., 2007), Omega speckles (Prasanth et al., 2000), Sumo bodies (reviewed in Heun, 2007), Sam68 nuclear bodies (Chen et al., 1999), clastosomes or proteasome bodies (Lafarga et al., 2002), FBX025 bodies (Manfiolli et al., 2008), cleavage bodies (Schul et al., 1996) and DDX1 bodies (Bleoo et al., 2001). Specific marker proteins identify all the mentioned domains.

Due to the fact that part of this thesis work has been developed around nuclear speckles, the structural and functional properties of this subnuclear compartment will be described in the next Section.

2.1 Nuclear speckles

Nuclear speckles or SFC (for Splicing Factor Compartment) are defined as 20-50 irregular dots that function as compartments of storage, assembly and modification of components of the RNA splicing machinery. Speckles are formed in regions that contain little or no DNA, although they tend to localize close to highly active transcription sites. By electron microscopy, two different morphological structures can be observed: i) interchromatin granule clusters (IGC), the equivalent structure to speckles seen by fluorescence microscopy. Each IGC is composed of a series of particles of 20-25 nm in diameter that seem to be connected in places by a thin fibril and ii) perichromatin fibrils, 3-5 nm fibrillar structures localized both at the periphery of IGCs and in nucleoplasmic regions away from them, which are the sites of nascent pre-mRNAs (reviewed in Lamond and Spector, 2003).

Diverse proteomic analyses have allowed defining the protein composition of nuclear speckles. Their main components are splicing factors, both snRNPs and serine/arginine-rich (SR) proteins, which are characterized by the presence of dipeptide RS repeats in the C-terminal region of the proteins. Other proteins found in this subnuclear compartment are transcription factors, 3'-RNA processing factors, translation factors, ribosomal proteins, structural proteins and several kinases and phosphatases. In addition, speckles also contain a Ser^2-phosphorylated form of the RNA polymerase II, lipids such as phosphatidylinositol 4,5-biphosphate and a population of poly(A)$^+$ RNAs of unknown function (Lamond and Spector, 2003; Mintz et al., 1999; Saitoh et al., 2004) (Figure I9).

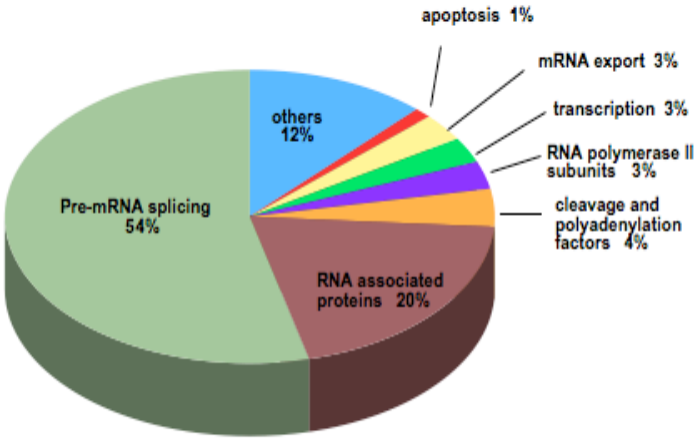


Figure I9: Profile of the composition analysis of nuclear speckles. From the 146 resultant proteins, 81% are involved in activities related to RNA metabolism (adapted from Saitoh et al., 2004).

Speckles are dynamic structures and their size, shape and number vary, both between different cell types and within a cell type, according to the levels of gene expression and in response to metabolic and environmental signals (Janicki et al., 2004; Misteli et al., 1997). When transcription is halted, either by the use of inhibitors or as a result of heat shock, splicing factors accumulate predominantly in enlarged and rounded speckles. On the contrary, when the expression of intron-containing genes increases or when the transcription levels are high during viral infection, the accumulation of splicing factors in speckles is reduced and they redistribute to nucleoplasmic transcription sites (reviewed in Lamond and Spector, 2003). Individual SFC components can therefore shuttle continually into and out the speckle and, at least for the splicing factor SF2/ASF, the exchange rate is known to be very rapid (Kruhlak et al., 2000; Phair and Misteli, 2000). However, the reduction in their movement rates when compared to freely diffusing molecules may result from numerous transient interactions both with large complexes and/or nuclear components, either within or outside speckles. This reasoning fits with the proposal that speckles form through a process of self-organization, in which protein-protein and protein-RNA interactions determine the

composition and the morphological structure, which is finally governed by steady-state dynamics (Misteli, 2001; Phair and Misteli, 2000).

The compartmentalization of splicing factors in the SFC confers several advantages: i) it allows to control the concentration of splicing factors in the nucleoplasm, so that optimize *in vivo* splicing efficiency; ii) it may act as a site of recycling and reactivation of splicing factors through reversible phosphorylation; iii) it regulates the relative amount of various splicing factors, modulating splice-site usage, and iv) it contributes to the functional coupling of transcription and splicing, through the assembly of complexes containing components of the transcription and splicing machineries (Misteli, 2000).

The exchange of factors between speckles and the nucleoplasmic pool is regulated by cycles of phosphorylation and dephosphorylation (Figure I10). Phosphorylation of the RS domain of SR splicing factors by several kinases has been shown to be necessary for the recruitment of SR proteins from nuclear speckles to sites of transcription/pre-mRNA processing, and for their association with the forming spliceosome (Misteli et al., 1998). In the case of the splicing factor SF2/ASF, there is an interplay between SRPK and CLK/STY kinases, since the sequential phosphorylation of a docking motif in SF2/ASF regulates the subcellular localization of the factor (Ngo et al., 2005). Interestingly, SRPKs, CLKs and DYRK kinases belong to the same branch of the CMGC group of protein kinases, which could point to some similarities in their functional roles. Moreover, the over-expression of CLK/STY, SRPK1 and DYRK1A kinases result in a complete redistribution of splicing factors from speckles to the diffuse nuclear pool (Alvarez et al., 2003; Colwill et al., 1996; Gui et al., 1994; Kuroyanagi et al., 1998; Misteli, 1998; Wang et al., 1998). On the other hand, the progression of the splicing reaction and the reenrolment of splicing factors in the SFC requires several dephosphorylation steps by speckle resident phosphatases (Bunce et al., 2008; Misteli, 1996; Yuan et al., 1998). This basal exchange rate of factors coupled with a mechanism to modulate it ensures that the required factors, in the correct phosphorylation state, are available to pre-mRNA transcripts at the sites of transcription and that the factors are sequestered out of the soluble nuclear pool into nuclear speckles when not functionally needed.

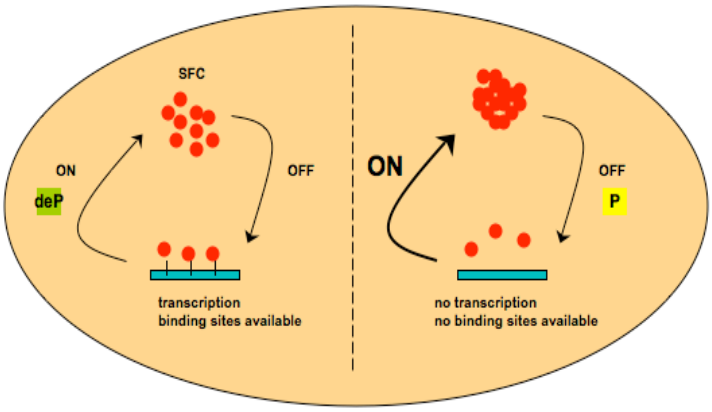


Figure I10: Nuclear speckles are in constant flux. The morphological appearance of the SFC is dependent on the on/off rates of its components, which are mainly regulated by phosphorylation (yellow square)/dephosphorylation (green square) cycles. Splicing factors continuously dissociate and reassociate with SFCs. After dissociation, they search for unspliced RNA to bind to, establishing a steady-state equilibrium. If no unspliced transcripts are available, the splicing factors spend less time outside of the compartment before they return and the net influx of proteins is increased. Splicing factors accumulate in the compartment, and it appears morphologically more round (adapted from Dundr et al., 2001).

2.2 Other subnuclear organelles

The **nucleolus** was the first subnuclear organelle to be identified. Each nucleus contains 1-5 nucleolus of 5-10 mm in diameter, formed on chromosomal loci (the so-called nucleolar organizer regions or NORs) that contain the genes coding for rRNA (Figure I11). The nucleolar structure is clearly differentiated into three regions: the fibrillar centers (the sites of rRNA genes transcription by the RNA polymerase I), the dense fibrillar component (where the rRNA is maturated) and the granular region (where the pre-ribosomal particles are assembled into the ribosome). The proteomic analysis of the nucleolus has identified a group of proteins that include RNA modifiying enzymes, helicases, translation factors, chaperones, kinases, nucleotide-binding proteins, ribonucleases and snoRNPs, among others (Andersen et al., 2002; Coute et al., 2006). The main function of the nucleolus is to act as the hallmark of rRNA synthesis, rRNA processing and assembly of ribosomal subunits (Spector, 1993). Other roles have been more recently proposed, such as the maturation of some RNAs, the sequestration of regulatory molecules, the maturation of the telomerase and some RNPs, the assembly of RNPs and the nuclear export (reviewed in Handwerger and Gall, 2006).

PML bodies (Figure I11), also known as ND10, PODs (PML oncogenic domains) or Kr bodies, are enriched in the PML protein, a tumor suppressor originally identified in acute promyelocytic leukaemia (Zhong et al., 2000). PML bodies are suggested to play a dual role in transcriptional regulation, both by sequestering gene regulatory proteins and by modulating the availability or activation status of transcription factors. In fact, several transcriptional regulators colocalize with PML in nuclear bodies, including Sp100, Daxx, pRB, p53, among others. PML bodies are associated to other functional roles such as the

mediation of DNA-damage response and the regulation of cellular processes like apoptosis and senescence (reviewed in Bernardi and Pandolfi, 2007).

Cajal bodies (Figure I11) are organelles that occur both free in the nucleoplasm and physically associated with histone and snRNA gene loci. They are highly dynamic structures that may play a role in snRNP transport, maturation or both and in the macromolecular assembly of spliceosomal subcomplexes (reviewed in Cioce and Lamond, 2005; Handwerger and Gall, 2006). Cajal bodies contain high concentrations of spliceosomal snRNPs, such as U7 snRNP, as well as several nucleolar antigens, including p80 coilin, fibrillarin, NOPP140 and U3 snoRNP.

Gems (Gemini of Cajal bodies) are coincident with or adjacent to Cajal bodies, depending upon the cell line examined (Figure I11). Gems have been characterized by the presence of the survival of motor neurons (SMN) protein and an associated factor, Gemin2. Mutations or reduced levels of SMN protein caused spinal muscular atrophy (SMA), a severe inherited neurodegenerative disease (reviewed in Spector, 2001). The SMN protein interacts with the Sm class of snRNP proteins and with a cellular protein called SIP1. The SMN-SIP1 complex plays an essential role in spliceosomal snRNP biogenesis (Fischer et al., 1997).

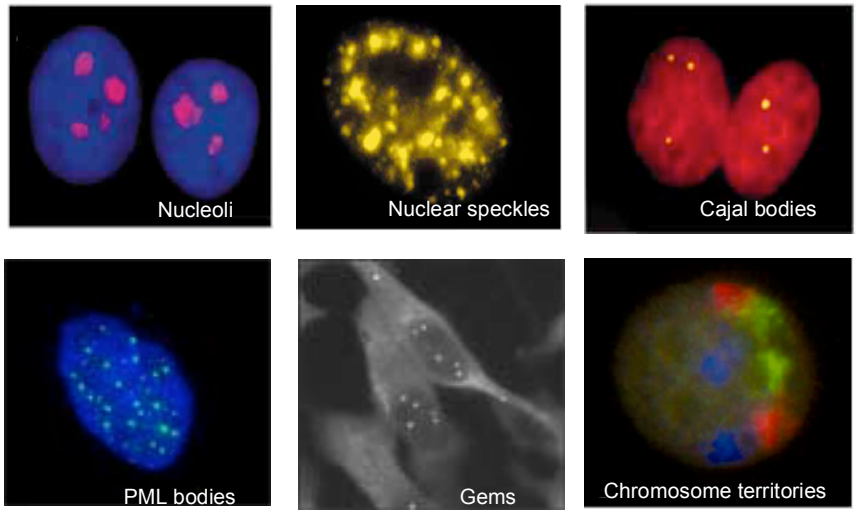


Figure I11: Fluorescent images of nuclear bodies. Nucleoli, nuclear speckles and Cajal bodies images adapted from (Misteli, 2000); PML bodies and Gems images adapted from the web page of A. Lammond lab (www.lammondlab.com); Chromosome territories of an interphase nuclei image adapted from (Dundr et al., 2001).

PcG bodies contain polycomb group proteins (i.e. RING1, BMI1 and hPc2) and have been observed associated with pericentromeric heterochromatin. These domains vary in number, size and protein composition and it is currently unclear whether they are storage compartments or are directly involved in silencing (Saurin et al., 1998).

The **perinucleolar compartment** (PNC) is associated with the surface of nucleoli and is though to play a role in RNA metabolism. It contains a series of small RNAs transcribed by RNA polymerase III and several RNA-binding proteins, including poly-pyrimidine-tract-binding

(PTB) protein. The perinucleolar compartmentC is found predominantly in transformed and cancer cells, but not in primary cells (reviewed in Kopp and Huang, 2005).

Paraspeckles are 10-20 discrete bodies in the interchromatin nucleoplasmic space that are often located adjacent to splicing speckles. They are composed by, at least, three RNA-binding protein components: PSP1, PSP2 and p54/nrb. The three paraspeckle proteins relocalise quantitatively to unique cap structures at the nucleolar periphery when transcription is inhibited (Fox et al., 2002).

Clastosomes are highly enriched in components of the ubiquitin-proteasome pathway and contain a high concentration of ubiquitin conjugates, the proteolytically active 20S core and the 19S regulatory complexes of the 26S proteasome, and protein substrates of the proteasome (reviewed in von Mikecz, 2006). Although detected in a variety of cell types, clastosomes are scarce under normal conditions; however, they become more abundant when proteasomal activity is stimulated. In contrast, clastosomes disappear when cells are treated with proteasome inhibitors. Protein substrates of the proteasome that are found concentrated in clastosomes include the short-lived transcription factors c-Fos and c-Jun, adenovirus E1A proteins, and the PML protein (Lafarga et al., 2002).

3. AMINO ACID REPEATS

Amino acid repeats are important for the evolution of proteins. Depending on their nature, they are broadly classified as perfect or mismatch repeats. The former class is subdivided into single amino acid repeats (reiteration of any single amino acid) and heteropeptide repeats (repeats with alternated different residues). The mismatch repeats consist of stretches with substituted conserved amino acids and form the most diverse class of repeats (Kalita et al., 2006). Over the past decade, the significant association between the expansion of specific homopopeptides, such as polyglutamine and polyalanine, and a disease phenotype has prompted many groups to further investigate them. The characteristics and functional consequences of the presence of amino acid repeats in proteins are discussed in this Section.

3.1 Structural and functional characteristics of single amino acid repeats

The presence of single amino acid repeats (SARs), also known as homopeptides or homopolymeric tracts, is a common feature of many eukaryotic proteins (Huntley and Golding, 2002). In the human genome, between 18-20% of proteins contain these repetitive sequences (Alba and Guigo, 2004; Karlin et al., 2002). Although most of them seem to be functionally neutral, recent evidence point to the possibility that repeats may play important functional or structural roles, as it will be discussed below.

Regarding their composition, SARs are generally less than 20 identical residues long and in mammals are primarily composed, in decreasing order, by Glu, Pro, Ala, Ser, Leu, Gly and Gln (Alba and Guigo, 2004; Faux et al., 2005; Huntley and Golding, 2002). The composition analysis of single amino acid repeat-containing proteins (SARPs) showed an overrepresentation of polar in comparison to hydrophobic tracts (Faux et al., 2005). Moreover, hydrophobic runs tended to be shorter, which may respond to the ability of hydrophobic SARPs to aggregate and form toxic fibrils (Oma et al., 2004). Within GENPEPT, 23% of all SARPs contain multiple repeats, again with a predominance of uncharged polar amino acids (Faux et al., 2005; Veitia, 2004). Moreover, these proteins show other unusual sequence features, such as multiple charge clusters (20-75 residues with significantly high specific charge content relatively to the charge composition of the whole protein) and significantly long uncharged segments.

Most proteins fold into only one particular conformation, determined by their amino acid sequence. In contrast, proteins with repeats frequently do not have unique stable 3D structures. It is broadly accepted that homopolymeric tracts form disordered structures (reviewed in Dunker et al., 2002; Kajava, 2001). This intrinsic disorder provides flexibility and

converts the repetitive region into a spacer element between individual folded domains, allowing a fine-tuning regulation of their movement relatively to each other and of the distance between adjacent domains. Regarding chemical modification processes, a side chain within a repetitive region facilitates substrate binding, because the disordered region can fold onto the modifying enzyme (Dunker et al., 2002). However, the main advantage of disordered regions is their ability to bind to multiple targets with high specificity and low affinity, an ideal property for transcription and signal transduction processes. For this reason, SARPs are usually found in association with large, multiprotein and/or nucleic acid complexes (Faux et al., 2005; Karlin et al., 2003), as represented in Figure I12.

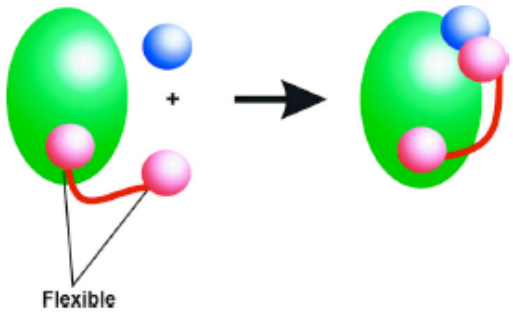


Figure I12: Schematic representation of a protein with amino acid repeats within a large multiprotein and/or nucleic acid complex. An example is shown where a two-domain protein (pink circles) within a macromolecular complex (green circle) functions via a flexible repeat (red) to recruit an additional binding partner (blue circle). Adapted from (Faux et al., 2005).

The main functions associated to homopolymeric tracts are the modulation of protein-protein interactions (Buchanan et al., 2004; Shimohata et al., 2000) and the regulation of gene transcriptional activity (Brown et al., 2005; Dunah et al., 2002; Friedman et al., 2007; Galant and Carroll, 2002; Gerber et al., 1994; Lanz et al., 1995).

With regard to the type of proteins that contain repeats, there is a clear bias towards transcription factors, kinases and developmental proteins (Alba and Guigo, 2004; Alba et al., 1999; Faux et al., 2005; Faux et al., 2007; Karlin et al., 2002; Karlin and Burge, 1996; Siwach et al., 2006) (Figure I13). These functional associations respond both to the positive contribution of the repeat to the function of the protein and to the permissiveness to incorporate types of SARs that cannot be accommodated in other types of proteins, such as structural or housekeeping proteins (Green and Wang, 1994; Karlin and Burge, 1996).

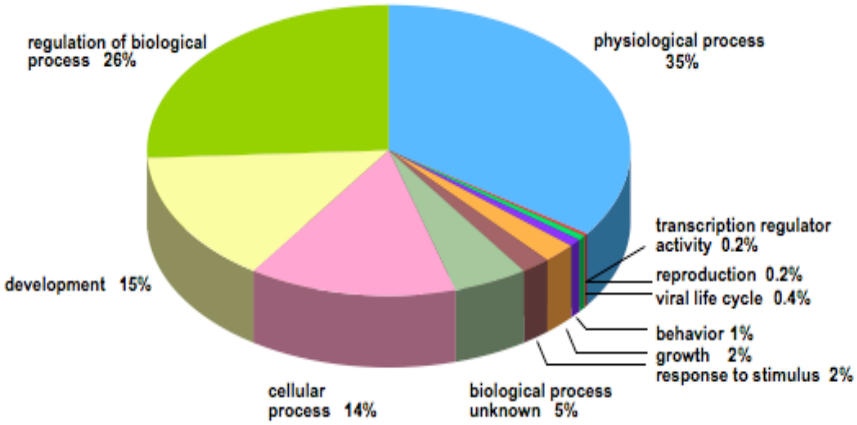


Figure I13: Distribution of biological processes as defined in Gene Ontology (GO) for SARPs where their repeats are conserved across human, mice and rat. The percentages reveal an overepresentation of proteins with roles related to physiological processes (35%), regulation of biological processes (26%) and development (15%) (adapted from Faux et al., 2007).

3.2 The appearance and evolution of amino acid repeats

SARs are originated by unequal crossing-over or replication errors resulting from the formation of unusual DNA secondary structures. Once a series of neighbouring codons end up being identical, small tandem repeats of a single codon would be then liable to rapid size change, with some undergoing expansion by trinucleotide slippage (reviewed in Pearson et al., 2005). At some point, this expansion would bring the repeat into a size range in which selection started to act due to deleterious effects on the phenotype, such as in the case of triplet repeat expansion diseases (see Section 3.3). If the repeat has acquired a function, and it is therefore not dispensable, its size will become frozen by the accumulation of point mutations, breaking up the pure codon repeats (while maintaining the amino acid repeat) and reducing the potential for slippage mutation and, therefore, evolutionary change (Hancock and Simon, 2005) (Figure I14).

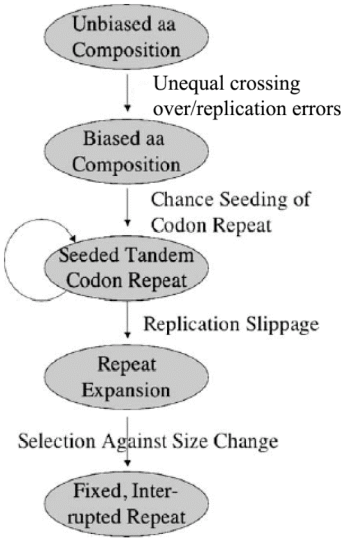


Figure I14: Schematic diagram of the evolution and fixation of amino acid repeats in proteins. The repetitive nature of trinucleotide repeats facilitates unequal crossing over and DNA replication slippage, which expand or contract the repeats in a short time span. Once the repeat appears, selective pressures at the amino acid level dictate the type, prevalence and length of the homopeptides (adapted from Hancock and Simon, 2005).

Most amino acid repeats occur in protein classes associated only with eukaryotes, more than prokaryotes or archaea, indicating that their formation is a relatively recent evolutionary event (Faux et al., 2005). The predominance of repeats in eukaryotic proteins respond to a greater need of protein-mediated signaling, regulation and cellular control, and highlights their importance in the reorganization of protein-protein interaction networks over the evolutionary time (reviewed in Hancock and Simon, 2005). Among eukaryotes, there are strong differences in the distribution of specific repeats, both at the length and at the composition level (Alba and Guigo, 2004; Faux et al., 2005; Kalita et al., 2006). However, the conservation of a given repeat in different species reveals its importance in the function of the host protein (Alba and Guigo, 2004). These conserved homopeptides lie within regions that are under stronger purifying selection (Faux et al., 2007; Mularoni et al., 2007) and tend to be encoded by mixed codon structures (Alba et al., 1999). It has been recently reported that slowly evolving proteins contain an unexpected large number of repeats, which tend to be conserved between organisms. As these proteins only rarely incorporate new repeat structures, many of their repeats are likely to be relatively old structures that have been preserved by selection (Mularoni et al., 2007), which again gives a clue about their functional importance. In contrast, non-conserved homopeptides are evolutionary labile repeats located in regions of proteins with relatively high rates of protein sequence changes and, since are encoded by pure codon repeats, they are formed most likely as a result of strand slippage in a recent fashion. This kind of repeats is mainly found in transcription factors and channel proteins (Alba et al., 1999; Hancock and Simon, 2005; Mularoni et al., 2007).

3.3 Variations in the length of amino acid tracts and their consequences

Although genetic variability of repeats provides a substrate for evolution (Kashi and King, 2006), uncontrolled expansions of such unstable regions within coding sequences have been extensively associated with several developmental and inherited neurodegenerative disorders (Gatchel and Zoghbi, 2005; Karlin et al., 2002), and with several types of cancer (Haberman et al., 2008), all of them referred as triplet expansion diseases (Table I3). In all these diseases, the context of the expanded repeat and the abundance, subcellular localization and interactions of the proteins and RNAs that are affected have key roles in disease-specific phenotypes (Gatchel and Zoghbi, 2005).

Repeat instability probably involves the formation of unusual DNA structures during DNA replication, repair and recombination (reviewed in Pearson et al., 2005). At the genomic level, the proposed consequences of hyperexpansion of DNA triplet repeats include loss or altered rate of transcription or translation, mRNA instability and aberrant DNA hairpin structures. At the protein level, the mechanisms suggested to underlie the pathogenic effects

of expanded tracts are the deregulation of transcriptional activity, either by loss-of-function or gain of abnormal function, where the protein behaves as a dominant negative (Friedman et al., 2007) and the formation of cytotoxic protein aggregates due to protein misfolding, exacerbated in the case of hydrophobic residues (Gatchel and Zoghbi, 2005; Menon and Pastore, 2006; Oma et al., 2004; Oma et al., 2005) (Table I3).

Table I3: Amino acid expansions and disease

Condition	Locus	Gene	Protein type	Expansion size
Hungtinton's disease	4p16.3	HD (hungtintin)	Transcription /intracellular trafficking	6-35Q> 36-121Q
Spinal and bulbar muscular atrophy	Xq11.2-q12	AR (androgen receptor)	Transcription factor	9-36Q> 38-62Q
Dentatorubral pallidoluysian atrophy	12p13.31	DRPLA (atrophin 1)	Unknown	6-35Q> 49-88Q
Spinocerebellar ataxia 1	6p23	ATXN1 (ataxin 1)	RNA binding?	6-44Q> 36-63Q
Spinocerebellar ataxia 2	12q24.1	ATXN2 (ataxin 2)	Actin-filament formation	15-31Q> 36-63Q
Spinocerebellar ataxia 3 (Machado-Joseph disease)	14q24.3-q32.2	ATXN3 (ataxin 3)	De-ubiquitylating activity	12-40Q> 55-84Q
Spinocerebellar ataxia 6	19p13.2-p13.1	CACNA1A	Calcium channel	4-18Q> 21-33Q
Spinocerebellar ataxia 7	3p21.1-p12	ATXN7 (ataxin 7)	SAGA complex	4-35Q> 37-306Q
Spinocerebellar ataxia 17	6q27	SCA17 (TBP)	Transcription factor	25-42Q> 47-63Q
Synpolydactyly type II	2q31.1	HOXD13	Transcription factor	15A> 22-29A
Cleidocranial dysplasia	6p21	RUNX2	Transcription factor	17A> 27A
Oculopharyngeal muscular dystrophy	14q11.2-q13	PABPN1	PolyA binding protein	10A> 11-17A
Holoprosencephaly	13q32	ZIC2	Transcription factor	15A> 25A
Blepharophimosis, ptosis and apicanthus inversus	3q23	FOXL2	Transcription factor	14A> 22-24A
Mental retardation X-linked	Xq26-q27	SOX3	Transcription factor	15A> 26A
Infantile spasm syndrome	Xp22.1-p21.3	ARX	Transcription factor	16A 18 or 23A
Partington syndrome	Xp22.1-p21.3	ARX	Transcription factor	12A> 20A
Congenital central hypoventilation syndrome	4p12	PHOX2B	Transcription factor	20A> 25-29A
Hand-foot-genital syndrome	7p15-p14	HOXA13	Transcription factor	18A> 24 or 26A
Pseudoachondroplasy	19p13.1	COMP	Extracellular matrix	5D> 7D
Multiple epiphyseal dysplasia	19p13.1	COMP	Extracellular matrix	5D> 6D

The expansion of polyglutamine (Gln) tracts results in a family of neurodegenerative disorders known as polyglutamine diseases, which are caused by a gain-of-function mechanism. These prototypical protein-misfolding disorders include Huntington disease and several types of atrophies and spinocerebellar ataxias (reviewed in Gatchel and Zoghbi, 2005; Shao and Diamond, 2007). Each of these disorders is characterized by the accumulation of the mutant protein in large intranuclear inclusions, due to the aberrant

interaction with normal cellular proteins, to inappropriate multimerization or to the formation of polar zippers between monomers (Figure I15A) (Karlin and Burge, 1996; Perutz et al., 1994). The principal toxic effects of the aberrantly folded protein may include alterations in transcription, metabolism or impairment of the proteasome and stress response pathways, among others (reviewed in Gatchel and Zoghbi, 2005; Shao and Diamond, 2007) (Figure I15B).

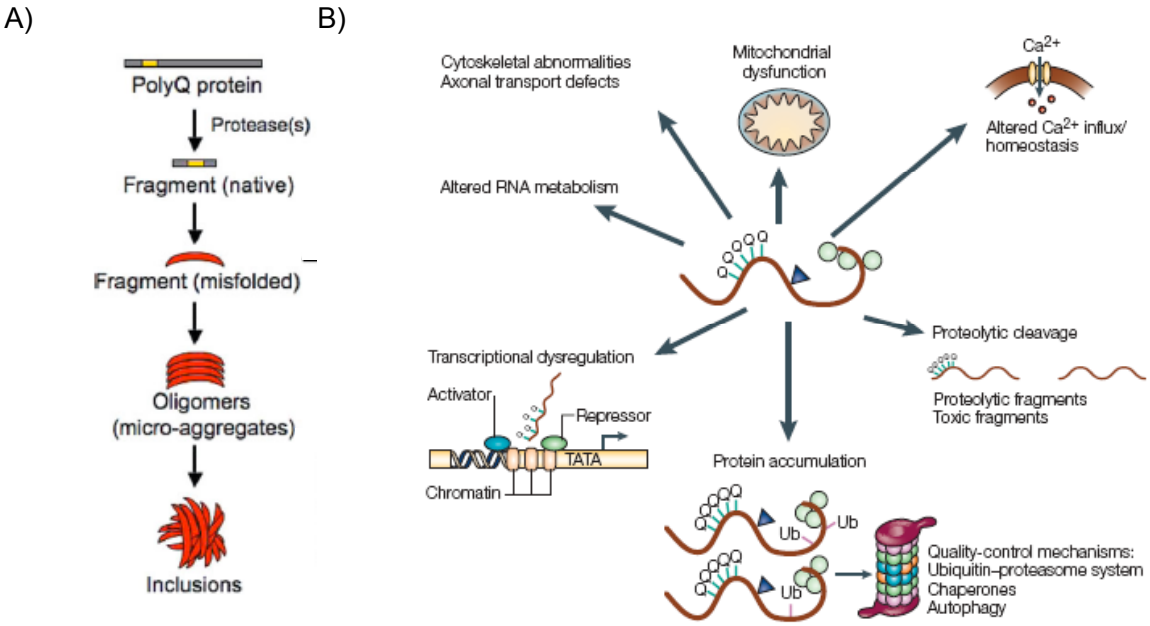


Figure I15: Pathogenesis of polyglutamine diseases. A) Many, but not all, polyglutamine diseases appear to be initiated by proteolytic cleavage to generate a toxic fragment. The expanded polyGln tract allows transition into a distinct conformation that may cause toxicity in several ways, as a monomer or by self-association to form toxic oligomers. The oligomers can assemble into larger aggregated species and ultimately are deposited in macromolecular intracellular inclusions (adapted from (Shao and Diamond, 2007). B) Expanded polyGln proteins might mediate pathogenesis through a range of mechanisms (adapted from Gatchel and Zoghbi, 2005).

Alanine (Ala) repeats represent the 16% of homopolymeric tracts and are mainly found in developmental transcription factors (Karlin et al., 2002). The overall consequence is an altered expression of downstream genes, and this, in turn, results in abnormal development (reviewed in Brown and Brown, 2004). In fact, polyAla expansions are associated to several human developmental disorders, including mental retardation and malformations of the brain, digits and other structures (reviewed in Brown and Brown, 2004). Present knowledge suggests that Ala tract expansions generally arise through unequal allelic homologous recombination, as opposed to replication slippage due to the imperfection of the trinucleotide repeats. In contrast, polyAla contractions have not been involved with certainty in human diseases. However, considering that many proteins with polyAla stretches interact with each other, the contraction of the tract in one protein may well modify the transcriptional

activity of a protein complex in both space- and time-dependant ways (reviewed in Amiel et al., 2004).

Aspartate (Asp) hyperexpansions have been related to two types of dysplasia and osteoarthritis (Delot et al., 1999; Kizawa et al., 2005). In the case of Asp repeats, both in-frame deletion of one codon and duplication of two codons within the perfect five trinucleotide repeats of the COMP gene are known to cause pseudoachondroplasia (Delot et al., 1999; Hecht et al., 1995). In this case, the unique feature is that both expansion and shortening of the repeat cause the same disease.

Although no other amino acid expansions have been described thus far, the tendency of the SARs to form aggregates (Oma et al., 2004) suggests that more expansion-related diseases will be discovered in the future.

OBJECTIVES

The general objective of this thesis has been to identify novel roles for the protein kinase DYRK1A within the cell nucleus. The importance of the His repeat in determining the subnuclear localization of the kinase has been used as a tool for approaching the problem. The main objective can be divided in the following subobjectives, with some of them appearing during the development of the thesis work.

I. To characterize the His repeat as a bona-fide targeting sequence that mediates the accumulation of DYRK1A in nuclear speckles.

II. To study the role of His repeats in the subcellular localization of other human proteins.

III. To identify the molecule/s which mediate the accumulation of DYRK1A in nuclear speckles.

IV. To characterize the interaction of DYRK1A with the transcription factor Brn-3b.

MATERIALS AND METHODS

1. PLASMIDS

In this Section, the construction of the plasmids generated during the thesis work is described in detail. It also includes the description and source of other plasmids used in some experiments.

1.1 Backbone vectors

pEGFP-C1 (Clontech)
pCDNA-3 (Invitrogen)
pCDNA-HA (de la Luna et al., 1999)
pGEX-5X3 (Amersham)
pGEX-5X2 (Amersham)
pGEX-6P1 (Amersham)

1.2 Expression plasmids for His-containing proteins

Generation of the plasmids expressing the different enhanced green fluorescent protein (EGFP) fusion proteins was done by PCR amplification of each of the open reading frames, using IMAGE Consortium cDNA clones (Lennon et al., 1996) as templates and specific primers. The identification number of the IMAGE clones and the sequence of the primers used are listed in Table M1. All the IMAGE clones were purchased from the RZPD German Resource Center for Genome Research. The general cloning procedure was the same as the one described in 2.2.

Taq polymerase from the Expand High Fidelity PCR System (Roche) was used in the amplification reactions, except in the case of PCR reactions with the GC-RICH PCR System. Brn-3b (fragment 1), FOXG1B, MafB and YY1 PCR products were obtained using the GC-RICH PCR System (Roche). HAND1 cDNA was amplified with the PCRx Enhancer System (Invitrogen). In the case of GSH2 the PCR reaction was carried out with the Expand Long Template PCR System (Roche). The DNA fragments were subcloned into the appropriate restriction sites of pEGFP-C1. The cDNA of Brn-3b was amplified separately in two PCR products (fragment 1: BglII/ApaI and fragment 2: ApaI/EcoRI) that were ligated together into the BglII and EcoRI sites of pEGFP-C1.

The general PCR reaction procedure is described in 2.3.

Name	Accession Nº	IMAGE clone ID	NIH gene ID	Forward primer	Reverse primer
CBX4ΔPB	EU439707	5493993	NM_003655	5'-**AGATCT**ATGGAGCTGCCAGC-3'	5'-CGGCTACACCGTCACGTAC-3'
DLX2	BC032558	5562689	NM_004405	5'-**AGATCT**ATGACTGGAGTCTTTG-3'	5'-TCTCCCTGGGGTTAGAAAATC-3'
FAM76A	BC025768	5211010	NM_152660	5'-**AGATCT**ATGGCGGCGCTCTAC-3'	5'-GTCTGTCATGGAGAGGTTATAG-3'
FAM76B	BC028727	4824833	NM_144664	5'-**AGATCT**ATGGCGGCCTCGG-3'	5'-CACATTACATACTCCTATCTCC
FOXG1B	BC050072	5284335	NM_005249	5'-**AGATCT**TGGGTGATGCTGGAC-3'	5'-TCCCAGGGATGTTAATGTATT-3'
GSH2	EU596451	2103455	NM_133267	5'-**GGATCC**ATGTCGCGCTCCTTCTATGTCGACTCG-3'	5'-GCCGCGGCCGAGCCAGGCTGCTGGGGCTGA-3'
HAND1	BC021190	3162118	NM_004821	5'-**GGATCC**ATGAACCTCGTGGG-3'	5'-CTCGGCTCACTGGTTTAACTC-3'
HOXA1	BC032547	5537563	NM_005522	5'-**GGATCC**ATGGACAATGCAAG-3'	5'-AGCCGCCTCAGTGGGAGG-3'
HOXA9	BC006537	2987903	NM_152739	5'-**GGATCC**ATGGCCACCACTGG-3'	5'-GCCCAAATGGCATCACTCGTC-3'
LMO6	BC016856	3846372	NM_006150	5'-**GGATCC**ATGTTCGCGCGTGG-3'	5'-CTGCCTTCAAGCCACGATGC-3'
MafA	BE676631	3296459	NM_201589	5'-**AGATCT**GCCTACGAGGCTTTC-3'	5'-GGGGCCGCCCCGGCCC-3'
MafB	BC028098	5242606	NM_005461	5'-**AGATCT**ATGGGCCGCGGAGCTG-3'	5'-CGGCCACGACTCACAGAAAG-3'
MECP2	BC011612	3956518	NM_004992	5'-**GGATCC**ATGGTAGCTGGGATG-3'	5'-GTAAAGTCAGCTAACTCTCTC-3'
MEOX2	BC017021	5209130	NM_005924	5'-**GGATCC**ATGGAACACCCGCTC-3'	5'-ATCATAAGTGCGCATGCTCTGAG-3'
MEOX2ΔHB	BC017021	5209130	NM_005924	5'**GGATCC**ATGGAACACCCGCTC-3'	5'TTTGCTCTTTGGTAAATGCTGTC-3'
NLK	BC064663	6527673	NM_016231	5'-**AGATCT**ATGGCGGCTTACAATG-3'	5'-CACCATCACTCCCACACCAG-3'
ONECUT1	EU532019	4372264	NM_004498	5'-**AGATCT**CCCACCATGACCATG-3'	5'-CTTCCTTCATGCTTTGGTAC-3'
OTX1	BC007621	3355563	NM_014562	5'-**AGATCT**GTTAGCATGATGTCT-3'	5'-ATTCCTGGGCTCACAAGACC-3'
OTX1ΔHB	BC007621	3355563	NM_014562	5'-**AGATCT**GGCCACCATCATCAC-3'	5'-ATTCCTGGGCTCACAAGACC-3'
PLK2	BC013879	3831747	NM_006622	5'-**GGATCC**ATGGAGCTTTTGCGG-3'	5'-**GGATCC**ATTCGAAAAGTCTTTCAGTTACATC-3'
POU4F2/Brn-3b	EU439706	DKFZp434P094Q	NM_004575	(1) 5'-**AGATCT**TCCCTGAACAGCAAG-3' (2) 5'-TCCCGGGCTGGCCCTG-3'	(1) 5'-TGGACACCACAGCGCCG-3' (2) 5'-CTTCTAAATGCCGGCGGAATA-3'
POU4F3/Brn-3c	BC112207	8327675	NM_002700	5'-**AGATCT**ATGATGGCCATGAACTC-3'	5'-TCAGTGGACAGCCGAATACTTC-3'
YY1	BC065366	5815774	NM_003403	5'-**AGATCT**TCAGCCATGGCCTCG-3'	5'-TTTCACTGGTTGTTTTTGGCC-3'
ZIC-3	EU532020	664181	NM_003413	5'-**GGATCC**ATGACGATGCTCCTG-3'	5'-TGTGTTTGTCCTCAGACGTACC-3'

Table M1: IMAGE clone number and primers used for the amplification of proteins with His repeats. In bold it is indicated the sequence recognized by the restriction enzyme. Brn-3b was amplified separately in two cDNA fragments (1 and 2), and afterwards cloned as a fusion of both parts into the GFP vector.

1.3 Plasmids expressing GFP fused to polyHis

To obtain plasmids expressing 5xHis, 6xHis, 7xHis, 8xHis and 9xHis protein segments fused to GFP, double stranded oligonucleotides were annealed and ligated into the BglII and EcoRI sites of the pEGFP-C1 expression vector. The oligonucleotides used were as follows (restriction sites appear in bold):

5xHis forward: 5'- **AGATCT**CACCATCACCATCACTAG**GAATTC**-3'

5xHis reverse: 5'- **GAATTC**CTAGTGATGGTGATGGTG**AGATCT**-3'

6xHis forward: 5'- **AGATCT**CACCATCACCATCACCATTAG**GAATTC**-3'

6xHis reverse: 5'- **GAATTC**CTAGTGATGGTGATGGTGATG**AGATCT**-3'

7xHis forward: 5'- **AGATCT**CACCATCACCATCACCATCACTAG**GAATTC**-3'

7xHis reverse: 5'- **GAATTC**CTAGTGATGGTGATGGTGATGGTG**AGATCT**-3'

8xHis forward: 5'- **AGATCT**CACCATCACCATCACCATCACCATTAG**GAATTC**-3'

8xHis reverse: 5'- **GAATTC**CTAGTGATGGTGATGGTGATGGTGATG**AGATCT**-3'

9xHis forward: 5'- **AGATCT**CACCATCACCATCACCATCACCATCACTAG**GAATTC**-3'

9xHis reverse: 5'- **GAATTC**CTAGTGATGGTGATGGTGATGGTGATGGTG**AGATCT**-3'

1.4 Brn-3b expression plasmids

- pFlag-Brn3b: human Brn-3b open reading frame was amplified separately in two PCR products (fragment 1: BglII/ApaI and fragment 2: ApaI/NotI) that were ligated together into the BglII and NotI sites of pCDNA-3 with a Flag tag.

- pFlag-Brn3bΔHis: generated by site-directed mutagenesis with primer 5'-TCGGTGCCCATCTCGCACCCTTGCGCGTTGGCGGGCACGCAGGCGCTGGAGGGCGAGCTGCTGGAGCACCTG-3' (deletion 172-186) on pFlag-Brn3b.

- pEGFP-Brn3b: described in section 1.2.

- pGEX-Brn3b: obtained by digestion of Flag-Brn3b with EcoRI/NotI and ligated into the EcoRI/NotI sites of pGEX-6P1. The fusion protein also included the Flag epitope.

- pGEX-Brn3b/1-215: Brn-3b fragment 1 (see Section 1.2) was digested from pGEM-T Easy vector with EcoRI and ligated into the EcoRI site of pGEX-5X3.

- pGEX-Brn3b/215-409: Brn-3b fragment 2 (see Section 1.2) was digested from pGEM-T Easy vector with EcoRI and ligated into the EcoRI site of pGEX-5X2.

- pGEX-Brn3b/1-105: generated by incorporating a STOP codon in pGEX-Brn3b/1-215 by site-directed mutagenesis with primer 5'-CCGAGCAATATATTCGGCGGGCTGGAT**T**AGAGTCGCTGGCCCGCGCCGAGGCTCTG-3'

- pGEX-Brn3b/105-215: generated by PCR amplification from pGEX-Brn3b/1-215 with primers 5'-AGTCTGCTGGCCCGCGCC-3' (forward) and 5'-TGGACACCACAGCGCCGT-3' (reverse) and cloned into the EcoRI site of pGEX-5X3.

- pGEX-Brn3bΔHis: generated by site-directed mutagenesis with primer 5'-TCGGTGCCCATCTCGCACCCTTGCGCGTTGGCGGGCACGCAGGCGCTGGAGGGCGAGCTGCTGGAGCACCTG-3' on pGEX-Brn3b.

- pGEX-Brn3b/S33A: generated by site-directed mutagenesis with primer 5'-ACTGCACAGCACC**G**CGCCGGGCTCCTCGGC-3' on pGEX-Brn3b.

- pGEX-Brn-3b/S46A: generated by site-directed mutagenesis with primer 5'-CCCTCGGCCAGC**G**CCCCCAGCAGCTCGAGC-3' on pGEX-Brn3b.

- pGEX-Brn-3b/T94A: generated by site-directed mutagenesis with primer 5'-GAGCCTGTCTTCCA**G**CCCCACCGAGCAATATATTC-3' on pGEX-Brn3b.

- pGEX-Brn-3b/S135A: generated by site-directed mutagenesis with primer 5'-ATCCACCCCACCAC**GC**CCCCTTCAAACCGG-3' on pGEX-Brn3b/1-215

- pGEX-Brn-3b/S198A: generated by site-directed mutagenesis with primer 5'-GCTGGAGCACCTGGCTCCCGGGCTGGCCCT-3' on pGEX-Brn3b/1-215

- pGEX-Brn-3b/S159A: generated by site-directed mutagenesis with primer 5'-GCCGCCTCTTCTTCAGCGGTGCCCATCTCGCACCCT -3' on pGEX-Brn3b/1-215

1.5 DYRK proteins expression plasmids

- pHA-DYRK1A (Alvarez et al., 2003): expression plasmid for human DYRK1A (754 amino acid isoform) with two HA-tags at its N-terminus.

- pHA-DYRK1A-K179R (Alvarez et al., 2003): expression plasmid for human DYRK1A with a mutation in the ATP-binding site that generates a catalytically inactive protein kinase.

- pHA-DYRK1A/1-522 (Alvarez et al., 2003): expression plasmid with a C-terminal deletion including the His repeat and the Ser-rich region.

- pHA-DYRK1A/1-474 (Alvarez et al., 2003): expression plasmid with a deletion in the C-terminal region including the PEST domain, the His repeats and the Ser-rich region.

- pHA-DYRK1A/1-377 (Alvarez et al., 2003): expression plasmid with a deletion in the C-terminal region including the NLS_2, the PEST domain, the His repeats and the Ser-rich region.

- pHA-DYRK1AΔN (Alvarez et al., 2003): expression plasmid encoding a mutant DYRK1A with a N-terminal deletion including the NLS_1 (147-754).

- pHA-DYRK1AΔHis (Alvarez, 2004): expression plasmid with an internal deletion comprising the His tract (590-616).

- pGFP-DYRK1A (Alvarez et al., 2003): DYRK1A fused to EGFP at its N-terminus.

- pGFP-DYRK1A/NLS_1mut (Alvarez, 2004): full-length DYRK1A with three mutations in the NLS_1 (K105/107A, R109A).

- pGFP-DYRK1AΔNLS$_1$ (generated by K. Arató in the lab): full-length DYRK1A with a deletion of the NLS$_1$ (residues 105-120).

- pGFP-DYRK1A/1-167 (Alvarez, 2004): fragment of the DYRK1A N-terminal region including the NLS$_1$.

- pHA-DYRK1B (generated by A. Raya in the lab): expression plasmid encoding human DYRK1B (629 amino acid form) with two HA-tags at its N-terminus.

- pGFP-DYRK1B (Becker and Joost, 1999): human DYRK1B (629 amino acid isoform) fused N-terminally to GFP.

- pGFP-DYRK1B/1-128 (Alvarez, 2004): DYRK1B N-terminal fragment comprising the NLS$_1$.

- pHA-DYRK2 (generated by A. Raya in the lab): human DYRK2 long isoform (600 amino acids) fused N-terminally to two HA-tags.

- pHA-DYRK3 (generated by A. Raya in the lab): human DYRK3 long isoform (587 amino acids) with two HA-tags in the N-terminal region.

- pHA-DYRK4 (generated by A. Raya in the lab): human DYRK4 (isoform of 519 amino acids) fused to two HA-tags in the N-terminal region.

1.6 Expression plasmids for TAPtag fusions

All the constructs were generated by subcloning GFP fusion proteins into the pCDNA3-TAPtag (pTAPtag) expression vector, a nice gift from J. Ortín (Centro Nacional de Biotecnología, CSIC, Madrid), as described below. Restriction sites (HindIII and NotI) appear in bold.

- TAPtag-GFP: the cDNA of GFP was amplified from pEGFP (Clontech) with primers 5'-**AAGCTT**GCTAGCGCTACCGGTC-3' and 5'-**GCGGCCGC**ACTTGTACAGCTC-3' and ligated into the HindIII and NotI restriction sites in pTAPtag.

- TAPtag-GFP-DYRK1A/378-616 (H+): the cDNA of GFP-DYRK1A/378-616 (Alvarez et al., 2003) was amplified with primers 5'-**AAGCTT**GCTAGCGCTACCGGTC-3' and 5'-**GCGGCCGC**GGTTACCCAAGGC-3' and ligated into the HindIII and NotI restriction sites in pTAPtag.

- TAPtag-GFP-DYRK1A/378-588 (H-): the cDNA of GFP-DYRK1A/378-588 (Alvarez et al., 2003) was amplified with primers 5'-**AAGCTT**GCTAGCGCTACCGGTC-3' and 5'-**GCGGCCGC**ATGCATTCTGTTG-3' and ligated into the HindIII and NotI restriction sites in pTAPtag.

1.7 Other plasmids

- pGL2-36 PRL (Gruber et al., 1997): control plasmid with a minimal promoter construct from the rat prolactin gene, containing a TATA element. It is used as control in reporter assays with Brn-3 factors.

- pGL2-3xBrn3a (Gruber et al., 1997): reporter plasmid with the minimal prolactin promoter plus 3 repeats of the Brn-3a recognition site.

- pFLAG-cyclin L2 (generated by A. Raya in the lab): human cyclin L2 fused N-terminally to the FLAG epitope.

- pMyc-cyclin T1 (Taube et al., 2002): expression plasmid for full-length human cyclin T1 with a MYC-tag at its N-terminus.

- pMyc-cyclin T1/1-590 (Taube et al., 2002): fragment of human cyclin T1 from residues 1 to 590 with a MYC-tag at its N-terminus.

- pGEX-CTD (Taube et al., 2002): bacterial expression plasmid encoding the 52 repeats of the heptapeptide YSPTSPS found in the human RNA polymerase II carboxi-terminal domain (CTD) fused N-terminally to gluthatione-S-transferase (GST).

- pYFP-SC35 (Prasanth et al., 2003), expression plasmid for a fusion yellow fluorescent protein (YFP) and human SC35, kindly provided by D. Spector (Cold Spring Harbor Laboratory, Cold Spring Harbor, USA).

-pEGFPC1-CBX4: expression plasmid for a fusion GFP and human CBX4, a kind gift from S. Aznar-Benitah (Center for Genomic Regulation-CRG, Barcelona, Spain).

- pEGFPN2-NKD2: expression plasmid for a fusion GFP and human NKD2, a nice gift from C. Li (Department of Medicine, Vanderbilt University Medical Center and Department of Veterans Affairs Medical Center, Nashville, USA)

- pTS-Brn3a (Fedtsova et al., 2008): bicistronic expression plasmid for a GFP and human Brn-3a.

-pEGFPC1-Brn3a, generated by digestion of pTS-Brn3a with EcoRI and BamHI and ligated into the pEGFP-C1 expression plasmid.

- pEGFPC1-HOXA1ΔHis, generated by site-directed mutagenesis with primer 5'-GCAGGGGGGTGCAGATCGGTTCGCCCCCCCAGCCGGCTACCTACCAGACT-3' (deletion 65-74) on pEGFPC1-HOXA1.

Plasmids pGEX-CTD, pMyc-cyclin T1 and pMyc-cyclin T1/1-590 were kindly provided by M. Peterlin (Howard Hughes Medical Institute, Deparments of Medicine, Microbiology and Immunology, University of California). The constructs pcDNA1-Brn3a, pGL2-36 PRL, pGL2-3xBrn3a and pTS-Brn3a were a gift from E. Turner (Department of Psychiatry, University of California, USA).

2. TECHNIQUES FOR DNA MANIPULATION

2.1. 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 Maxi Kit (Qiagen) following manufacturer's instructions.

DNA samples were sequenced at the Sequencing Service of the Universitat Pompeu Fabra using the Big-Dye terminator 3.0 reactive (Applied Biosystems) and the following PCR reaction (final volume of 10 μl):

- 2 μl Big-Dye 3.0 reactive
- 2 μl of purified DNA
- 1 μl of primer (3.2 μM)
- 5 μl H_2O

PCR sequencing conditions:

- Initial denaturation step: 1 min at 94ºC
- Denaturation step: 30 s at 94ºC
- Primer annealing step: 30 s at 50ºC } 28 cycles
- Extension step: 4 min at 60ºC

PCR reactions were purified using Sephadex G-50 (Pharmacia) columns. Briefly, 800 μl of Sephadex-G50 equilibrated in H_2O were added to a Centristep column and centrifuged for 1 min at 2,600*x*g. After discarding the supernatant, the column was washed with 10 μl of H_2O; the PCR reaction was added and recovered by centrifugation. Finally, DNAs were dried without heat in a Speed-vac for 10 min.

2.2 Cloning of a DNA fragment into an expression vector

In this section, all the procedures followed to clone a fragment of DNA into any of the expression vectors used are described. In general, these steps were followed:

1) PCR amplification of the cDNA
2) Ligation into a pGEM-T Easy vector (Promega)
3) Transformation into competent bacteria
4) Colony screening by PCR and sequencing of positive clones
5) Digestion with restriction enzymes, DNA fragment elution and ligation into the expression vector digested with the same enzymes
6) Transformation into competent bacteria
7) Colony screening by PCR

8) DNA extraction and confirmation by restriction enzyme digestion and/or sequencing of the fusions.

When possible, DNA fragments were prepared by restriction enzyme digestion; in those cases, steps 1-5 were omitted.

2.2.1 PCR amplification

PCR reactions were carried out in a total volume of 25 μl as follows:

- 2.5 μl commercial buffer 10x (with $MgCl_2$)
- 4 μl dNTPs (1.25 mM)
- 0.5 μl Expand High Fidelity PCR System (Roche)
- 1 μl forward primer (10 μM)
- 1 μl reverse primer (10 μM)
- 1 ng template DNA
- H_2O

PCR conditions:

- Initial denaturation step: 1 min at 94ºC
- Denaturation step: 30 s at 94ºC
- Primer annealing step: 30 s at 58ºC
- Extension step: 40 s at 72ºC } 30 cycles
- Final extending step: 10 min at 72ºC

2.2.2 Cloning into pGEM-T Easy Vector System (Promega)

PCR products (2 μl) were directly ligated to a pGEM-T Easy Vector following manufacturer instructions. Ligation reactions were incubated overnight at 4ºC.

2.2.3 Transformation into competent Escherichia coli (E. coli)

All DNA transformations were done using the *E. coli* strain XL1-blue F' (supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lac- F' [proAB+ lacIq lacZ DM15 Tn10 (tetr)]). Frozen competent cells were thawed and kept for 20 min on ice; 5 μl of each ligation were added to 50 μl of cells and kept for 20 min on ice. Cells were incubated for 90 s at 42ºC, put on ice for 5 min, resuspended in 200 μl of SOC media (20 g/l triptone, 5 g/l yeast extract, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) and incubated for 30-45 min in a shaker at 37ºC. Finally, cells were plated on LB (Luria Broth medium)-agarose plates containing antibiotics (ampicillin, 100 μg/ml); kanamycin, 30 μg/ml).

2.2.4 Colony screening by PCR amplification

Positive colonies were inoculated in 100 μl of LB+antibiotic and incubated for 1 h at 37ºC. PCR reactions were performed as described before in a final volume of 10 μl, using 1 μl of culture and specific primers. Finally, DNA from positive clones were prepared as described in section 2.1 and sequenced.

2.2.5 Preparation of DNA fragments and ligation into the expression vector

DNA (2-20 μg) was digested with restriction enzymes in a final volume of 100 μl for 2 h at the recommended temperature for each enzyme. Reactions were loaded in Low Melting Agarose (ECOGEN) gels (1-2%, depending of the size of the fragment)-1xTAE (2 mM Tris, 1 M acetic acid, 50 mM ethylendiamine tetracetic acid [EDTA]) - 0.1 μg/ml ethidium bromide (BrEt) gel and run in 1xTAE at 80 V. DNA was visualized with an UV transilluminator and the band of the expected molecular weight was cut. Bands were processed as follows: agarose was melted for 10 min at 65ºC, and NaCl plus EDTA were added to final concentrations of 100 mM and 10 mM, respectively. DNA was extracted by adding 1 vol of phenol (Sigma) and precipitated with 2 vol of 100% ethanol at –20ºC for 12-16 h.

To dephosphorylate the digested vector, DNA was treated with alkaline phosphatase (shrimp, Roche) for 1 h at 37ºC and purified by phenol-chloroform (1:1; Sigma) extraction.

Ligations were carried out using a molar ratio of 1:3 (vector:insert) in a final volume of 10 μl with 1 μl of T4 DNA ligase (Roche). The reactions were incubated overnight at 16ºC and 5 μl of each one were transformed into competent bacteria as described previously (2.1.3).

2.3 Site-directed mutagenesis

Site-directed mutagenesis has been used to generate point mutants of a given amino acid, to introduce stop codons or to generate small deletions using the commercial kit Quickchange Multi Site-Directed Mutagenesis Kit (Stratagene), following manufacturer's instructions. All the primers have been purchased with a phosphorylated 5' end (Bonsai Technologies), as required by the mutagenesis reaction. All the mutants have been checked by DNA sequencing of the complete sequence.

3. CELL CULTURE

The cell lines used in this work are as follows:

- U2-OS cells, epithelial cell line derived from human osteosarcoma

- HEK-293 cells, epithelial cell line derived from human embrionic kidney
- HEK-293T cells, epithelial cell line derived from human embrionic kidney transformed with SV40 T-Antigen
- HeLa cells, epithelial cell line derived from human cervix carcinoma.
- CV-1 cells, fibroblast cell line derived from African green monkey kidney

All these cell lines were supplied by the ATCC (American Type Culture Collection, http://www.atcc.org). Cells were cultured in Dulbecco´s modified Eagle medium (DMEM) (Invitrogen) containing 10% fetal bovine serum (FBS) (Invitrogen) and supplemented with antibiotics (100 units/ml penicillin and 100 µg/ml streptomycin) (Invitrogen) in 5% CO_2 atmosphere at 37ºC.

3.1 Cell transfection

Cells were transfected with 3 µg (35-mm plates), 6 µg (60-mm plates) or 20 µg (100-mm plates) of DNA by the calcium phosphate precipitation method (Graham and van der Eb., 1973), as described in Current Protocols in Molecular Biology (vol. 2, chapter 9). The optimal pH of the Hepes buffer for each cell line was tested by transfection with a GFP-expressing construct and quantification of fluorescent cells. The DNA-calcium phosphate precipitate was removed after 16 h of incubation by washing the cells with phosphate-buffered saline (PBS) and adding fresh DMEM medium. For transitory expression experiments, cells were harvested in PBS at 48 h after transfection and processed accordingly to the purpose of the experiment.

3.2 Generation of stable cell lines

Transfected cells were selected by incubating with the antibiotic G418 (Gibco). Selection of cells started at 48 h post-transfection by adding 750 µg/ml of G418 to the medium. Medium was replaced every 3 days, and the antibiotic concentration decreased to 500 µg/ml according to the level of cell death until a final concentration of 250 µg/ml. G418-resistant cells were trypsinized and plated at a dilution of one cell per well in a 96-well plate to obtain clones coming from a single cell, that were gradually amplified and tested for the expression of the fusion protein. Stable cell lines were maintained in DMEM, 10% FBS, penicillin/streptomycin and 250 µg/ml G418.

3.3 Cell treatments

Treatment of HeLa cells with the RNA polymerase II inhibitor alpha-amanitin (50 mg/ml; Sigma) and with the nuclear export inhibitor leptomycin B (10 ng/ml; Sigma) was carried out for 5 h at 37ºC. Stock solutions were as follows:

- alpha-amanitin: 1 mg/ml in H_2O
- leptomycin B: 5 mg/ml in 70% methanol

4. TECHNIQUES FOR PROTEIN MANIPULATION

4.1 Western blot analysis

Total cell extracts were prepared by resuspension of cell pellets in 2x sample loading buffer (100 mM Tris-Cl pH 6.8, 200 mM dithiothreitol [DTT], 4% [w/v] sodium dodecyl sulphate [SDS], 20% [v/v] glycerol, 0.2% [w/v] bromophenol blue) and heated for 5 min at 98ºC. Protein samples coming from other extraction methods or reactions (see later) were mixed with 6x sample loading buffer and heated for 5 min at 98ºC. All samples were resolved on SDS-polyacrylamide gels of different acrylamide percentage (depending on the molecular weight of the protein) 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) and protein transfer was checked by staining with Ponceau S (Sigma).

Transferred membranes were blocked with 10% skimmed milk (Central Lechera Asturiana) diluted in PBS-0.1% Tween-20 (Sigma-Aldrich) for 1 h at room temperature (RT) or 16 h at 4ºC, and later incubated with the primary antibody diluted in 5% skimmed milk in PBS-0.1% Tween-20 for 1 h at RT or 16 h at 4ºC. Three washes of 10 min in PBS-0.1% Tween-20 eluted non-bound primary antibody. Membranes were incubated with the secondary antibody conjugated to horseradish peroxidase (HRP) diluted in 5% skimmed milk in PBS-0.1% Tween-20 for 1 h at RT. After two 20 min-washes in PBS-0.1% Tween-20, membranes were revealed with SuperSignal West Pico Chemiluminescent Substrate (Pierce) and finally exposed to a film in a dark room or in the Chemiluminescer Image Analyzer (Fujifilm) with the LAS3000-pro software (Fujifilm).

In the case of the phosphospecific antibodies H5 and H14, blocking and antibody incubations were performed with 3% of bovine serum albumin (BSA) diluted in TBS-0.1% (10 mM Tris-HCl pH 7.5, 100 mM NaCl)-Tween-20. Finally, washes were done with TBS-0.1% Tween-20.

The primary antibodies used in this work are listed in Table M2. Secondary antibodies were HRP-conjugated rabbit anti-mouse, HRP-conjugated goat anti-rabbit and HRP-conjugated rabbit anti-goat (DAKO; 1:2000).

Primary antibody	Host	Working dilution	Commercial brand
Anti-actin	rabbit	1:1000	Sigma
Anti-Brn-3b	goat	1:500/1:5000	Santa Cruz
Anti-Calretinin	rabbit	1:2000	Swant Antibodies
Anti-DYRK1A	mouse	1:500/1:1000	Abnova
Anti-Flag (M2)	mouse	1:10000	Sigma
Anti-GFP (JL-8)	mouse	1:1000	BD Biosciences
Anti-GST (HRP-conjugated, B14)	mouse	1:5000	Santa Cruz
Anti-HA (HA.11)	mouse	1:2000	Babco
Anti-lamin B (H-90)	rabbit	1:200	Santa Cruz
Anti-myc (A14)	rabbit	1:1000	Santa Cruz
Anti-RNA polymerase II (8WG16)	mouse	1:1000	Covance
Anti-RNA polymerase II (N20)	rabbit	1:200	Santa Cruz
Anti-RNA polymerase II (H5)	mouse	1:500	Covance
Anti-RNA polymerase II (H14)	mouse	1:500	Covance

Table M2: Properties and working dilution of the primary antibodies used in Western blots. In the case of the anti-Brn3b and anti-DYRK1A antibodies, the lower and higher dilutions correspond to the detection of the endogenous and transfected protein, respectively.

4.2 Immunofluorescence

Cells were plated in 35-mm dishes ($7x10^5$ cells) containing 10-mm round glass coverslips, transfected and processed 48 h post-transfection. All the procedure was done at RT as follows: cells were washed in PBS, fixed with 4% (w/v) paraformaldehyde in PBS (15 min), permeabilized with 0.1% Triton-X-100 in PBS (10 min) and blocked with 10% FBS in PBS (30 min). After being washed in PBS, cells were incubated with the primary antibody diluted in PBS-1% FBS for 1 h. Then, coverslips were gently washed in PBS-1% FBS and incubated with the secondary antibody (listed below) for 45 min in darkness. Finally, coverslips were washed repeatedly in PBS-1% FBS and mounted using Vectashield (Vector Laboratories) plus 0.2 mg/ml 4',6-diamidino-2-phenylindole (DAPI) or TO-PRO-3 (Molecular Probes) in PBS. Secondary antibodies were Alexa 647-conjugated goat anti-mouse (Molecular Probes, 1:400), Alexa 555-conjugated donkey anti-mouse (Molecular Probes, 1:400), Alexa 488-conjugated donkey anti-goat (Invitrogen, 1:400) and fluorescein isothiocyanate (FITC) conjugated goat anti-rabbit (Southern Biotechnology, 1:400).

Preparations were observed with a confocal inverted microscope Leica SP2 of the Confocal Imaging Service at the Universitat Pompeu Fabra (*http://www.upf.edu/cexs/sct/index.htm?opcio=6*). GFP was excited with the 488 nm line of the Argon laser and IgG Alexa 633 was excited with a 647 nm HeNe laser. In some cases,

preparations were observed in a Leica DMR fluorescence microscope, and photos were taken using a Leica DC500 camera and the software Leica IM1000 Image Manager. Images were processed with Adobe Photoshop.

The primary antibodies used in this work are listed in Table M3.

Primary antibody	Host	Working dilution	Commercial brand
Anti-Brn-3b	goat	1:200	Santa Cruz Biotechnoogy
Anti-GFP	rabbit	1:500	Molecular Probes
Anti-DYRK1A	rabbit	1:500	Home made
Anti-SC35	mouse	1:100	BD Pharmingen
Anti-ubiquitin	mouse	1:50	Santa Cruz Biotechnology
Anti-RNA polymerase II (8WG16)	mouse	1:200	Babco

Table M3: Properties and working dilution of the primary antibodies used in indirect immunofluorescence experiments.

4.3 *In vivo* cell imaging

$7x10^5$ HeLa cells growing on MatTek plates (MatTek Corporation) were transfected with 1 µg of pGFP-HOXA1 and 250 ng of pYFP-SC35. Sixteen h after transfection, cells were transferred to an environmental control box (EMBLEM Technology Transfer) mounted onto the stage of an inverted Leica TCS SP5 confocal microscope at the CRG Imaging Facility (http://pasteur.crg.es/portal/page/portal/Internet/03_SERVICIOS/3485C8485B15B1A3E0401 2AC0E010D1D). GFP was excited with the 488 nm line and YFP with the 514 line of the Argon laser. Cells were treated with alpha-amanitin (50 mg/ml) for 5 h at 37ºC. Time-lapse images were acquired at 63x every 5 min and processed with the LAS (Leica Application Suite) AF software.

4.4 Nuclear fractionation

The procedure was done according to (Schreiber et al., 1989). Briefly, pellets from $2x10^6$ transfected 293T cells were resuspended in 400 µl of buffer A (10 mM Hepes pH 7.9, 10 mM KCl, 0.1 mM EDTA, 0.1 mM ethylene glycol tetraacetic acid [EGTA], 1 mM DTT, 0.5 mM phenylmethylsulphonyl fluoride [PMSF], 10 mM sodium pyrophosphate [NaPPi], 1 mM sodium orthovanadate [$NaVO_4$], 2 mM sodium fluoride [NaF] and a cocktail of protease inhibitors [Complete Mini EDTA-free, Roche]) and incubated for 15 min on ice. Then, 25 µl of 10% Nonidet P-40 (NP-40) were added, cells were vortexed for few seconds and centrifuged for 2 min at 16,000xg at 4ºC. Supernatants (soluble in 10 mM salt fraction) were stored in ice and nuclear pellets were resuspended in 200 µl of buffer B (20 mM Hepes pH 7.9, 400 mM

NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM PMSF, 10 mM NaPPi, 1 mM $NaVO_4$ and the cocktail of protease inhibitors) and incubated in a rotating wheel for 15 min at 4ºC. Samples were centrifuged at 16,000*x*g for 5 min at 4ºC and nuclear supernatants were obtained. Finally, insoluble pellets were resuspended in loading sample buffer. Samples were analysed by Western blot.

4.5 Immunoprecipitation assays

For the preparation of soluble cell extracts, $2x10^6$ transfected HEK-293T cells were washed twice in cold PBS and lysed in buffer C (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 10% glycerol, protease inhibitor cocktail, 10 mM $NaVO_4$ and 30 mM NaPPi. Supernatants were obtained by centrifugation, diluted with one volume of buffer C without NP-40 plus 3 mg of rabbit polyclonal anti-GFP antibody (Molecular Probes) or 2 mg of mouse monoclonal anti-Flag antibody (Sigma) and incubated for 5 h at 4ºC. Equilibrated protein G-Sepharose beads (Amersham Biosciences) were added for 1 h, and then beads were washed extensively with lysis buffer C. Finally, both lysates and immunoprecipitates were resuspended in 6x loading sample buffer and analysed by Western blot.

In the case of the co-immunoprecipitation experiments with RNA polymerase II, nuclear extracts (section 4.4) were diluted with cytoplasmic extracts to have a final salt concentration of 150 mM NaCl and were then incubated with either rabbit polyclonal anti-RNA polymerase II N20 antibody (3 μg) or with rabbit IgGs (Sigma, 3 μg) and kept rolling at 4ºC for 3 h. Equilibrated protein G-sepharose beads (15 μl) were added and incubated rolling at 4ºC for 1 h more. Finally, beads were extensively washed with buffer D (50 mM Hepes pH 7.9, 150 mM NaCl, 0.1% Triton-X-100, 5 mM EDTA, 5 mM DTT), resuspended in 6x loading sample buffer and analyzed by SDS-PAGE and Western blot.

4.6 Nucleic acid binding assays

Binding assays were performed with either soluble cell extracts from transfected cells or with bacterially expressed recombinant gluthatione-S-transferase (GST)-fusion proteins. Trasfected cells were directly scrapped in DNA-lysis buffer (10 mM Tris-HCl pH 7.4, 1% NP-40, 100 mM NaCl, 0.05 mM EDTA, 10% glycerol, 3 mM DTT, 1 mM $NaVO_4$, 1 mM PMSF, 2 mM NaF and a protease inhibitor cocktail; 400 μl/$2x10^6$ cells). Cells were incubated for 30 min in ice, centrifuged at 16,000*x*g for 20 min at 4ºC and supernatants stored on ice.

For the binding assays, 15 μg of cellulose (Sigma), cellulose-coupled single stranded-DNA (Sigma) and cellulose-coupled double stranded-DNA (Amersham Biosciences) were washed 3 times with 5 M NaCl to eliminate the DNA that was weakly bound to the column. Then, columns were equilibrated by washing 3 times with 1 ml of DNA-binding buffer (10 mM

Tris-HCl pH 7.4, 0.1% Triton-X-100, 100 mM NaCl, 0.05 mM EDTA, 1 mM $MgCl_2$ and 3 mM DTT). In the case of poly(U)-Sepharose columns (Amersham), the resin was directly equilibrated without the 5 M NaCl treatment. Soluble cell extracts or purified fusion proteins were added and incubated for 1 h rolling at 4ºC. Finally, samples were washed with cold DNA-binding buffer, resuspended in 2x loading sample buffer and analysed by SDS-PAGE and Western blot.

4.7 Purification of bacterially produced GST-fusion proteins

4.7.1 Expression of the fusion protein

A fresh colony of *E. coli* BL-21 (DE3) pLysS strain containing the plasmid of interest was inoculated in 5 ml of 2xYTA medium (16 g/l triptone, 10 g/l yeast extract, 5 g/l NaCl, 100 μg/ml ampicillin) and grown overnight in a shaker at 37ºC. The miniculture was diluted 1:100 in 250 ml of 2xYTA and grown at 37ºC with shaking until the OD_{600} was between 0.6 and 0.8; then, isopropyl-beta-D-thiogalactopyranoside (IPTG; 0.1 mM for GST-DYRK1A and GST-Brn-3b, 1 mM for GST-CTD) was added to induce the expression of the fusion protein and further incubated with shaking (8 h at 20ºC for GST-DRYK1A; 4 h at 37ºC for the rest of the proteins).

4.7.2 Affinity purification

Bacterial cultures (250 ml) were centrifuged for 10 min at 3,000*x*g and 4ºC and pellets were resuspended in 10 ml of lysis buffer (10 mM Tris-HCl pH 8, 100 mM NaCl, 0.5% NP-40, 1 mM EDTA and a cocktail of protease inhibitors) plus 1 mg/ml lysozime and incubated on ice for 30 min. The cellular suspension was sonicated with 3 pulses (5 s for GST-DYRK1A; 15 s for the rest of the proteins) and 10% amplitude in a digital Branson Sonifier-250, and centrifuged at 10,000*x*g for 15 min at 4ºC. The supernatant was incubated with 200 μl of lysis buffer-equilibrated glutathione-Sepharose beads (Amersham Biosciences) for 1-2 h rolling at 4ºC. Finally, beads were washed 3 times with 5 ml of cold lysis buffer. Fusion proteins were either eluted (see below) or resuspended in 400 μl lysis buffer plus inhibitors.

Both protein concentration and purity were estimated by SDS-PAGE and Coomassie blue staining using BSA as standard.

4.7.3 Elution

GST-fusion proteins were sequentially eluted from glutathione-Sepharose beads with a 1:1 ratio of elution buffer (10 mM reduced glutathione in 50 mM Tris-HCl pH 8) and incubated for 10 min at RT in a rotating wheel. Protein concentration was determined by a colorimetric reaction (Bradford prepared with Brilliant Blue G, Sigma). The most concentrated fractions

were mixed and dialysed overnight at 4ºC against 500 ml of dialysis buffer (30 mM Tris-HCl pH 8, 70 mM KCl and 1 mM DTT) or against kinase buffer (section 4.9), and finally stored with 10% glycerol at –80ºC.

4.8 Pull-down assays

GST-pull down assays were carried out with *in vitro* transcribed and translated proteins using the TnT T7 Coupled Reticulocyte Lysate System (Promega) following manufacturer's instructions. Briefly, synthesis reactions were performed in a final volume of 50 μl using 1 μg of template DNA and 20 μCi of [^{35}S]-radiolabelled methionine (1000 Ci/mM) (Amersham Biosciences) and incubated for 90 min at 30ºC. Equivalent amounts of synthesized proteins were incubated with 5 μg of purified unfused GST or GST-fusion proteins bound to glutathione beads plus 300 μl of binding buffer (20 mM Hepes-KOH, 200 mM KCl, 0.1% Triton-X-100, 0.05% NP-40, 5 mM EDTA, 0.3% BSA and 5 mM DTT) for 3 h rolling at 4ºC. Beads were extensively washed with 1 ml of cold washing buffer (20 mM Hepes-KOH, 500 mM KCl, 0.1% Triton-X-100, 0.05% NP-40, 5 mM EDTA and 5 mM DTT), resuspended in 2x loading sample buffer and resolved in a SDS-PAGE gel. The gel was stained with Coomassie Blue to ensure that equivalent amounts of GST proteins were used. The stained gel was dried and exposed to a phosphoimager screen for 1 h or to a film for 16 h. Images were obtained in a phosphoimager (FLA-500, Fujifilm).

4.9 In vitro kinase assays

For *in vitro* kinase (IVK) assays using GST-DYRK1A as enzyme, the fusion protein was affinity-purified and dialyzed against kinase buffer (50 mM Hepes pH 7.4, 5 mM $MgCl_2$, 5 mM $MnCl_2$). For radioactive kinase assays and to avoid the appearance of autophosphorylation bands, GST-DYRK1A (50 ng) was first incubated in kinase buffer plus 0.5 mM DTT and 100 μM ATP for 20 min at 30ºC in a final volume of 25 μl. Then, 1 μCi of [^{32}P]-γ-ATP (3000 Ci/mmol) and the substrate were added to the mix to a final volume of 50 μl, and the samples further incubated for 30 min at 30ºC.

When the IVK assays were performed with immunocomplexes from transfected cells, anti-GFP immunocomplexes were split into two aliquots, one for Western blotting and the other for IVK. Immunocomplexes were washed twice with kinase buffer plus 0.5 mM DTT, and incubated for 20 min at 30ºC in 30 µl of kinase buffer with a final concentration of 10 µM ATP and [g^{32}P]-ATP ($6.5x10^{-3}$ µCi/pmol).

Reactions were stopped by addition of 2x loading sample buffer, and samples were resolved by SDS-PAGE and then stained with Coomassie blue. ^{32}P incorporation was detected by autoradiography of dried gels. Images were obtained in a phosphoimager (FLA-

500, Fujifilm) and quantified with the Image Gauge v4.0 software (Fujifilm).

4.10 Dephosphorylation assays

Cell extracts were prepared by resuspending cells in phosphatase buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% NP-40, 2 mM $MgCl_2$, 2 mM PMSF and a cocktail of protease inhibitors) and clarification by centrifugation at 16,000*x*g for 15 min at 4ºC. Three reaction mixes were prepared:

1) 40 µl of extract
2) 40 µl of extract + 30 U of alkaline phosphatase (Alkaline Phosphatase from Calf thymus, Sigma)
3) 40 µl of extract + 30 U of alkaline phosphatase + 40 mM NaPPi

Reactions took place for 30 min at 30ºC in a final volume of 75 µl. In reaction 3, the phosphatase was pre-incubated with NaPPi for 30 min at 30ºC. The reactions were stopped by adding 6x loading sample buffer and analysed by Western Blot.

4.11 Luciferase Reporter assays

10^5 CV-1 cells were transfected and lysed 48 h after with 150 µl of 5x Reporter Lysis Buffer (Promega). Lysates were incubated at 4ºC for 10 min and centrifuged at 16.000*x*g for 10 min. 15 µl of the extracts were used to mesure luciferase activity in a Berthold luminometre with the Luciferase Assay System kit (Promega). At the same time, 50 µl of the extracts were used to mesure β-galactosidase activity: extracts were incubated with 50 µl of β-galactosidase buffer (200 mM $Na_3(PO4)_2$, 2 mM $MgCl_2$, 100 mM β-mercaptoethanol, 1.33 mg/ml ortho-nitrophenyl-b-D- galactopyranoside) at 37ºC for 20 min. The colorimetric assay was mesured at OD_{420}.

4.12 Protein quantification

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

5. BIOINFORMATIC TOOLS

5.1 Search in the databases

Sequence search, both DNA and protein, has been performed basically using the public database from the NCBI (*National Center for Biotechnology Information*), through the Entrez system. This inclcudes information from different sources (GenBank, RefSeq,…). This

database has been also used for the search and query of bibliographic references with PubMed.

Web site: http://www.ncbi.nlm.nih.gov/entrez/

5.2 DNA sequence analysis

The alignment and comparison of both DNA and protein sequences has been carried out with the BLAST (*Basic Alignment Search Tool*) and Blast2Sequences programs from NCBI.

Web site: http://www.ncbi.nlm.nih.gov/BLAST/

Web site: http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi

The analysis of the chromatograms of DNA sequences obtained from the Genomic Sequencing Service at the Universitat Pompeu Fabra has been done with the program 4Peaks (Freeware; *www.mekentosj.com*).

For the translation from DNA to amino acid sequences and the generation of plasmid maps the program MacVector 3.5 (Accelrys) has been used. The identification of restriction sites and the generation of restriction maps have been performed with the programs DNA Strider and MacVector.

5.3 Protein analysis

The identification of protein domains and putative phosphorylation sites has been carried out with the program PROSITE, sited at the ExPASy (*Expert Protein Analysis System*) from the Bioinformatic Institute of Switzerland (BIS), or with the program SMART (*Simple Modular Architecture Research Tool*).

Web site: http://www.expasy.org/prosite/

Web site: http://smart.embl-heidelberg.de

The alignment of amino acid sequences was performed with the multiple sequence alignment program.

Web site: http://www.ebi.ac.uk/Tools/clustalw2/index.html

The prediction of disordered regions in polyHis-containing proteins was carried out using the IUPRED software.

Web site: http://iupred.enzim.hu/index.html

The building of interaction networks between proteins with His repeats and their distribution by functional categories were performed with the Ingenuity Pathway Analysis software.

Web site: http://www.ingenuity.coml

The prediction of phosphorylated sites was carried out with the Phosphosite website.

Web site: http://www.phosphosite.org/

RESULTS

1. The His repeat in DYRK1A as a targeting signal to the nuclear speckle compartment

1.1 The background

As previously introduced, DYRK1A localizes both in the nucleus and in the cytoplasm of the cells. In the nucleus, the kinase accumulates in nuclear speckles via its His repeat (Alvarez et al., 2003), as demonstrated by co-localization with an endogenous SFC marker, the splicing factor SC35 (Fu and Maniatis., 1990). This repetitive region was reported to be both necessary and sufficient for the targeting to the SFC (Alvarez, 2004), as demonstrated by localization studies with a DYRK1A mutant skipping the His tract (DYRK1AΔHis) and with a GFP fusion of the His fragment (DYRK1A 590-616). Figure 1 summarizes these results.

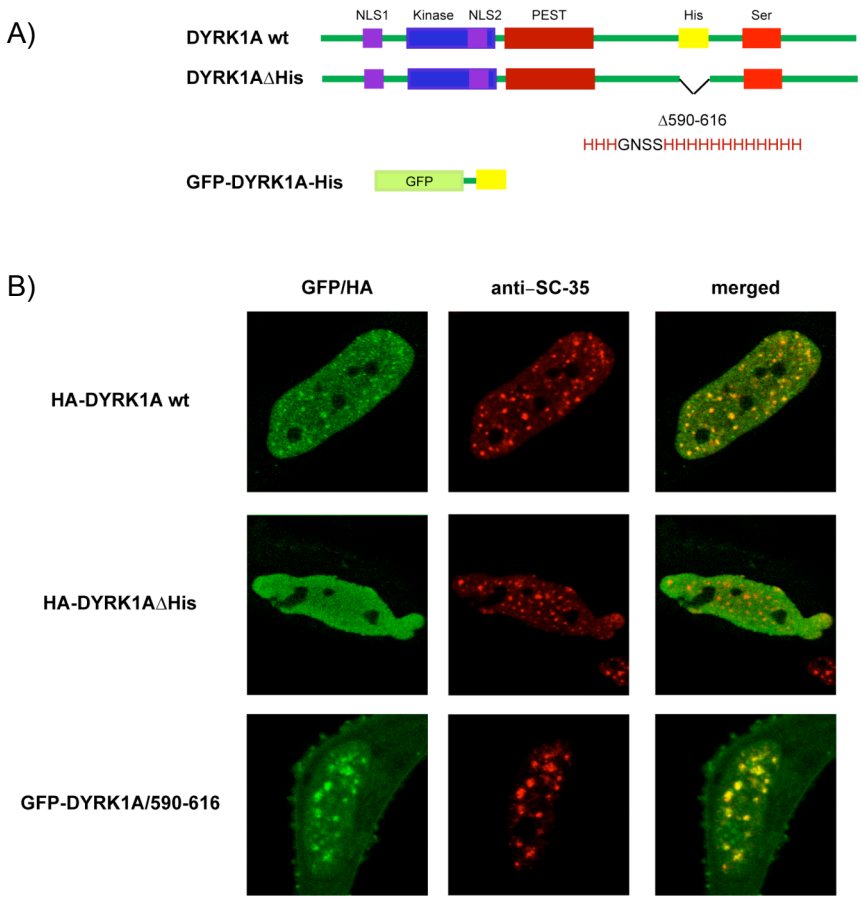


Figure 1: The His repeat is necessary and sufficient to localize DYRK1A to the nuclear speckles compartment. A) Scheme representing the constructs used. B) HeLa cells were transfected with the expression plasmids pHA-DYRK1A, pHA-DYRK1AΔHis and pGFP-DYRK1A/590-616. At 48 h post-transfection, the localization of the GFP fusions was assessed by direct fluorescence, whereas HA-DYRK1A and SC35 were detected by immunofluorescence staining with anti-DYRK and anti-SC35 antibodies, respectively. FITC-conjugated goat anti-rabbit and Alexa 647-conjugated goat anti-mouse were used as secondary antibodies. All the images were taken by confocal microscopy.

Given that these results were obtained by transient transfections of expression plasmids, the next sections are dedicated to show experiments that try to rule out non-physiological effects due to protein over-expression.

1.2 The dot-like pattern of staining in DYRK1A is not a result of intranuclear protein aggregation

The presence of homopolymeric tracts of amino acids has been associated to increased insolubility in proteins (Oma et al., 2004). In the case of polyHis-containing proteins, the formation of diffuse aggregates within the cytoplasm of cells overexpressing a GFP fusion with 26 His residues has been described (Oma et al., 2004). Moreover, polyHis expansions in the transcription factor HOXA1 resulted in nuclear protein aggregates (Paraguison et al., 2005). The HOXA1 protein present in aggregates is highly ubiquitinated, as it has been also described also for other homopeptides involved in repeat expansion diseases (Calado et al., 2000). In these cases, the co-localization with ubiquitin was used as a marker of protein misfolding and aggregation. To rule out the possibility that the nuclear punctuated staining observed in DYRK1A was due to protein aggregation, an immunostaining assay with an anti-ubiquitin antibody was performed. As shown in Figure 2, no co-localization was observed between the fusion protein and endogenous ubiquitin, indicating that the nuclear foci are not the result of protein aggregation.

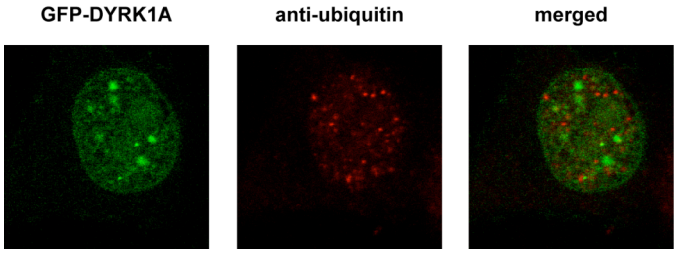


Figure 2: The dot-like staining pattern of DYRK1A does not overlap with ubiquitin-enriched nuclear aggregates. HeLa cells were transfected with the expression plasmid encoding GFP-DYRK1A. At 48 h post-transfection, cells were immunostained with an anti-ubiquitin antibody and analyzed by direct fluorescence (left panel, green) and by indirect immunofluorescence (middle panel, red). A merged image is also shown (right panel). Note that no co-localization of the DYRK1A nuclear speckles with the signal from the ubiquitin antibody is detected.

1.3 The accumulation of DYRK1A in nuclear speckles is not due to protein over-expression

To provide evidence of the staining pattern not being the result of protein over-expression, the subnuclear localization was assessed on stable cell lines expressing GFP-DYRK1A/388-616 (H+), a protein fragment that includes the His repeat and that previously shown is able to target GFP to the SFC (Alvarez et al., 2003). Unfortunately, this experiment could not be

performed with full-length DYRK1A due to the impossibility to generate stable cell lines. These difficulties were very likely due to the cell cycle arrest induced by the expression of the protein (S. Aranda and S. de la Luna, unpublished results). A stable cell line expressing the fusion GFP-DYRK1A/388-588 (H-), which lacks the His domain, was used as a negative control. The scheme of the fusion proteins used in the experiment is summarized in Figure 3A. The analysis by direct fluorescence showed that no differences existed between the nuclear localization of the stable versus the transiently expressed DYRK1A fragment (Figure 3B, central column). The diffuse nucleoplasmic staining of the construct without the His tail was neither altered (Figure 3B, right column).

To further ensure that the nuclear dots corresponded to nuclear speckles, an immunofluorescence staining with anti-SC35 was performed. As illustrated in Figure 3C (right column), both the transient and the stably expressed DYRK1A-H+ fusion co-localized with SC35, indicating that the accumulation in the SFC is an intrinsic property of the kinase and has nothing to do with protein over-expression.

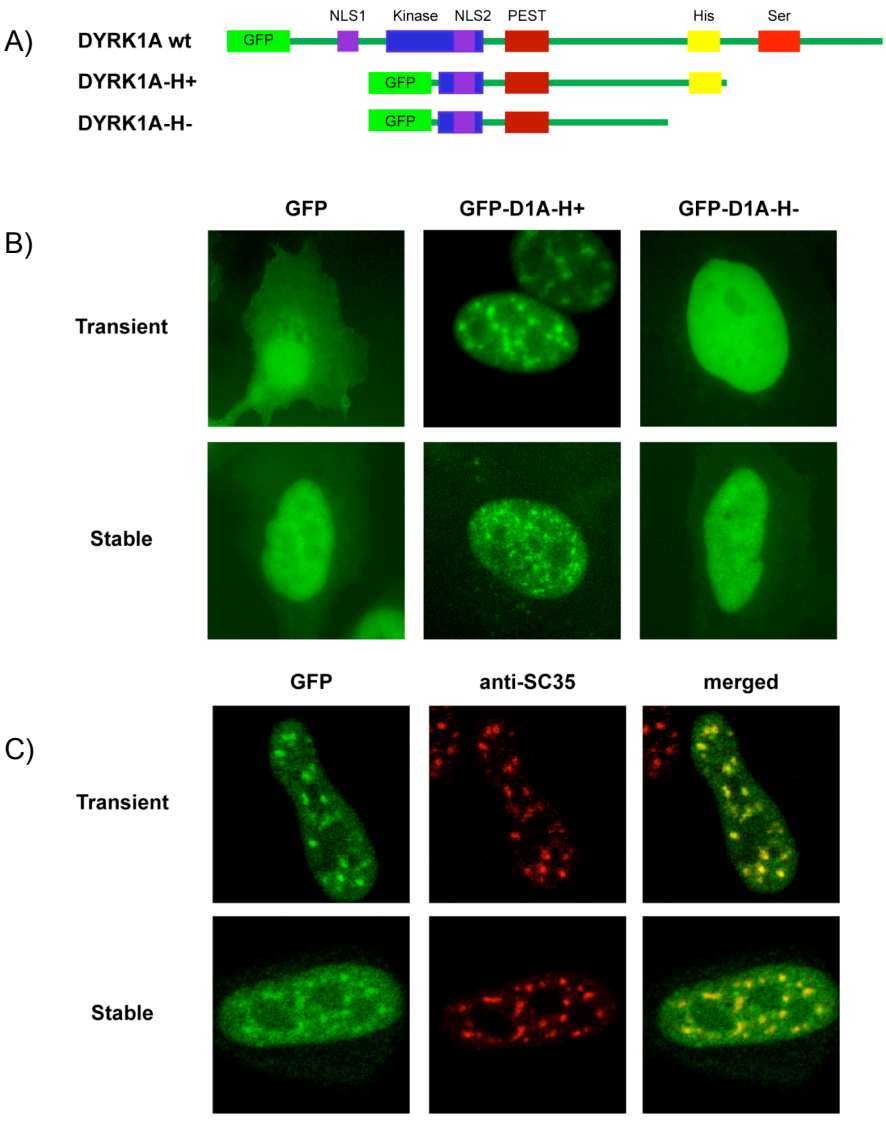


Figure 3: The phenotype of the localization in nuclear speckles of DYRK1A is not a result of protein over-expression. A) Scheme showing the GFP fusions used in the experiment, with (H+) or without (H-) the His repeat. B) U2-OS cells were transiently transfected with plasmids expressing GFP, GFP-DYRK1A-H+ or GFP-DYRK1A-H- fusions. Stable U2-OS cells expressing the same fusion proteins were platted simultaneously. At 24 h, cells were analyzed by direct fluorescence. Unfused GFP was used as a general negative control. C) U2-OS transiently expressing GFP-DYRK1A-H+ or stable cells expressing the identical construct in which the localization of the fusion protein was analyzed by direct fluorescence. Co-localization analysis was performed by immunostaining with anti-SC35 antibody. Alexa 647-conjugated goat anti-mouse was used as secondary antibody.

1.4 DYRK1A follows the behaviour of endogenous SFC proteins along the cell cycle

It is known that SC35-positive speckles undergo through dynamic changes during the cell cycle: they are lost during prophase and they begin to reform in the cytoplasm of telophase cells (Figure 4A, adapted from (Spector et al., 1991)).The stable cell lines allowed the analysis of the subcellular localization of the fusion protein GFP-DYRK1A-His+ during the cell cycle. Figure 4B shows that the behaviour of the fusion protein DYRK1A-H+ completely followed the dynamics described for other SFC-resident proteins, and therefore indicates that the polyHis-segment is a *bona fide* targeting signal to the nuclear speckles compartment.

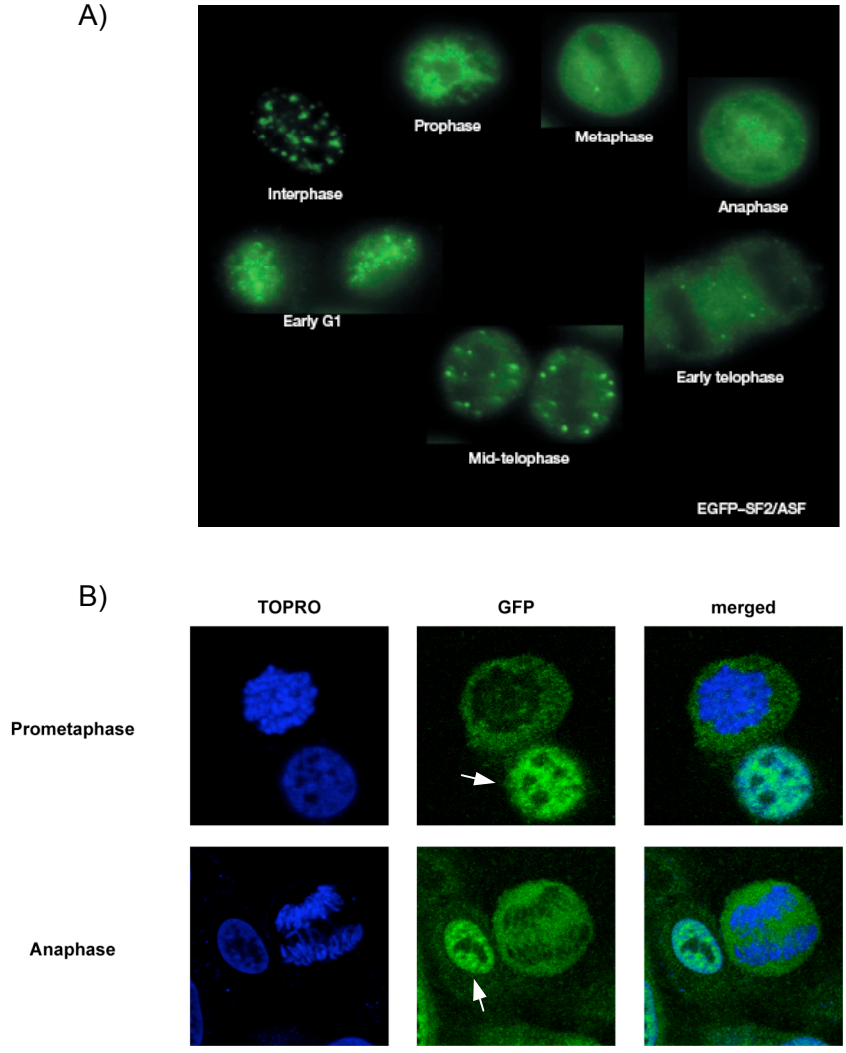


Figure 4: DYRK1A mimics the behaviour of an endogenous SFC component during mammalian cell cycle. A) The speckle cell cycle. The interphase speckled pattern disperses as cells enter prophase of mitosis. From metaphase to mid-telophase, mitotic interchromatin granules (MIGs) appear in the cytoplasm and increase in number and size. During mid–late telophase, factors leave MIGs and enter the daughter nuclei, and speckles form during early G1 phase. HeLa cells that are expressing enhanced green fluorescent protein (EGFP)–pre-messenger RNA splicing factor/alternative splicing factor (SF2/ASF) are shown (adapted from Spector et al., 1991). B) U2-OS stable cell line expressing GFP-DYRK1A-H+ growing in coverslips was analyzed by direct fluorescence (central panel). DNA was counterstained with TOPRO to distinguish interphasic versus mitotic cells (left panel). Note that in both interphasic nuclei (white arrows), GFP-DYRK1A-H+ is expressed in discrete foci compatible with nuclear speckles, whereas during mitosis (prophase, upper panel; anaphase, lower panel) the staining is diffused through the cytoplasm as a consequence of SFC disassembly.

2. PolyHis repeats are a novel nuclear speckle-targeting signal

2.1 A repeat of 6 His residues is sufficient to direct a heterologous protein to the SFC

Since His repeats are present in several mammalian proteins, it would be possible that they act as general SFC-targeting signals. To initially test this, the localization of chimeric proteins of GFP and His tracts was analyzed. Plasmids expressing GFP fusion proteins with 5, 6, 7, 8 or 9 His were generated and their subcellular localization was assessed by direct fluorescence and by immunostaining with anti-SC35. The comparison of the distribution of these fusion proteins did not reveal significant differences in the staining pattern between GFP and GFP-5xHis (Figure 5). However, from 6xHis on, a linear dependence between the accumulation in the SFC and the length of the His-tract was detected (Figure 5). This suggests that a minimum of 6 His residues would be sufficient to target a protein to the SFC.

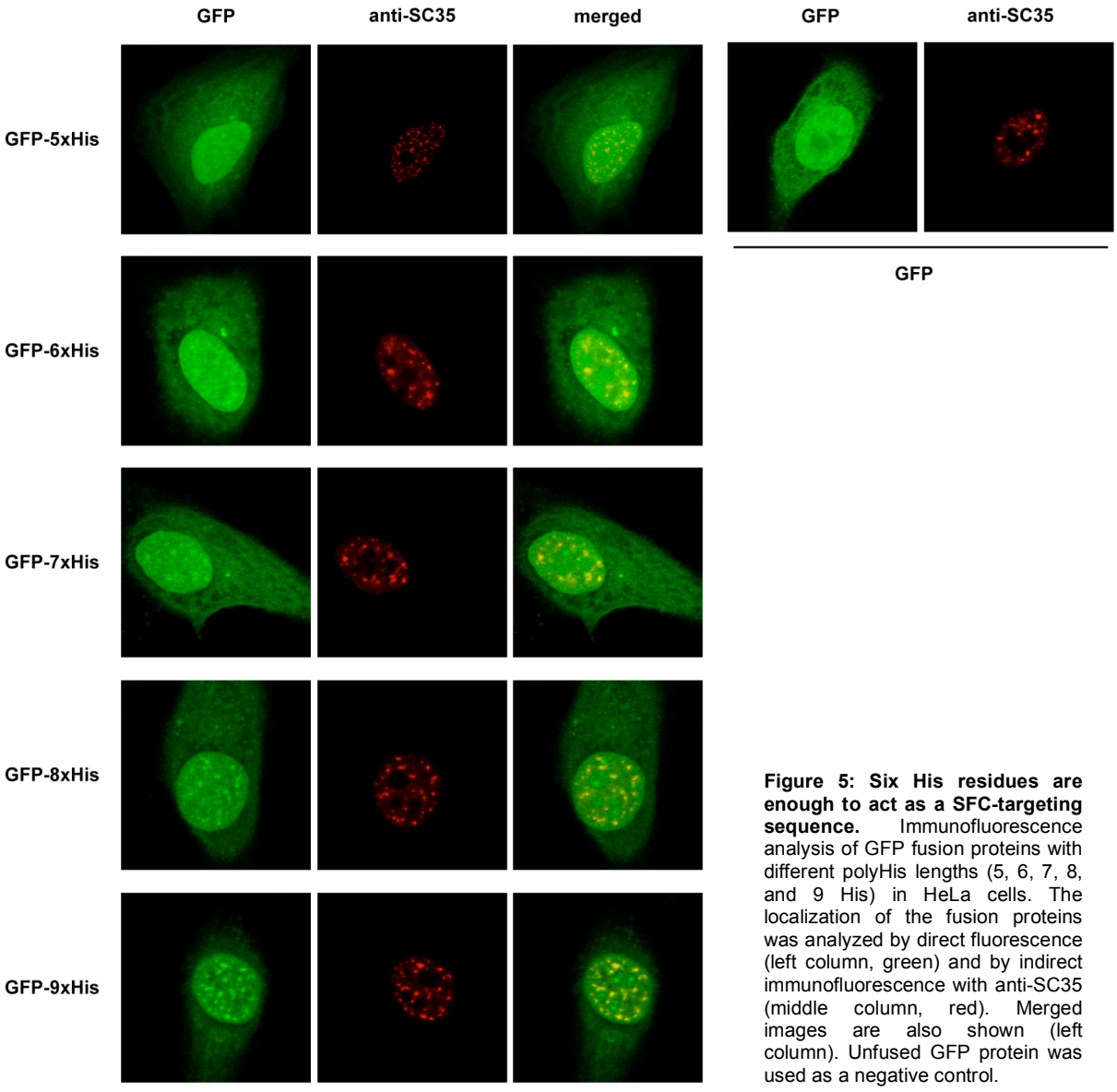
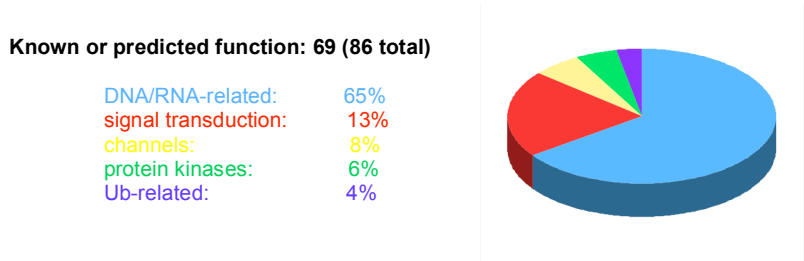


Figure 5: Six His residues are enough to act as a SFC-targeting sequence. Immunofluorescence analysis of GFP fusion proteins with different polyHis lengths (5, 6, 7, 8, and 9 His) in HeLa cells. The localization of the fusion proteins was analyzed by direct fluorescence (left column, green) and by indirect immunofluorescence with anti-SC35 (middle column, red). Merged images are also shown (left column). Unfused GFP protein was used as a negative control.

2.2 His repeats are overrepresented in nuclear proteins

To extrapolate these results to real proteins, we first search for human proteins containing His repeats in their primary sequence. For a typical protein of 400 amino acids and average composition, a run of an individual amino acid is statistically significant if it is 5 or more residues long (Karlin, 1995). Therefore, a bioinformatics screen on the Ensembl database (Ensembl 48, based on NCBI36; Flicek et al., 2008) was performed to find out all human proteins containing at least one His repeat of size 5 or longer. This work was done in a collaboration with Mar Albà's group of the Universitat Pompeu Fabra/Institut Municipal d'Investigació Mèdica in Barcelona (http://evolutionarygenomics.imim.es/index.html). The search identified 86 Ensembl genes (Table 1), most of them related to DNA/RNA synthesis or processing, suggesting that they should have a nuclear phase (Figure 6A, Table 1). Several protein families were represented in the screening: POU-domain transcription factors, Maf transcription factors, naked cuticle proteins and several Ca^{2+}-channels. Moreover, a subset of four protein kinases was also present: DYRK1A, nemo-like kinase (NLK), polo-like kinase 2 (PLK2) and BMP2-inducible kinase (BMP2K). The analysis of the GO functions in the dataset encoding polyHis-containing proteins confirmed the enrichment in transcription factors and developmental proteins. The Ingenuity Pathways Analysis application was used to distribute polyHis proteins into several functional categories (Figure 6B).

A)



B)

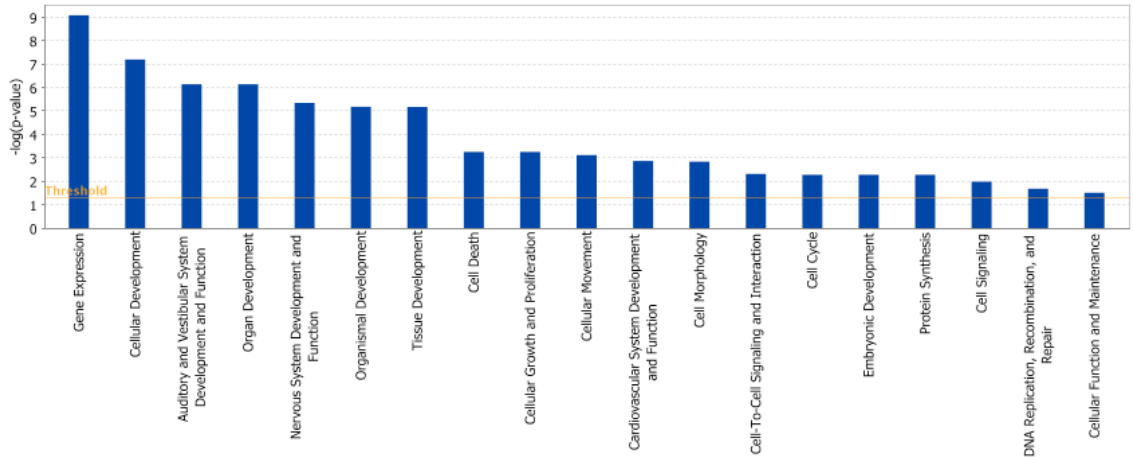


Figure 6: Distribution of the proteins in the dataset according to their cellular function. A) GO functions were grouped by similarity and expressed as a percentage of the total number of proteins (86). B) The Ingenuity Pathway Analysis software was used to build a bar diagram representing the main functional categories represented in the protein data set. The value of the bars is given as the negative logarithm of the p-value (Fisher's Exact Test p-value). The threshold line (yellow) indicates a p-value of 0.01.

Table 1: List with the whole dataset of 86 proteins containing 5 or more His residues.

ID	Description	Predicted subcellular location[a]	Type
AMOT	angiomotin	Plasma Membrane	unknown
ANKRD57	ankyrin repeat domain 57	Nucleus	unknown
ARID1A	AT rich interactive domain 1A (SWI-like)	Nucleus	transcription regulator
ATN1	atrophin 1	Nucleus	transcription regulator
AUTS2	autism susceptibility candidate 2	Unknown	unknown
AXIN2	axin 2 (conductin, axil)	Cytoplasm	intracellular signalling
BEAN	brain-expressed associated with NEDD4	Membrane	unknown
BMP2K	BMP2 inducible kinase	Nucleus	protein kinase
BRD4	bromodomain containing 4	Nucleus	transcription regulator
BRPF1	bromodomain and PHD finger containing, 1	Nucleus	transcription regulator
BTBD11	BTB (POZ) domain containing 11	Membrane	unknown
C10orf140	chromosome 10 open reading frame 140	Unknown	unknown
C21orf58	chromosome 21 open reading frame 58	Unknown	unknown
CACNA1	calcium channel, voltage-dependent, N type, alpha 1B subunit	Plasma Membrane	ion channel
CACNA1B	calcium channel, voltage-dependent, N type, alpha 1B subunit	Plasma Membrane	ion channel
CACNA1G	calcium channel, voltage-dependent, T type, alpha 1G subunit	Plasma Membrane	ion channel
CACNA1H	calcium channel, voltage-dependent, T type, alpha 1H subunit	Plasma Membrane	ion channel
CBL	Cas-Br-M (murine) ecotropic retroviral transforming sequence	Cytoplasm	intracellular signalling
CBX4	chromobox homolog 4 (Pc class homolog, *Drosophila*)	Nucleus	transcription regulator
CCNT1	cyclin T1	Nucleus	transcription regulator
CHD8	chromodomain helicase DNA binding protein 8	Nucleus	DNA helicase
CSPG3	neurocan	Extracellular Space	cell adhesion
DLGAP3	discs, large (*Drosophila*) homolog-associated protein 3	Cytoplasm	unknown
DLX2	distal-less homeobox 2	Nucleus	transcription regulator
DLX6	distal-less homeobox 6	Nucleus	transcription regulator
DYRK1A	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A	Nucleus	protein kinase
ERC2	ELKS/RAB6-interacting/CAST family member 2	Cytoplasm	unknown
FAM76B	family with sequence similarity 76, member B	Unknown	unknown
FBXO41	F-box protein 41	Unknown	protein ubiquitination
FOXB2	forkhead box B2	Nucleus	transcription regulator
FOXF2	forkhead box F2	Nucleus	transcription regulator
FOXG1B	forkhead box G1	Nucleus	transcription regulator
GATA6	GATA binding protein 6	Nucleus	transcription regulator
GSX2	GS homeobox 2	Nucleus	transcription regulator
GTF2A1	general transcription factor IIA, 1, 19/37kDa	Nucleus	transcription regulator
HAND1	heart and neural crest derivatives expressed 1	Nucleus	transcription regulator
HOXA1	homeobox A1	Nucleus	transcription regulator
HOXA9	homeobox A9	Nucleus	transcription regulator
HRC	histidine rich calcium binding protein	Cytoplasm	unknown
HRCT1	histidine-rich carboxyl terminus 1	Membrane	unknown
IQSEC2	IQ motif and Sec7 domain 2	Unknown	intracellular signalling
KCNN2	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 2	Plasma Membrane	ion channel
LMO6	prickle (*Drosophila*) homolog 3	Unknown	unknown
LRCH1	leucine-rich repeats and calponin homology (CH) domain containing 1	Unknown	unknown
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene homolog (avian)	Nucleus	transcription regulator
MAFA	v-maf musculoaponeurotic fibrosarcoma oncogene homolog A (avian)	Nucleus	transcription regulator
MAFB	v-maf musculoaponeurotic fibrosarcoma oncogene homolog B (avian)	Nucleus	transcription regulator

MECP2	methyl CpG binding protein 2 (Rett syndrome)	Nucleus	transcription regulator
MEOX2	mesenchyme homeobox 2	Nucleus	transcription regulator
MEPCE	methylphosphate capping enzyme	Nucleus	transcription regulation
NA	neurocanthocytosis	Unknown	unknown
NKD1	naked cuticle (*Drosophila*) homolog 1	Cytoplasm	intracellular signalling
NKD2	naked cuticle (*Drosophila*) homolog 2	Cytoplasm	intracellular signalling
NLK	nemo-like kinase	Nucleus	protein kinase
NR4A3	nuclear receptor subfamily 4, group A, member 3	Nucleus	transcription regulator
NUFIP2	nuclear fragile X mental retardation protein interacting protein 2	Nucleus	RNA binding
ONECUT1	one cut homeobox 1	Nucleus	transcription regulator
ONECUT2	one cut homeobox 2	Nucleus	transcription regulator
OTX1	orthodenticle homeobox 1	Nucleus	transcription regulator
PBXIP1	pre-B-cell leukemia homeobox interacting protein 1	Nucleus	transcription regulator
PLK2	polo-like kinase 2	Unknown	protein kinase
POU3F1	POU class 3 homeobox 1	Nucleus	transcription regulator
POU3F3	POU class 3 homeobox 3	Nucleus	transcription regulator
POU4F1	POU class 4 homeobox 1	Nucleus	transcription regulator
POU4F2	POU class 4 homeobox 2	Nucleus	transcription regulator
PRDM13	PR domain containing 13	Nucleus	unknown
PRRT1	proline-rich transmembrane protein 1	Membrane	unknown
RBM15B	RNA binding motif protein 15B	Nucleus	RNA binding
RBM33	RNA binding motif protein 33	Nucleus	RNA binding
RHOBTB2	Rho-related BTB domain containing 2	Unknown	GTPase
RNF111	ring finger protein 111 (or Arkadia)	Unknown	ubiquitin E3 ligase
SHANK1	SH3 and multiple ankyrin repeat domains 1	Membrane	intracellular signaling
SLC39A6	solute carrier family 39 (zinc transporter), member 6	Plasma Membrane	transporter
SORBS2	sorbin and SH3 domain containing 2	Cytoskeleton	intracellular signaling
SYNGAP1	synaptic Ras GTPase activating protein 1 homolog	Plasma Membrane	intracellular signaling
TAF2	TAF2 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 150kDa	Nucleus	transcription regulator
TBX1	T-box 1	Nucleus	transcription regulator
TNKS	tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase	Nucleus	poly(ADP-ribose) polymerase
USP34	ubiquitin specific peptidase 34	Unknown	ubiquitin protease
VGLL3	vestigial (*Drosophila*) like 3	Nucleus	transcription regulator
YY1	YY1 transcription factor	Nucleus	transcription regulator
ZCCHC14	zinc finger, CCHC domain containing 14	Nucleus	transcription regulator
ZIC2	Zic family member 2 (odd-paired homolog, *Drosophila*)	Nucleus	transcription regulator
ZIC3	Zic family member 3 heterotaxy 1 (odd-paired homolog, *Drosophila*)	Nucleus	transcription regulator
ZNF281	zinc finger protein 281	Nucleus	transcription regulator
ZNF609	zinc finger protein 609	Nucleus	transcription regulator

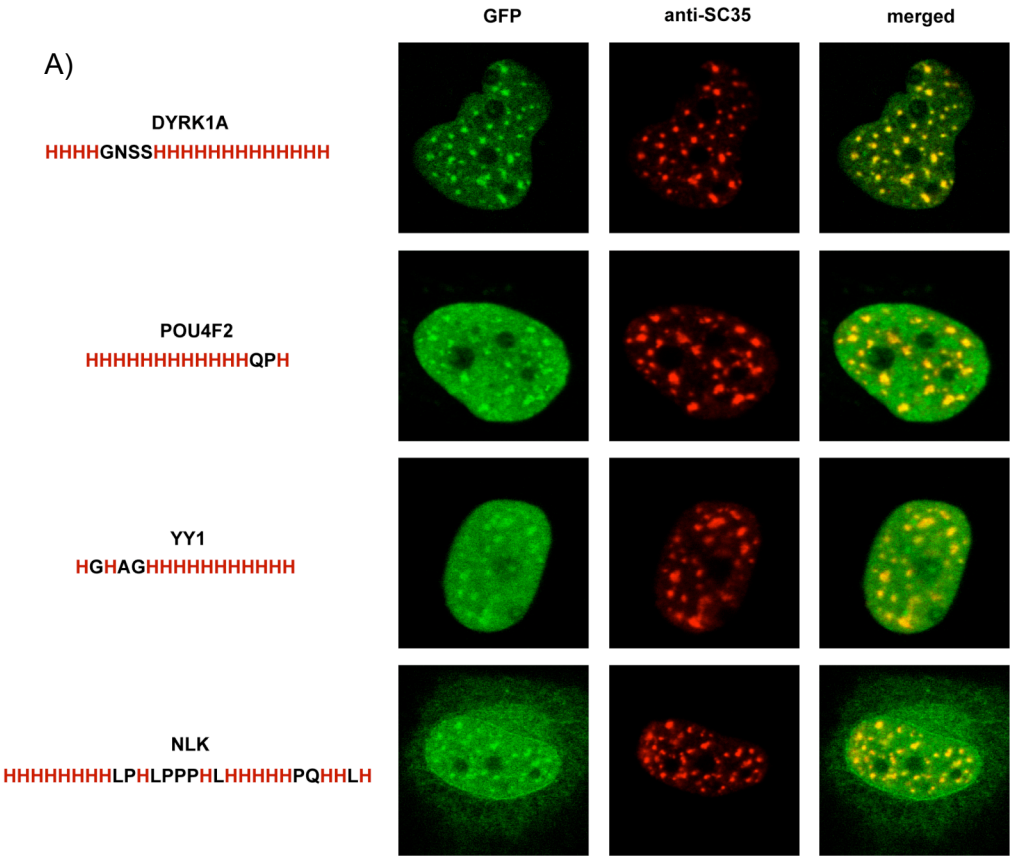
Proteins are named by alphabetical order. The ID corresponds to HUGO official name. [a]Predicted subcellular localization according to GENE entries. Proteins analyzed in this study are shown in bold.

Regarding the features of the His repeats, two different kind of tracts were observed: i) pure repeats, represented by His residues alone, as for instance in the case of the transcription factor HOXA1 (HHHHHHHHHH); and ii) extended repeats, made by at least one His pure repeat of size 5 or longer, additional His residues and other “interrupting” residues

(often P, Q, G, S and A). This is the case of, for example, the transcription factor OTX1 (HHHHHPHAHHPLSQSSGHHHHHHHHHH).

2.3 The His repeat is a novel SFC-directing sequence

Most of the polyHis-containing proteins are predicted to be nuclear and, therefore, the possibility existed to be targeted to the SFC. From the group of predicted nuclear proteins, proteins with pure His repeats of different lengths and several proteins with extended repeats were chosen to analyze their subcellular localization. The open reading frame of the candidates was cloned as GFP fusion proteins in a mammalian expression vector and the distribution of the fusion proteins was analyzed by direct fluorescence. In addition, the presence of nuclear speckles was revealed by indirect immunofluorescence with anti-SC35. Similarly to the behaviour previously described for cyclin T1 and DYRK1A, other polyHis-containing proteins such as the transcription factors POU4F2 or YY1, or the protein kinase NLK, also showed a nuclear punctuated staining that co-localized with SC35 (Figure 7A). These results highly suggest that the His repeat might act as a SFC-localization signal. Moreover, the His tract fulfils the requirement of being necessary since a mutant with a deletion covering the His-repeat in, for instance, the kinase DYRK1A (DYRK1AΔHis) and in the transcription factor POU4F2 (POU4F2ΔHis), showed a diffuse nuclear staining and no accumulation in SC35 positive foci (Figure 7B).



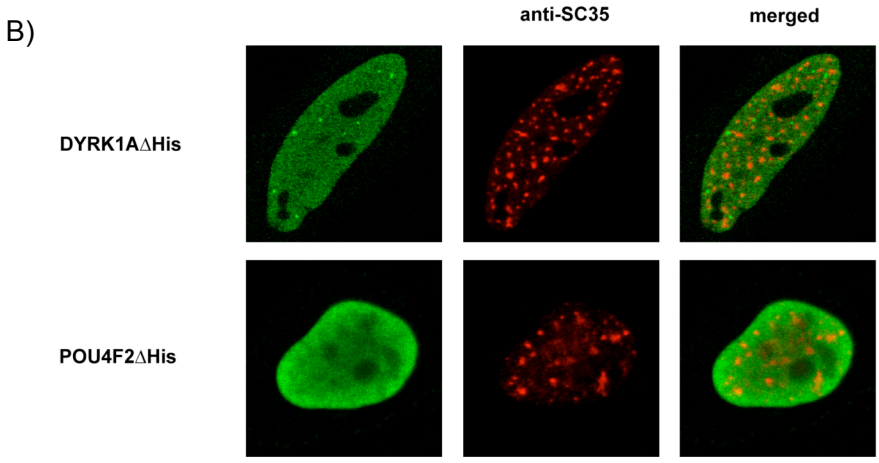


Figure 7: The His-repeat is a novel SFC-accumulating signal. A) HeLa cells were transfected with expression plasmids for the fusion proteins GFP-DYRK1A, GFP-POU4F2, GFP-YY1 and GFP-NLK. At 48 h post-transfection, cells were immunostained with anti-SC35 antibody to visualize nuclear speckles (middle column, red). GFP fusion proteins were directly visualized by fluorescence microscopy (left column, green). Merged images are also shown (right column). B) HeLa cells were transfected with the expression plasmids pHA-DYRK1AΔHis and pFlag-POU4F2ΔHis. At 48 h after transfection, cells were immunostained with anti-DYRK1A or with anti-POU4F2 and FITC-conjugated goat anti-rabbit or Alexa 488-conjugated donkey anti-goat (left column). In both cases anti-SC35 and Alexa 647-conjugated goat anti-mouse or Alexa 555-conjugated donkey anti-mouse, respectively, were used to detect nuclear speckles (middle column).

2.4 Homologous proteins without His repeats fail to localize in nuclear speckles

The analysis of the set of polyHis-containing proteins revealed that some of them belong to families of proteins with members that, in spite of being highly homologous, do not contain a His-repeat in their primary sequence. Examples for these groups of paralogous proteins have been summarized in Table 2. Thus, it was reasonable to think that if the His-tract was responsible for the speckle accumulation phenotype, then the paralogous without polyHis would not accumulate in this subnuclear compartment. To confirm this hypothesis, the localization of the pair of paralogous proteins FAM76A/FAM76B was analyzed. The sequence alignment illustrates the high degree of conservation between both proteins, only lost in the region comprising the His stretch (Figure 8A). As shown in Figure 8B, only FAM76B accumulated in the SFC, whereas the paralogous without the His-tract, FAM76A, presented a diffuse nucleoplasmic staining. A similar result was also obtained with the pairs DYRK1A/DYRK1B and POU4F2/POU4F3 (Figure 8C and D). Therefore, these findings support the proposal that protein segments with polyHis represent a sequence necessary for localizing proteins in the SFC.

Table 2. List of paralogous proteins including or not His-repeats in their amino acid sequence

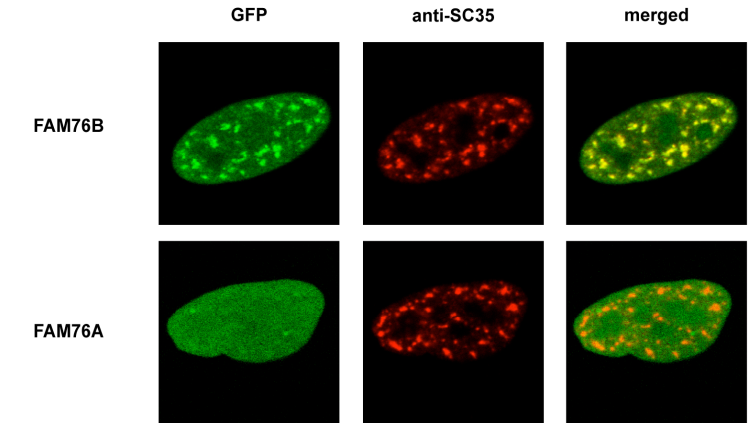
With His tracts	Without His tracts
BMP2K (NP_942595)	BMP2KL (XP_293293)
DLX2 (NP_004396)	DLX1 (NP_835221)
DYRK1A (NP_001387)	DYRK1B (NP_004705)
FAM76B (NP_653265)	FAM76A (NP_689873)
FOXB2 (NP_001013757)	FOXB1 (NP_036314)
FOXF2 (NP_001443)	FOXF1 (NP_036314)
GATA6 (NP_005248)	GATA4 (NP_002043) GATA5 (NP_536721)
GSH2 (NP_573574)	GSH1 (NP_663632)
HAND1 (NP_004812)	HAND2 (NP_068808)
HOXA1 (NP_005513)	HOXB1 (NP_002135)
HOXA9 (NP_689952)	HOXB9 (NP_076922) HOXC9 (NP_008828)
MafA (NP_963883) MafB (NP_005452) cMaf (NP_005351)	NRL (NP_006168)
MEOX2 (NP_005915)	MEOX1 (NP_004518)
NKD2 (NP_149111)	NKD1 (NP_149110)
NR4A3/Nor1 (NP_008912)	NR4A1/Nur77 (NP_775180) NR4A2/Nurr1 (NP_006177)
ONECUT1 (NP_004489) ONECUT2 (NP_004843)	ONECUT3 (NP_001073957)
OTX1 (NP_055377)	OTX2 (NP_068374)
PLK2 (NP_006613)	PLK (NP_005021) PLK3 (NP_004064)
POU3F1/Oct-6 (NP_002690) POU3F3/Brn-1 (NP_006227)	POU3F2/Brn-2 (NP_005595) POU3F4/Brn-4 (NP_000298)
POU4F1/Brn-3a (NP_006228) POU4F2/Brn-3b (NP_004566)	POU4F3/Brn-3c (NP_002691)
PRICKLE3/LMO6 (NP_006141)	PRICKLE1/RILP (NP_694571) PRICKLE2 (NP_942559)
VGLL3 (NP_005513)	VGLL2 (NP_872586)
YY1 (NP_003394)	YY2 (NP_996806)
ZIC2 (NP_009060) ZIC3 (NP_003404)	ZIC1 (NP_003403)

His-containing proteins are listed in the left column. Paralogous proteins are listed in the right column. All the members are grouped by protein families. RefSeq IDs are given in brackets.

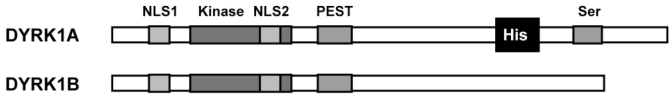
A)

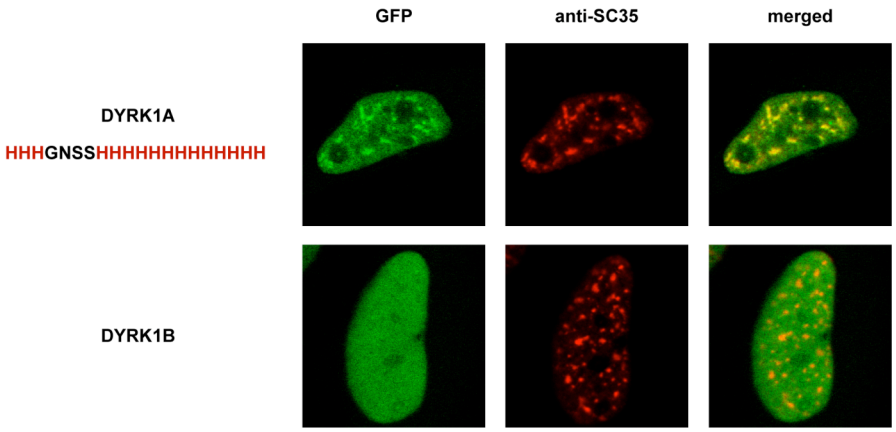
```
FAM26B     4   SALYACTKCTQRYPFEELSQGQQLCKECRIAHPIVKCTYCRSEFQQESKTNTICKKCAQN   63
FAM26A     2   AALYACTKCHQRFPFEALSQGQQLCKECRIAHPVVKCTYCRTEYQQESKTNTICKKCAQN   61

FAM26B    64   VKQFGTPKPCQYCNIIAAFIGTKCQRCTNSEKKYGPPQTCEQCKQQCAFDRKEEGRRKVD  123
FAM26A    62   VQLYGTPKPCQYCNIIAAFIGNKCQRCTNSEKKYGPPYSCEQCKQQCAFDRKDD-RKKVD  120

FAM26B   124   GKLLCWLCTLSYKRVLQKTKEQRKSLGSSHSNSSSSSLTEKDQHHPKHHHHHHHHHRHS  183
FAM26A   121   GKLLCWLCTLSYKRVLQKTKEQRKHLSSSSRAGH----QEKEQ----------------  159

FAM26B   184   SSHHKISNLSPEEEQGLWKQSHKSSATIQNETPKKKPKLESKPSNGDSSSINQSADSGGT  243
FAM26A   160   -----YSRLSGGGHYN--SQKTLSTSSIQNEIPKKKSKFESITTNGDSFSPDLALDSPGT  212

FAM26B   244   DNFVLISQLKEEVMSLKRLLQQRDQTILEKDKKLTELKADFQYQESNLRTKMNSMEKAHK  303
FAM26A   213   DHFVIIAQLKEEVATLKKMLHQKDQMILEKEKKITELKADFQYQESQMRAKMNQMEKTHK  272

FAM26B   304   ETVEQLQAKNRELLKQVAALSKGKKFDKSGSILTSP  339
FAM26A   273   EVTEQLQAKNRELLKQAAALSKSKKSEKSGAI-TSP  307
```

B)

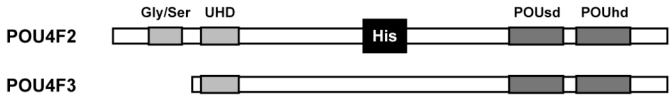

GFP
anti-SC35
merged
FAM76B
FAM76A

C)


NLS1
Kinase
NLS2
PEST
His
Ser
DYRK1A
DYRK1B


GFP
anti-SC35
merged
DYRK1A
HHHGNSSHHHHHHHHHHHHH
DYRK1B

D)


Gly/Ser
UHD
His
POUsd
POUhd
POU4F2
POU4F3

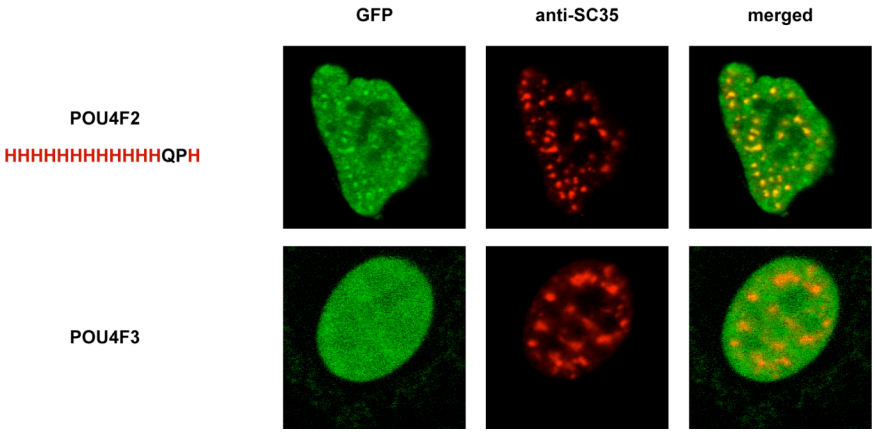

GFP
anti-SC35
merged
POU4F2
HHHHHHHHHHHHQPH
POU4F3

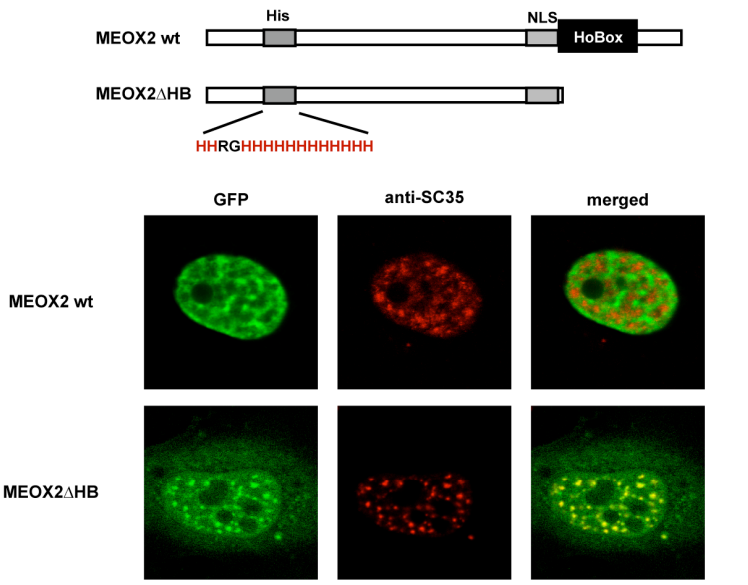
Figure 8: The presence of a His repeat dictates the differential subcellular localization of paralogous proteins. A) Primary sequences alignment of the pair of paralogous FAM76B (NP_653265; hypothetical protein LOC143684) and FAM76A (NP_689873; hypothetical protein LOC199870), obtained with the multiple sequence alignment program Blast2Sequences. His residues in FAM76B are highlighted in red. B) HeLa cells were trasfected with a GFP-expression plasmid encoding FAM76B (upper panel) or FAM76A (lower panel). C and D) HeLa cells were transfected with plasmids expressing GFP fusions with the DYRK family of protein kinases DYRK1A and DYRK1B (C) and the POU family of transcription factors POU4F2 and POU4F3 (D). A schematic representation of each pair of paralogues is presented. A) NLS= nuclear localization signal; Kinase= kinase domain; PEST= PEST sequences; His= His-repeat; Ser= serine-rich region. B) Gly/Ser= segment rich in glycine and serine; UHD= upstream homology domain in POU family members; POUsd= POU specific domain; POUhd= POU homeodomain. In B-D C) cells were processed for immunofluorescence analysis at 48h after transfection. Subcellular localization of the fusion proteins was analyzed by direct fluorescence, and their accumulation in nuclear speckles was followed by immunostaining with anti-SC35 antibody.

2.5 Subcellular localization of polyHis-containing proteins depends on other domains present in the proteins

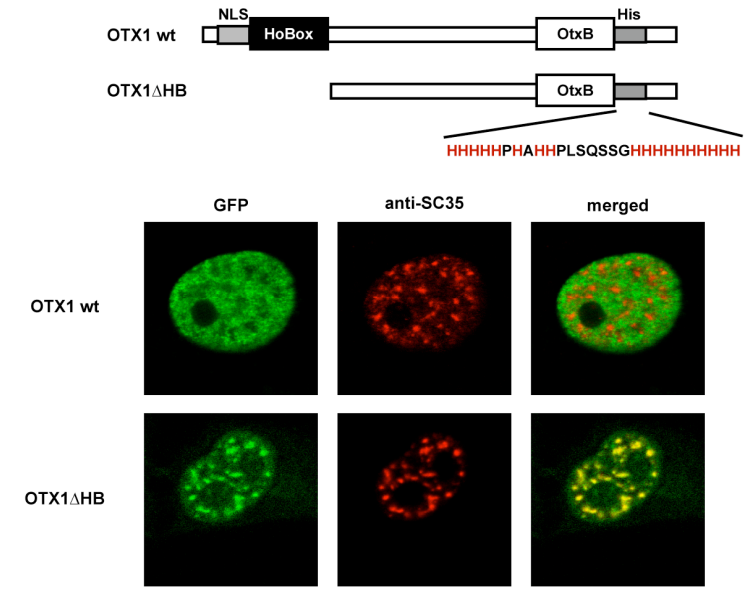
The nuclear localization analysis of the chosen His-containing proteins revealed that some of them did not accumulate in the SFC. In general, these proteins contained other domains in their primary sequence, such as DNA-binding and/or protein-protein interaction domains. For instance, the transcription factors MEOX2 and OTX1 harbour a homeobox DNA-binding domain at their C- and N-terminal regions, respectively (Figure 9). Therefore, the hypothesis suggesting that the SFC accumulation of these proteins might be influenced by the presence of this interacting domain was formulated. Deletion mutants skipping the DNA binding domain of MEOX2 and OTX1 were generated, and their subcellular localization assessed. Both wild type proteins presented the typical dispersed staining for most transcription factors, compatible with active transcription sites (Wei et al., 1999) (Figure 9A and B, upper panels). However, the mutant proteins in which the homeobox domain was eliminated, MEOX2ΔHB and OTX1ΔHB, showed a complete co-localization with SC35-positive speckles (Figure 9A and B, lower panels). Moreover, the different localization of the His-tract in both proteins, N-terminal for MEOX2 and C-terminal for OTX1, suggests that position does not influence on the targeting of the proteins to the nuclear speckles.

The CtBP2 and RING1-interacting domains in the C-terminal region of CBX4 mediate the accumulation of this protein in subnuclear foci known as polycomb bodies (Kagey et al., 2005). A CBX4 deletion mutant without these two regions was generated, and its subcellular localization assessed. Accordingly, the wild type protein presented a nuclear staining with foci that were not positive for SC35 (Figure 9C). However, the mutant protein in which the two domains were not present clearly co-localized with SC35 (Figure 9C). Altogether, these results confirm that some of the polyHis-containing proteins could accumulate in nuclear speckles depending on their ability to be targeted to other subnuclear compartments such as active chromatin or other nuclear bodies and suggest, therefore, that competition between distinct protein regions dictates the subnuclear localization of the proteins at the steady state.

A)


His
NLS
MEOX2 wt
HoBox
MEOX2ΔHB
HHRGHHHHHHHHHHHH
GFP
anti-SC35
merged
MEOX2 wt
MEOX2ΔHB

B)


NLS
His
OTX1 wt
HoBox
OtxB
OTX1ΔHB
OtxB
HHHHHPHAHHPLSQSSGHHHHHHHHHH
GFP
anti-SC35
merged
OTX1 wt
OTX1ΔHB

C)

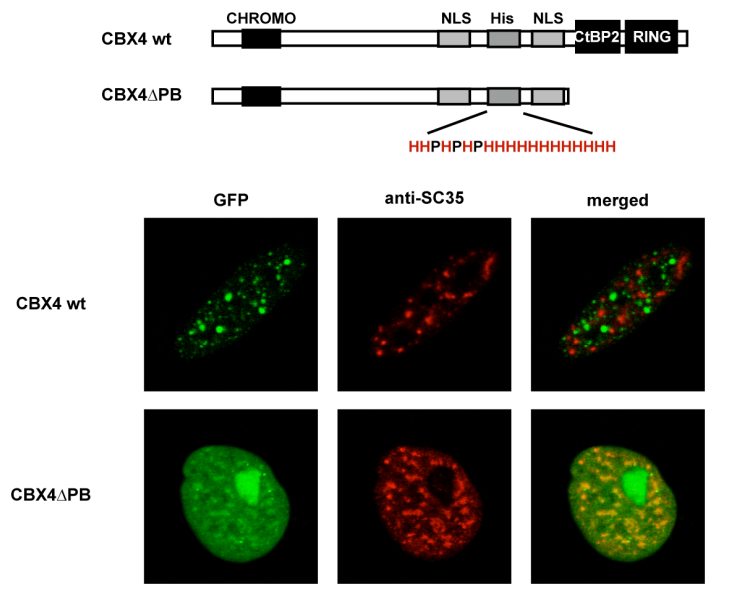
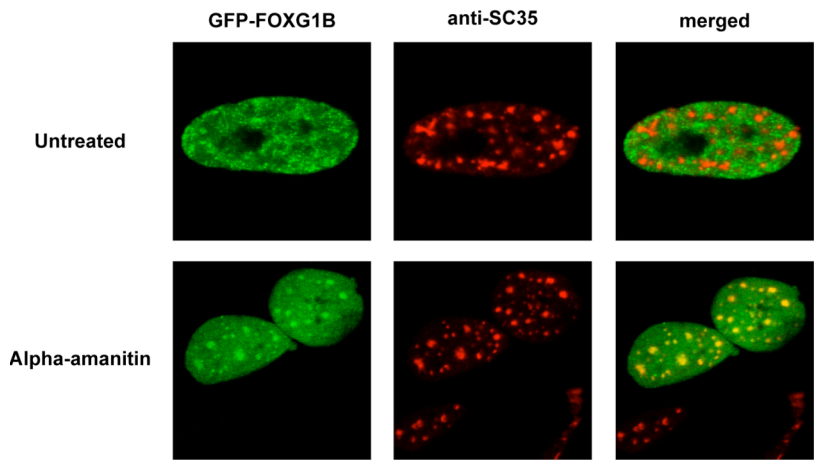

CHROMO
NLS
His
NLS
CBX4 wt
CtBP2i
RING
CBX4ΔPB
HHPHPHPHHHHHHHHHHH
GFP
anti-SC35
merged
CBX4 wt
CBX4ΔPB

Figure 9: The accumulation in nuclear speckles of some transcription factors with polyHis runs depends on their interaction with DNA and/or other proteins. HeLa cells were transfected with the expression plasmids for GFP-MEOX2 wild type or the mutant GFP-MEOX2ΔHB (A), GFP-OTX1 wild type or GFP-OTX1ΔHB (B) and GFP-CBX4 wild type or GFP-CBX4ΔPB (C). At 48 h post-transfection, the nuclear localization of both proteins was analyzed by direct fluorescence (left columns, green) and their co-localization with nuclear speckles by immunostaining with anti-SC35 antibody (middle columns, red). See scheme: His: His repeat; NLS: nuclear localization signal; HoBox: homeobox domain; OtxB: Otx box; CtBP2 and RING1: CtBP2 and RING1- interacting domains.

2.6 Transit through nuclear speckles is a dynamic property of polyHis-containing proteins

In mammalian cell lines the structure and function of speckles is sensitive to the transcriptional state of the cell (reviewed in Lamond and Spector, 2003). When cells are treated with RNA polymerase II transcription inhibitors, there is a decrease in the splicing activity and a redistribution of SFC components, which are recruited to larger and rounded-up nuclear speckles (Zeng et al., 1997). Considering that most of the His-containing proteins were transcription factors and that DNA binding activity influenced speckle accumulation, it was wondered whether the absence of SFC accumulation could be reverted by using the inhibitor of RNA polymerase II alpha-amanitin. Figure 10 shows the results obtained with the transcription factors FOXG1B, HOXA1 and HOXA9. The three proteins showed nuclear staining, but no accumulation in nuclear speckles in basal conditions (Figure 10A, B and C, untreated). Upon alpha-amanitin treatment, SC35 redistributed in less but larger speckles as expected (Figure 10A, B and C, anti-SC35 panels). Interestingly, the three transcription factors also redistributed from their diffuse nucleoplasmic staining to a punctuated one, where they overlapped with SC35 (Figure 10A, B and C). To confirm the requirement of the His-repeat for this relocalization, a mutant HOXA1ΔHis was prepared and the same type of experiments performed. In this case, the treatment with alpha-amanitin failed to cause the accumulation of the mutant protein in nuclear speckles (Figure 10D).

A)



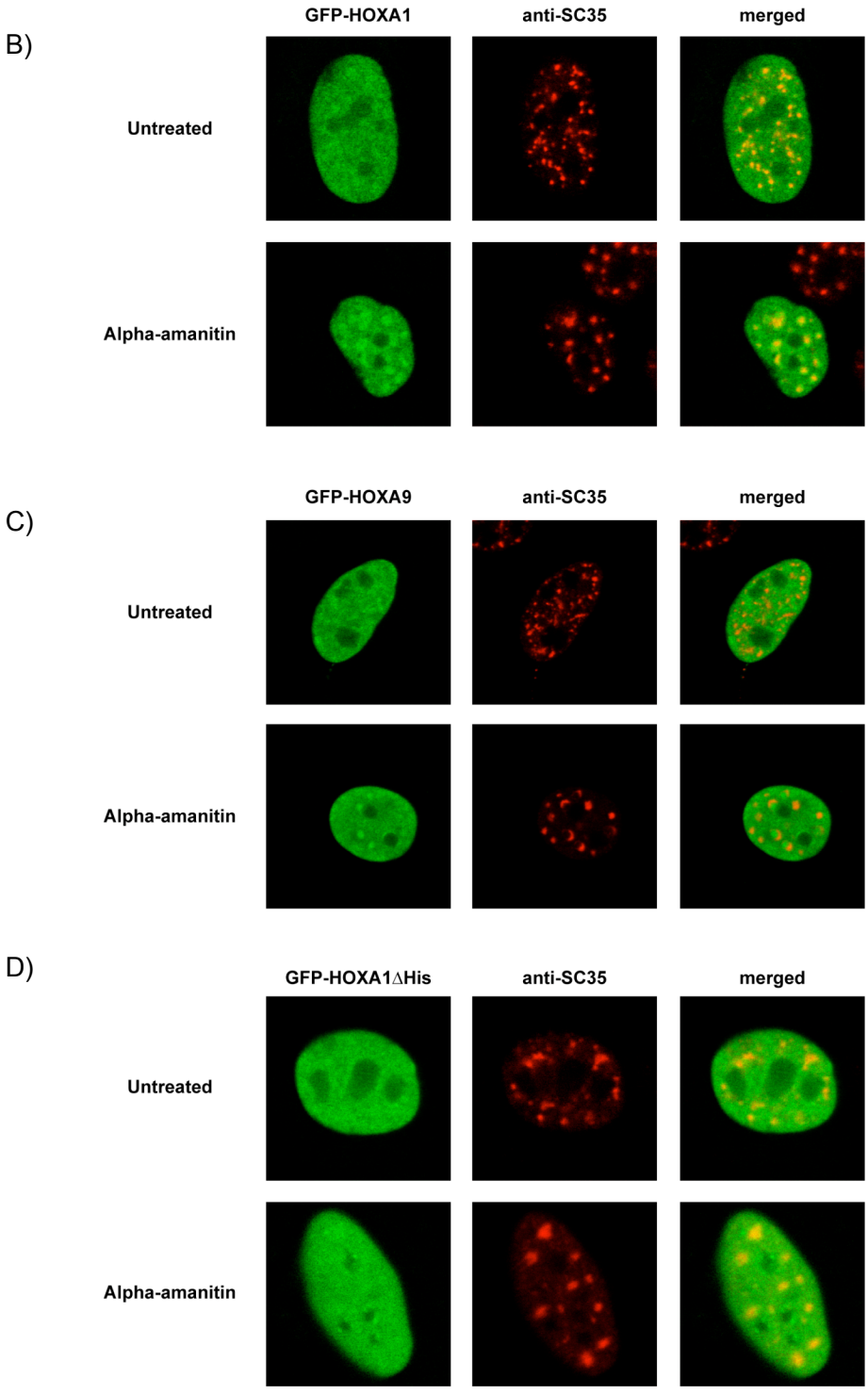


Figure 10: **The transcriptional state of the cell determines the accumulation in the SFC of some polyHis-containing transcription factors.** HeLa cells were transfected with the expression plasmids encoding GFP-FOXG1B (A), HOXA1 (B), HOXA9 (C) and (D) HOXA1ΔHis. At 48 h post-transfection, cells were treated with alpha-amanitin for 5 h to inhibit transcription, and processed for immunofluorescence with anti-SC35. The progressive enlargement of the speckles was used as a control of the treatment.

Finally, the dynamic changes in the localization of HOXA1 were also analyzed by *in vivo* cell imaging that showed a quick relocalization of this protein in response to the RNA

polymerase II inhibitor (video images at http://davinci.crg.es/speckles_GFP-HOXA1_a.mov and http://davinci.crg.es/speckles_YFP-SC35_a.mov).

The distribution of the signalling molecule PRICKLE (also known as LMO6) was also analyzed. The interest on this protein stems from its exclusive cytosolic staining, with total exclusion of the nucleus (Figure 11A, untreated cells). However, after alpha-amanitin treatment, a small proportion of the protein appeared within the nucleus, where it clearly co-localized with SC35 (Figure 11A, treated cells). This observation suggests that, although being predominantly cytoplasmic, PRICKLE3 has the ability to shuttle in and out of the nucleus. As shown in Figure 11B, cell treatment with leptomycin B, an inhibitor of CRM1-dependent nuclear export, caused PRICKLE3 complete re-localization to the nucleus. This observation confirms that PRICKLE3 has a nuclear phase and further supports the results obtained with the alpha-amanitin treatment.

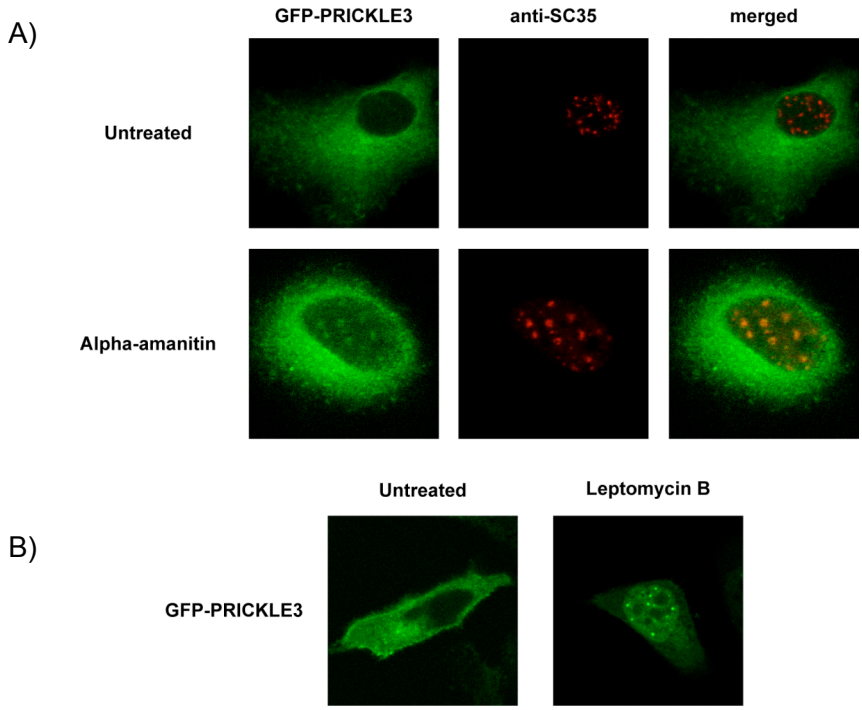


Figure 11: PRICKLE6 confirms the ability to shuttle of some His-containing proteins. HeLa cells expressing GFP-PRICKLE3 fusion protein were mock-treated or treated with alpha-amanitin (A) or leptomycin B (B) for 5 h, at 24 h after transfection. The subcellular localization of the fusion protein was analyzed by direct fluorescence. Note that PRICKLE3 is detected in the cytosol in untreated cells, but accumulates in the nucleus, nucleoplasm and nuclear speckles in response to the nuclear export inhibitor.

All the results obtained in the analysis of the subcellular localization of 22 polyHis-containing proteins are summarized in Table 3. Fifteen out of 22 analyzed showed a nuclear staining compatible with accumulation in nuclear speckles, indicating that proteins with His repeats have the property to dynamically localize in the splicing factor compartment.

Table 3: Summary of the results obtained in the subcellular localization analysis of polyHis-containing proteins.

Name	His tract	Protein domains	Function
Cyclin T1	513-HPSNHHHHHNHHSHKHSH-530	cdk binding domain	Transcription regulator
POU4F2	172-HHHHHHHHHHHHQPH-186	POU domain (254-328) Homeobox (346-405)	Transcription factor Differentiation and survival of retinal ganglion cells
YY1	65-HGHAGHHHHHHHHHHH-80	Zinc finger (296-320; 325-347;353-377;383-407)	Transcription factor Regulation of development and differentiation
DYRK1A	590-HHHHGNSSHHHHHHHHHHHHH-610	Kinase domain (159-479)	Ser/Thr protein kinase Regulator of cell proliferation and differentiation
NLK	14-HHHHHHHHLPHLPPPHLHHHHHPQHHLH-42	Kinase domain (126-415)	Ser/Thr protein kinase Regulator of Wnt-signaling pathways
FAM76B	167-HHPKHHHHHHHHHHRHSSSHH-187	Not found	Unknown
GSH2	124-HAH HHHHPPQHHHHHH-139	Homeobox (203-261)	Transcription factor Telencephalic development
HOXA1	65-HHHHHHHHHH-74	Homeobox (229-291)	Transcription factor Hindbrain segmentation
HOXA9	84-HHHHHH-89	Homeobox (207-267)	Transcription factor Positional identity on the anterior/posterior axis
MEOX2*	64-HHRGHHHHHHHHHHH-79	Homeobox (186-248)	Transcription factor Somite development
OTX1*	275-HHHHHPHAHHPLSQSSGHHHHHHHHHH-301	Homeobox (36-96) Otx-box (247-274)	Transcription factor Brain development
HAND1	8-HHHHHHHPHPAH-20	Helix-loop-helix (103-151)	Transcription factor Cardiac morphogenesis
CBX4*	380-HHPHPHPHHHHHHHHHHH-398	Chromodomain (16-69) CtBP2-interacting domain (470-475) RING2-interacting domain (540-558)	Chromatin modification SUMO E3-ligase
FOXG1B	33-HHASHGHHNSHHPQHHHHHHHHHH-57	Fork-head domain (179-269) PLU-1-interacting domain (375-411) FAST2-intracting domain (314-372)	Transcription factor Regulator of telencephalon morphogenesis
PRICKLE3	513-HHHHNHHHHHNRH-525	PET domain (73-178) LIM domain (186-243)	Unknown
DLX2	309-HHHHHHH-315	Homeobox (157-210)	Transcription factor Forebrain differentiation
POU4F1	100-HHHHHHHHH-108	POU domain (279-291) Homeobox (306-319) Homeobox (389-412)	Transcription factor Differentiation and survival of sensory neurons
ZIC3	87-HHHHHHHHHHH-97	Zinc finger (300-322; 328-352; 358-382; 388-410)	Transcription factor Determination of left-right asymmetry
ONECUT1	124-HHHHHHHHHHHPHH-138	CUT domain (283-369) Homeobox (385-477) CREB-interacting domain (327-331)	Transcription factor Pancreas specification
MAFA	184-HHHGAHHAAHHHHAAHHHHHHHHHHSHGGAGHGGGAGHH-219	Maf-N (111-145) Basic leucine zipper (253-316)	Transcription factor Regulator of insulin gene expression
MAFB	131-HHHHHHHHHPHPHHAYPGAGVAHDELGPHAHPHHHHHH-167	Maf_N (80-114) Basic leucine zipper (209-303)	Transcription factor Regulator of lineage-specific haematopoiesis
MEC2P	366-HHHHHHH-372	Methyl-CpG-binding domain (94-168)	Transcription repressor

White rows = proteins that accumulate in nuclear speckles under basal conditions; clear grey rows = proteins that accumulate in nuclear speckles after alpha-amanitin treatment; *deletion of DNA-binding/protein-protein interaction motifs was also tested; grey rows = proteins that do not localize in speckles.

3. Molecular characterization of the His run in DYRK1A

Once confirmed that His repeats may act as a general SFC-targeting signal, the question was how the amino acid run was performing this function. Because of the nature of nuclear speckles components, the ability of His runs to target proteins to SFC could be mediated either by binding to nucleic acids, such as the RNA present in the speckle, or by protein-protein interactions with resident proteins. In fact, the physicochemical properties of His make it a suitable amino in both scenarios. It is described that the accumulation of His residues in an extended conformation, such as a β-strand, displays a positive charge or charge gradient that could mediate protein-protein, protein-DNA or protein-substrate interactions, or even orient protein subunits via electrostatic interactions (Karlin et al., 2003).

In the case of nucleic acids, His could mediate direct binding to DNA or RNA through hydrogen bonds from its side chains to phosphate groups. Moreover, His residues are found in Zn-fingers coordinating Zn ions. Zn-fingers are functional domains implicated in the interaction with nucleic acids (Laity et al., 2001) and present in many families of DNA and RNA binding proteins (Gamsjaeger et al., 2007).

His runs has also been described as a protein interaction surface in the case of the transcriptional regulator cyclin T1 with the carboxyl terminal domain (CTD) of the RNA polymerase II, the transcriptional regulator Human I-mfa Domain-Containing protein (HIC) and the growth factor granulin (Hoque et al., 2003; Taube et al., 2002; Young et al., 2003). In addition, DYRK1A interacts through the His repeat with Sprouty2 (Aranda et al., 2008) and the transcription factor Brn-3b (see Section 7 of this work). Some of these reports demonstrate the possibility that His residues participate in intermolecular Zn-finger structures through Zn atoms coordination (Aranda et al., 2008; Young et al., 2003).

To approach these questions, and trying to find out how His-tracts work as speckles recruiting elements, DYRK1A was used as a model system. The results obtained will be presented in the following sections.

4. DYRK1A binds nucleic acids

4.1 The background

We had some indication from previous results obtained by M. Alvarez suggesting that DYRK1A could bind nucleic acids. When GFP-DYRK1A-expressing cells were treated with DNAse I (Alvarez, 2004), the dot-like staining of nuclear DYRK1A disappeared, while SC35-positive speckles remained untouched (Figure 13, adapted from Alvarez, 2004).

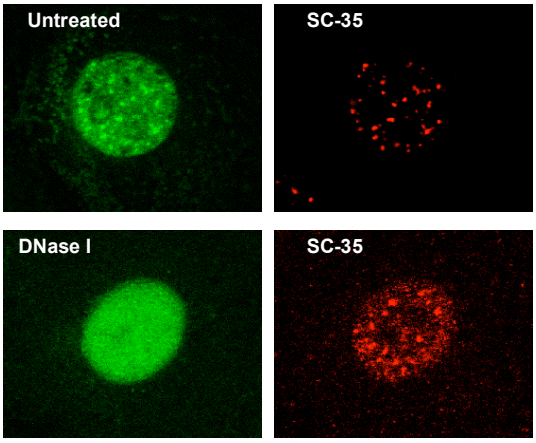


Figure 13: The SFC accumulation of DYRK1A is sensitive to DNAse I treatment. Immunofluorescence analysis of GFP-DYRK1A-overexpressing COS-7 cells, in which permeabilized cells are treated with DNAse I before fixation. This treatment leaves the SFC untouched as seen by the integrity of the SC-35 staining (Spector et al., 1991), but causes a change in DYRK1A accumulation (Alvarez, 2004 #305).

Since there is no DNA within the SFC (Spector et al., 1991), the explanation for the lost of DYRK1A staining is not straightforward. One possibility is that the digestion with DNAse generates a pool of putative binding sites for DYRK1A, which would displace it from its localization in nuclear speckles. This would favor the hypothesis of an interaction between DYRK1A and DNA, maybe established through the His stretch.

4.2 DYRK1A interacts with single-stranded nucleic acids

To address this question, a DNA-affinity chromatography was carried out with soluble extracts of U2-OS cells expressing HA-DYRK1A. Single and double stranded DNA coupled to cellulose matrix were used; free cellulose was also used as a control of unspecific binding. As shown in Figure 14, DYRK1A bound to single stranded DNA (ssDNA) but not to double stranded DNA (dsDNA) or cellulose.

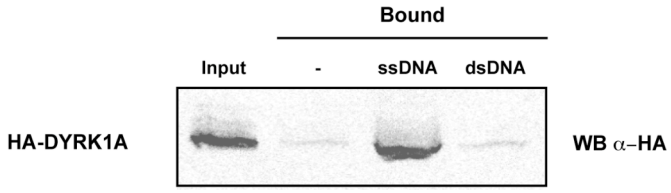


Figure 14: DYRK1A specifically binds to ssDNA. Soluble extracts from U2-OS cells expressing HA-DYRK1A were incubated with equal amounts of cellulose (-), ssDNA-cellulose (ssDNA) or dsDNA-cellulose (dsDNA) for 1 h at 4ºC. Matrices were repeatedly washed, resuspended with sample buffer and analyzed by Western blot with anti-HA.

The ability of DYRK1A to bind ssDNA suggested that it could also interact with RNA. Therefore, we tested DYRK1A retention in poly(U)-Sepharose columns. The enrichment of DYRK1A in the RNA-bound fraction (Figure 15) confirms the ability of the kinase to bind single strand nucleic acids, both DNA and RNA.

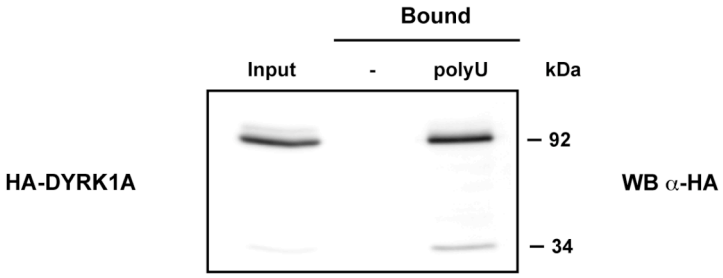


Figure 15: DYRK1A binds to RNA. Soluble extracts from HA-DYRK1A-expressing U2-OS cells were incubated with cellulose (-) or poly(U)-Sepharose for 1 h at 4ºC. Matrices were washed, resuspended in sample buffer and analyzed by Western blot with anti-HA antibody. Input represents 10% of the sample. The position of the marker proteins (in kDa) is indicated.

4.3 The interaction of DYRK1A with DNA is direct

The interaction between DYRK1A and DNA could be either direct or through other DNA-binding proteins present in the cell extracts. To test this possibility, we used bacterially purified DYRK1A protein. As shown in Figure 16, DYRK1A could interact with ssDNA, indicating that directly DYRK1A binds nucleic acids.

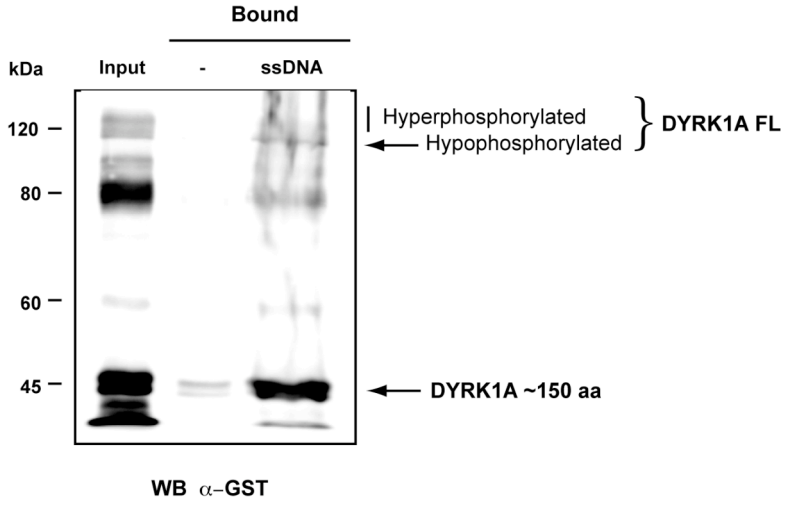


Figure 16: DYRK1A binds directly to DNA. Purified GST-DYRK1A (1 μg) was incubated 1 h at 4ºC with equal amounts of cellulose (-) or ssDNA-cellulose (ssDNA). The bound protein was visualized by Western blot with anti-GST antibody. Arrows indicate full-length DYRK1A (FL) and a truncated product of about 150 amino acids. Hyper and hypophosphorylated forms of DYRK1A are also marked. Input corresponds to 10% of the total sample. The position of the marker (in kDa) is indicated.

A closer observation of the results presented in Figure 16 allows proposing a few more conclusions. First, as DYRK1A autophosphorylates itself, the bacterially expressed full-length protein always appears as a doublet of both hypo- and hyperphosphorylated forms (Figure 16, black arrow). In the DNA-bound complexes, the enrichment in the higher mobility bands is more apparent, suggesting that hypophosphorylated DYRK1A interacts preferentially with DNA. The increase in negative charges may induce an electrostatic repulse towards the nucleic acid preventing the interaction. Second, binding of a short N-

terminal fragment is clearly detected (Figure 16, bottom arrow). It is worth mentioning that a similar result was obtained when binding to polyU was tested (see Figure 15). These results suggest that the nucleic acid-binding region of DYRK1A is very likely located within the first 150 amino acid, and furthermore that the His repeat might not participate in the interaction as initially proposed.

4.4 DYRK1A binding to single-stranded nucleic acid is not dependent on the His repeat

To test whether the His run was necessary for the binding of DYRK1A to DNA/RNA, cell extracts expressing a HA-DYRK1A mutant lacking the His repeat (ΔHis) were used for the nucleic acid-affinity chromatography. As shown in Figures 17B and C, the mutant DYRK1AΔHis interacted with single-stranded nucleic acids as wild type DYRK1A did. It is worth to note again that enrichment in a shorter DYRK1A C-terminally deleted fragment in the bound fraction is clearly apparent (see around marker 34 kDa in Figure 17C). Thus, this result rules out the His repeat as a DNA or RNA binding domain.

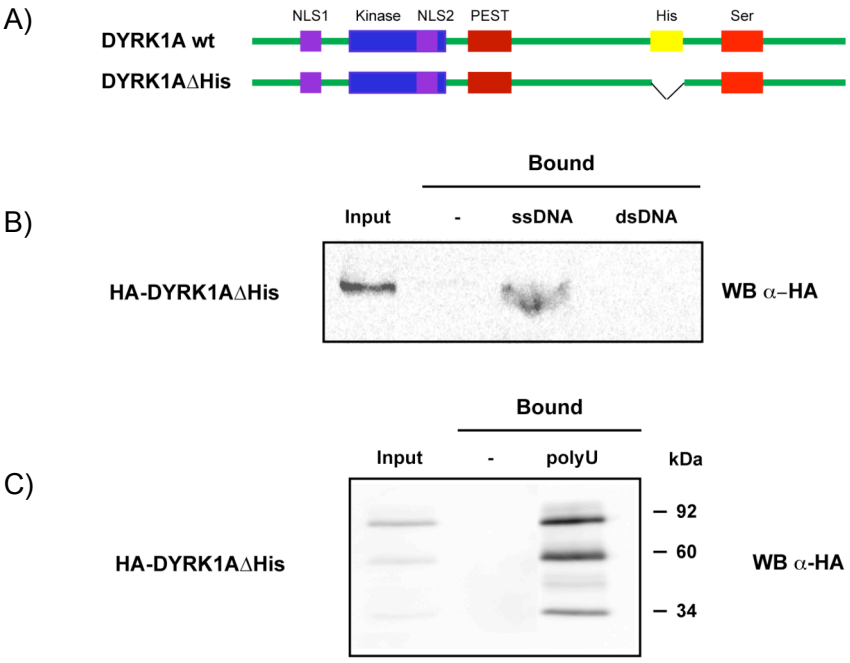


Figure 17: DYRK1A binding to DNA is not mediated by the His repeat. A) Scheme showing the proteins used. NLS: nuclear localization signal; His: His repeats; Ser: Ser/Thr rich region. Soluble extracts from U2-OS expressing HA-DYRK1AΔHis were incubated with equal amounts of cellulose (-), ssDNA-cellulose (ssDNA), dsDNA-cellulose (dsDNA) (B) or poly(U)-sepharose (C) for 1 h at 4ºC. After several washes, matrices were resuspended in sample buffer, resolved by SDS-PAGE and analyzed by Western blot with anti-HA antibody. Input represents 10% of the sample.

4.5 The interaction with DNA is conserved in the homologous protein DYRK1B

DYRK1A and DYRK1B are the closest members of the DYRK family in mammals. Both protein kinases are nuclear (Becker et al., 1998; Leder et al., 1999), due to the existence of the two conserved NLSs (Alvarez, 2004; Alvarez et al., 2003) (Figure 18A). However, whereas DYRK1A accumulates in speckles, DYRK1B localizes diffusely over the

nucleoplasm (Figure 8C). These differences in subnuclear distribution are due to the lack of the His repeat in DYRK1B. To find out whether the ability to bind DNA was conserved in DYRK1B, soluble cell extracts overexpressing HA-DYRK1B were incubated with cellulose or ssDNA cellulose-coupled matrices. Similarly to DYRK1A, DYRK1B was present in the DNA-bound fraction (Figure 18B). The result further supports that the His run is not involved in the interaction with nucleic acids, and suggest that the interacting region should be located in a conserved region between DYRK1A and DYRK1B.

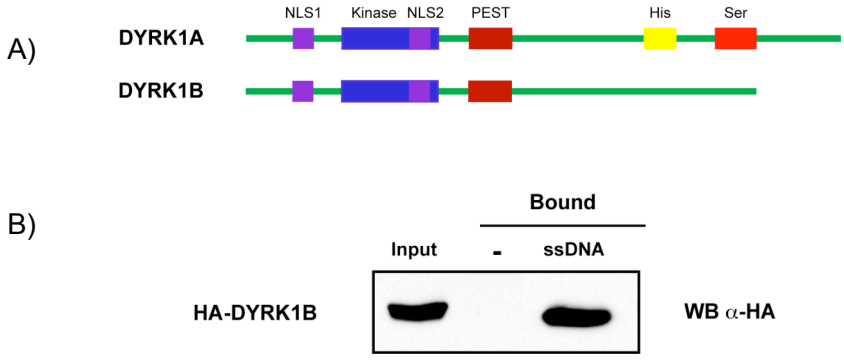


Figure 18: DYRK1B binds to DNA. A) Scheme comparing sequence similarities between DYRK1A and DYRK1B. B) Soluble extracts from of U2-OS cells expressing HA-DYRK1B were incubated with cellulose (-) or ssDNA (ssDNA) matrices for 1 h at 4ºC and the bound protein analyzed by Western blot with anti-HA antibody. Input represents 10% of total extract.

4.6 The DNA-binding domain of DYRK1A is located in the N-terminal region of the protein

The presence of a protein band of smaller size in the nucleic acid bound fractions (Figures 15-17), helped in narrowing the DNA-interacting region in DYRK1A. In the GST-DYRK1A bacterial preparations, several C-terminal deleted proteins co-exist with the full-length protein, very likely due to truncated synthesis or proteolysis. As Figure 16 shows, all the shortened forms were still able to bind DNA. The smallest one, which contained approximately the first N-terminal 150 amino acids, also bound DNA as efficiently as the full-length protein (see band around 45 kDa marker in Figure 16). This observation suggests that the DNA-binding domain in DYRK1A is located within the N-terminal region of the protein.

To confirm this hypothesis, DNA-binding experiments were performed using a DYRK1A mutant that spans the first 167 amino acids, expressed as a GFP fusion protein (Figure 19A). This protein segment bound DNA as efficiently as full-length DYRK1A (Figure 19B).

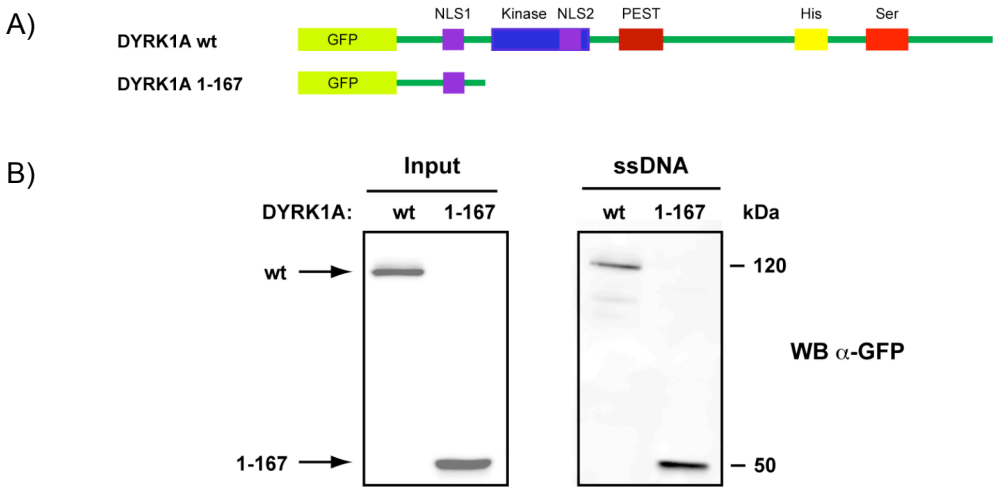


Figure 19: The N-terminal region of DYRK1A interacts with DNA. A) Scheme showing the proteins used in panel B. B) Soluble extracts of U2-OS cells expressing GFP-DYRK1A full length (wt) or the first 167 amino acids (1-167) were incubated with ssDNA-affinity matrices for 1 h at 4ºC and the bound protein was detected by Western blot with an anti-GFP antibody. Input represents 10% of total extract.

Since the homology between DYRK1A and DYRK1B is maintained at the N-terminal region of the proteins, we wondered whether binding to DNA in DYRK1B mapped also within its N-terminus. To this end, a DNA-binding assay with a DYRK1B N-terminal fragment (1-128) equivalent to that used for DYRK1A (Figure 20A) was carried out. As shown in Figure 20B, the result indicated that DYRK1B/1-128 was as able to bind ssDNA as the full-length protein, and allows concluding that the region that mediates the interaction with single strand nucleic acids is conserved in both DYRK1A and DYRK1B kinases.

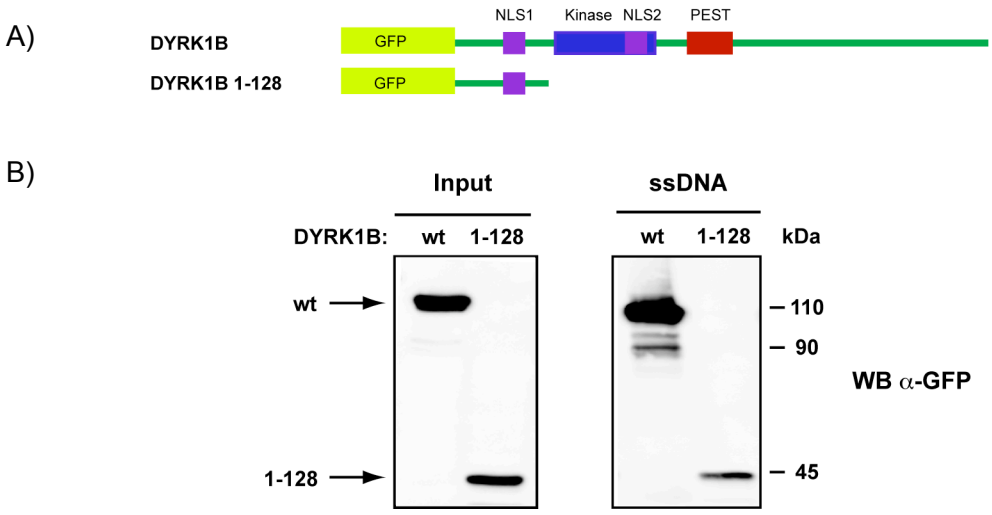


Figure 20: The DNA-binding domain is also restricted to the N-terminal part of the protein in DYRK1B. A) Scheme showing proteins used in panel B. B) Soluble extracts of U2-OS expressing GFP-DYRK1B full length (wt) and an N-terminal fragment (1-128) were incubated with ssDNA-affinity columns for 1 h at 4ºC and the bound protein was detected by Western blot with an anti-GFP antibody. Input represents 10% of total extract.

4.7 DYRK1A NLS_1 is not involved in the interaction with nucleic acids

The DYRK1A fragment spanning amino acids 1-167 contains one of the DYRK1A nuclear localization signals (NLS_1). The NLS_1 (amino acids 105 to 125) is a bipartite NLS responsible for most of the nuclear import of the kinase (Alvarez et al., 2003). Moreover, the homolog fragment in DYRK1B (1-128) also contains a NLS that mediates the nuclear localization of the protein (Alvarez, 2004). It has been described that in 90% of the DNA-interacting proteins, their NLSs act both as nuclear directing sequences and as DNA binding regions (Cokol et al., 2000). Thus, it was plausible to think that DYRK1A and DYRK1B NLS_1 were mediating the interaction of the kinases with nucleic acids. To test this, the DNA binding experiments were carried out with a DYRK1A mutant protein in which the NLS_1 was deleted (amino acids 105-125) (Figure 21A). As shown in Figure 21B, both DYRK1A wt and mutant forms were able to bind both ssDNA and poly(U), indicating that the bipartite NLS_1 is not mediating the interaction between DYRK1A and single stranded nucleic acids.

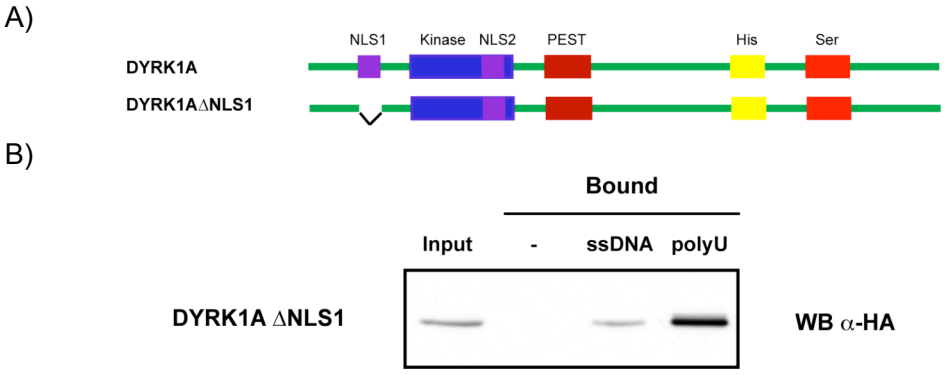


Figure 21: DYRK1A NLS_1 does not mediate the interaction with single-stranded nucleic acids. A) Scheme representing the constructs used in the analysis, with the deletion of the entire NLS_1 (residues 105-125). B) Soluble extracts of U2-OS cells expressing HA-DYRK1A ΔNLS_1 were incubated ssDNA or poly(U)-sepharose for 1 h at 4ºC. Bound protein were detected by Western blot with anti-HA antibody. Input represents 10% of total extracts

In summary, the results obtained in this section demonstrate the ability of DYRK1A to interact with single-stranded nucleic acids, both DNA and RNA. The interaction between DYRK1A and nucleic acids is direct and mediated by a region located within the first 167 amino acids of the kinase. This region is conserved in the homologous kinase DYRK1B, where it maintains its role as nucleic acid-binding sequence. Finally, the DYRK1A His repeat is not involved in binding to single stranded nucleic acids and, therefore, targeting of the kinase to nuclear speckles should not be due to RNA binding.

5. DYRK1A binds the CTD of the RNA polymerase II

5.1 The hypothesis

In the search of speckle-resident proteins that could act as a DYRK1A-anchoring point, the RNA polymerase II turned out to be a good candidate. The RNA polymerase II large subunit has a distinct domain at its C-terminus known as C-terminal domain (CTD), which is formed by several repetitions of the heptapeptide YSPTSPS (Corden et al., 1985). CTD physical and chemical properties turn it into a surface of interaction with proteins involved in RNA synthesis and processing (reviewed in Cho, 2007; Egloff et al., 2007). Thus, it is thought to play a central role in the coupling process by which pre-mRNA processing occurs co-transcriptionally (Maniatis and Reed, 2002; Meinhart et al., 2005). The protein domain activities are finely regulated through reversible phosphorylation of serine residues within the heptapeptide, which generates different populations of RNA polymerase II: hypo- and hyperphosphorylated. The phosphorylation events occur sequentially during the transcriptional process (reviewed in Hirose and Ohkuma, 2007). In fact, the phosphorylation of Ser^2 or Ser^5 of the consensus heptapeptide has different functional consequences, and therefore it is used as a tool to place a kinase function along the transcriptional process (Prelich, 2002): it is broadly accepted that Ser^5 kinases are implicated in the initiation process, while Ser^2 kinases are related to the elongation step (Prelich, 2002). One of these CTD-kinases is known as Positive Elongation Factor b (P-TEFb), a protein complex comprised by the transcriptional regulator cyclin T1 and cyclin-dependent kinase 9 (CDK9) (Bres et al., 2008; Marshall et al., 1996). As part of the complex PTEF-b, cyclin T1 is necessary for the processive elongation of the RNA polymerase II during RNA transcription (Price, 2000). Interestingly, cyclin T1 localizes in nuclear speckles and interacts with the RNA polymerase II CTD through its His-repeat (Alvarez et al., 2003; Taube et al., 2002).

Considering that RNA polymerase II itself accumulates in nuclear speckles (Bregman et al 1995; Mortillaro et al., 1996) and that most of the proteins containing polyHis segments and able to accumulate in the SFC were transcription factors, we hypothesized that the CTD of RNA polymerase II may represent the recruiting element for the His-repeat containing proteins for targeting to the nuclear speckles.

5.2 DYRK1A interacts with the CTD of the RNA polymerase II *in vitro*

To test whether DYRK1A could interact with the CTD, the approach used by M. Peterlin's group (Taube et al., 2002) to study the interaction between cyclin T1 and the CTD was chosen as a model. Several control proteins were included in the experiments, such as cyclin T1 and cyclin L2, which were taken as positive and negative controls, respectively (Figure

22A). The latest only binds to the hyperphosphorylated form of the CTD (Berke et al., 2001), and given that proteins synthesized in bacteria do not contain phosphorylated residues, cyclin L2 should be unable to interact with a bacterially expressed CTD. Cyclin T1 and L2 behaved as expected in CTD-pull-down assays (Figure 22B). Interestingly, DYRK1A interacted with the CTD of the RNA polymerase II even more efficiently than cyclin T1 (Figure 22C, left).

Then, DYRK1A kinase activity requirement for the interaction was tested. For that, the ability to bind the CTD of the DYRK1A kinase-dead form K179R was analyzed. As shown in Figure 22C, both proteins bound the CTD equally, suggesting that the interaction between DYRK1A and the CTD is independent of the kinase activity of the protein.

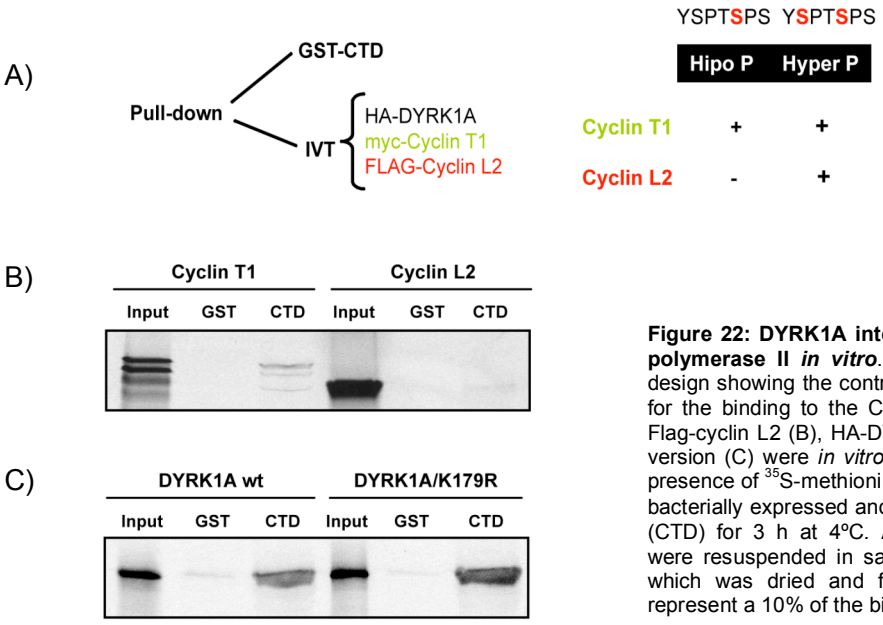


Figure 22: DYRK1A interacts with the CTD of the RNA polymerase II *in vitro*. A) Scheme of the experimental design showing the control proteins and their requirements for the binding to the CTD. B and C) Myc-cyclin T1 (B), Flag-cyclin L2 (B), HA-DYRK1A wt (C) and K179R mutant version (C) were *in vitro* transcribed and translated in the presence of ^{35}S-methionine and then incubated with 5 μg of bacterially expressed and purified GST (GST) or GST-CTD (CTD) for 3 h at 4ºC. After extensively washing, beads were resuspended in sample buffer and loaded in a gel, which was dried and finally exposed to a film. Inputs represent a 10% of the binding.

5.3 Mapping of the CTD-interacting domain in DYRK1A

To identify the region or regions in DYRK1A responsible for the binding to the polymerase, a set of DYRK1A deletion was used (Figure 23A). Almost all deletion mutants, including the mutant with a deletion in the His repeat, presented an equivalent binding to the CTD as DYRK1A wt did (Figure 23B). Only the shortest mutant (1-377) hardly bound to the CTD. The results demonstrate that the His repeat of DYRK1A is not involved in the interaction with the CTD. Furthermore, they suggest that the region responsible for the interaction with the CTD of the RNA polymerase II is comprised between amino acids 377 and 474.

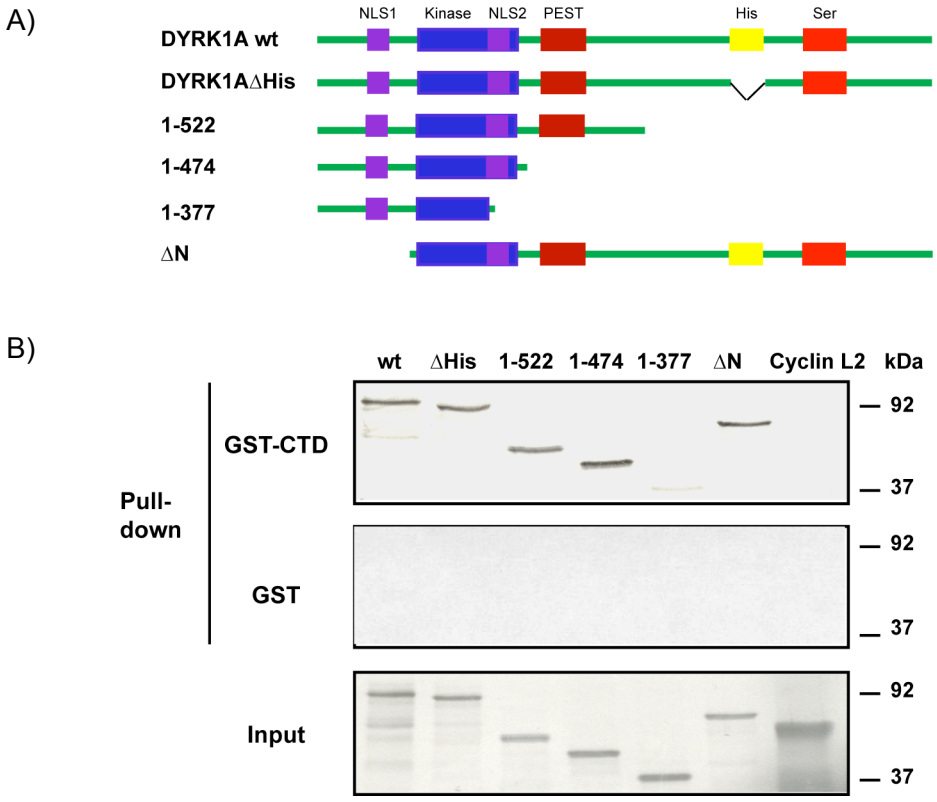


Figure 23: The CTD-interacting region in DYRK1A may map within the kinase domain. A) Schematics of the DYRK1A deletion mutants used in the pull-down assay. The main features are shown. B) Pull-down assays were performed by incubating *in vitro* translated HA-DYRK1A deletion mutants and cyclin L2 with 5 μg of GST or GST-CTD for 3 h at 4ºC. Beads were then washed and the binding was analyzed by autoradiography. No binding was detected to the GST beads (middle panel). Input (bottom panel) represents 10% of total extracts.

This result clearly demonstrates that the interaction between both proteins does not depend on the His-rich domain of DYRK1A. However, the possibility of DYRK1A being a CTD kinase prompted us to further explore the interaction.

5.4 DYRK1A interacts with the RNA polymerase II *in vivo*

First, the interaction between DYRK1A and the RNA polymerase II was investigated. For this, a co-immunoprecipitation assay using nuclear extracts of HEK-293T cells transfected with a HA-DYRK1A expressing plasmid was performed. Non-transfected cells and cells transfected with a plasmid expressing full-length Myc-cyclin T1 were used as negative and positive controls, respectively. RNA polymerase II was immunoprecipitated with N20, an antibody that recognizes the N-terminal region of the large subunit, or with rabbit IgGs as a control of specificity. As shown in Figure 24, DYRK1A was detected in the N20 immunocomplexes, but not in the rabbit IgG ones, indicating that the kinase is present in RNA polymerase II complexes *in vivo*. It is worth to note that the ratio between input and protein present in the immunocomplexes for DYRK1A is very similar to that of cyclin T1. Since the N20 antibody

recognizes all RNA polymerase II species, it cannot be distinguished whether DYRK1A is in specific complexes containing a distinct subpopulation of the polymerase, either hypophosphorylated and/or hyperphosphorylated in Ser^5 and/or Ser^2. Given that DYRK1A binds directly to the CTD, the presence in RNA polymerase complexes could indicate that the kinase interacts with RNA polymerase II.

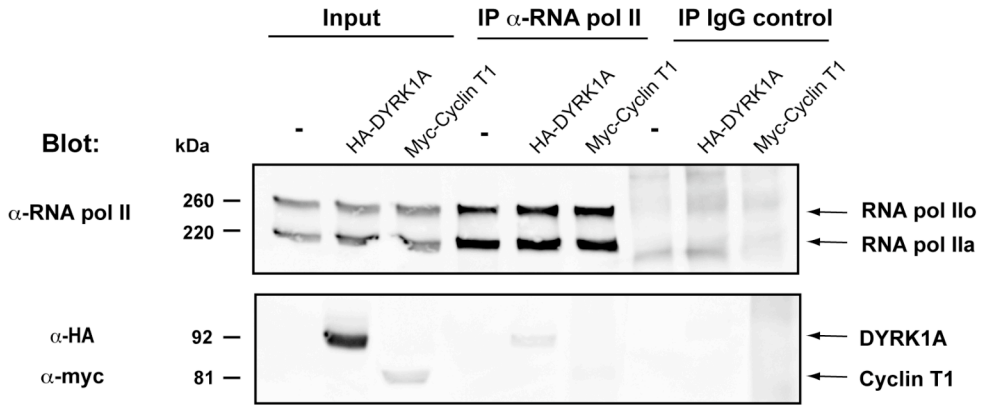


Figure 24: DYRK1A is in complexes with RNA polymerase II. HEK-293T cells were transiently transfected with an empty vector (-), pHA-DYRK1A or pMyc-Cyclin T1. At 48 h after transfection, soluble nuclear extracts were prepared and incubated with protein G-beads coupled to a rabbit IgG (IP IgG control) or to the N20 antibody (IP α-RNA pol II) for 3 h at 4ºC. Beads were extensively washed and finally resuspended with sample buffer. The presence of DYRK1A and cyclin T1 in the immunoprecipitates was assessed by Western blot with anti-HA and anti-myc antibodies, respectively. Both the hypo- (RNA pol IIa) and hyperphosphorylated (RNA pol IIo) forms of the RNA polymerase II are shown. The presence of DYRK1A and cyclin T1 is indicated with an arrow. Input represents 10% of total extracts.

Second, the behaviour of the two proteins upon subcellular fractionation was analyzed. For this reason, HEK-293T cells transfected with HA-DYRK1A were subjected to an extraction protocol in the presence of low and high salt concentrations (see experimental design in Figure 25B). Both soluble and insoluble fractions were analyzed with anti-HA and anti-N20 antibodies to detect DYRK1A and RNA polymerase II (hypo- and hyper-phosphorylated forms), respectively. As previously described, the insoluble fraction in high salt was enriched in hyperhosphorylated RNA polymerase II (Mortillaro et al., 1996), where it forms part of macromolecular complexes related to transcription and RNA processing (Figure 25B, P2 fraction). DYRK1A was distributed all along the fractions analyzed, including the pellet of the high salt solubilization (Figure 25B). The detection of DYRK1A and the RNA polymerase II in the same subcellular fractions strongly suggests that both proteins share the same intracellular localization and then have the opportunity to interact.

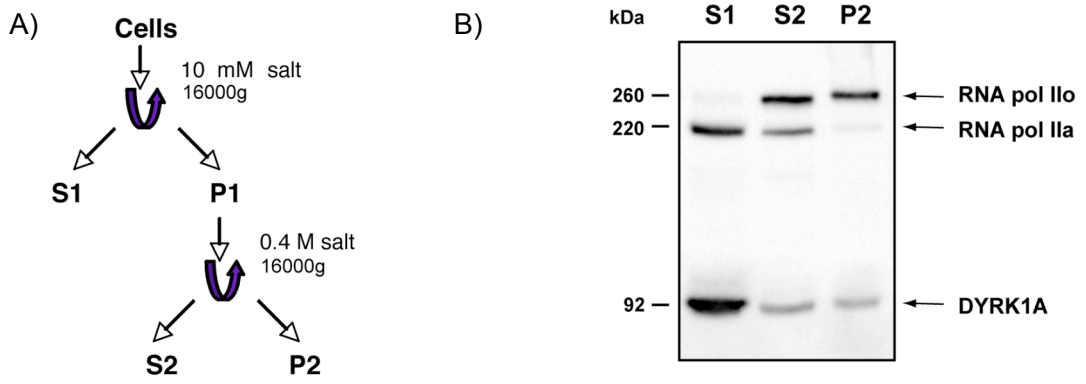


Figure 25: DYRK1A and the RNA polymerase II co-fractionate. A) Schematic representation of the subcellular fractionation assay. B) HEK-293T cells were transiently transfected with HA-DYRK1A. At 48 h post-transfection, cells were processed as described in Materials and Methods (section 4.4). The samples were analyzed by immunoblotting with anti- HA and anti-N20 antibodies to detect DYRK1A and RNA polymerase II, respectively. The N20 antibody detects both hypo- (RNA pol IIa) and hyper-phosphorylated (RNA pol IIo) forms of the RNA polymerase II.

Finally, to test the subcellular co-localization, an immunofluorescence staining was performed in HeLa cells transfected with a GFP-DYRK1A-expressing construct. The detection of endogenous RNA polymerase II was carried out with 8WG16 antibody, which recognizes the hypophosphorylated form of the protein. The result showed that DYRK1A and RNA polymerase II partially co-localized in some foci within the nucleus (Figure 26).

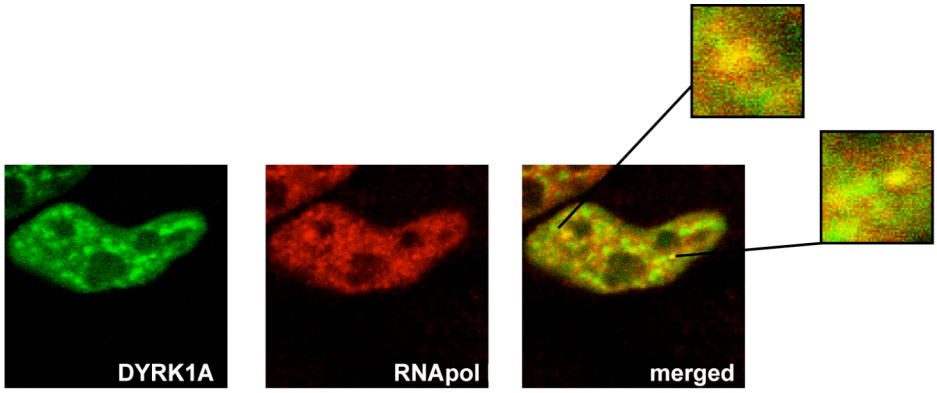


Figure 26: DYRK1A and the RNA polymerase II are localized in the same subcellular compartment. HeLa cells were grown on coverslips and transiently transfected with GFP-DYRK1A and analyzed by direct fluorescence (left panel) and indirect immunofluorescence with 8WG16 and an Alexa 647-labelled anti-mouse secondary antibody to detect hypophosphorylated RNA polymerase II (central panel). Images including the merged (right panel) were taken by confocal microscopy. The inlets show amplified images of areas in which co-localization is clearer.

5.5 DYRK1A phosphorylates the CTD of the RNA polymerase II

As explained above, the serine residues of the consensus heptapeptide, Y**S$_2$**PT**S$_5$**PS, which forms the CTD of the RNA polymerase II can be phosphorylated. To date, some CTD-kinases have been reported to phosphorylate specific Ser residues at a specific step during

transcription (reviewed in Palancade and Bensaude, 2003; Prelich, 2002). At least three of the CTD kinases are cyclin-dependent kinases: CDK7, CDK8 and CDK9, which generally associate with cyclins H, C and T, respectively. In addition, during stress response the CTD can also be phosphorylated by ERK kinases (Bonnet et al., 1999). All these kinases differ in their specificity towards the target serine residue in the CTD. Kinase specificity may not only be achieved by CTD recognition at the kinase active site, but also by CTD binding to kinase-associated factors, as cyclin T1 in P-TEFb (Taube et al., 2002). It is thought that these factors can act as bridges between signalling cascades and the transcription machinery.

The CTD heptapeptide constitutes a putative phosphorylation substrate for DYRK1A since it is a proline-directed kinase (Himpel et al., 2000). To study if DYRK1A could phosphorylate the CTD, *in vitro* kinase assays were performed with both proteins expressed and purified from bacteria. As shown in Figure 27A, DYRK1A efficiently phosphorylated the RNA polymerase II CTD, causing a striking change on the electrophoretic mobility of the protein (Figure 27B). This is likely related to the repetitive nature of the CTD region and would mean that DYRK1A incorporation of phosphates could involve several Ser residues in the CTD. In fact, it has been reported that the CTD can be phosphorylated in more than 50 sites *in vivo*, with one phosphate per repeat as an average (Palancade and Bensaude, 2003).

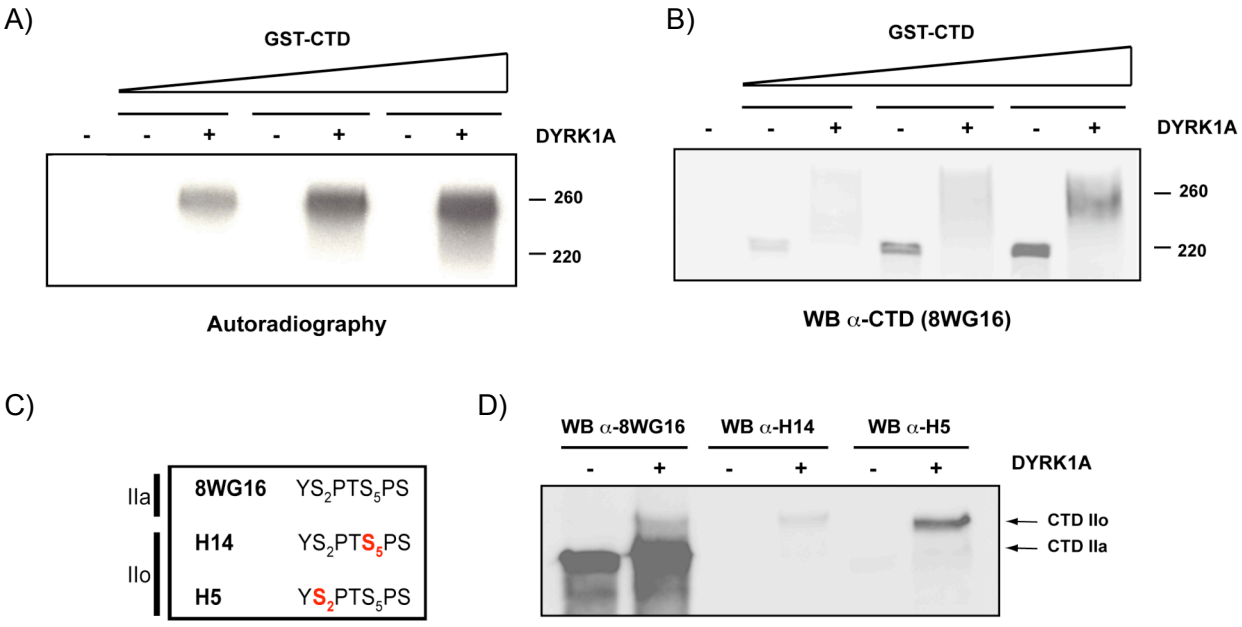


Figure 27: DYRK1A phosphorylates the CTD of the RNA polymerase II. A) *In vitro* kinase assay where increasing amounts of GST-CTD (100 ng, 250 ng and 500 ng) were incubated with [^{32}P]-γ-ATP in the presence of 50 ng of unfused GST or GST-DYRK1A. The reaction was performed for 30 min at 30ºC, and samples were analyzed by SDS-PAGE and autoradiography. B) An *in vitro* kinase assay was performed in the presence of cold ATP. Samples were analyzed by Western blot with the 8WG16 antibody, which recognizes the hypophosphorylated CTD and can slightly cross-react with the hyperphosphorylated one. C) Schematic representation of the recognition abilities of the antibodies used. The red colour has been used as an indication of phosphorylation. IIo: RNA polymerase II hypophosphorylated form; IIa: RNA polymerase II hyperphosphorylated form. D) DYRK1A phosphorylates the CTD in Ser^{2}. An *in vitro* kinase assay was performed in the presence of cold ATP and 500 ng of

GST-CTD with 50 ng of unfused GST (-) or with 50 ng of GST-DYRK1A (+) for 30 min at 30ºC. Equivalent amounts of each reaction were loaded in triplicate and analyzed by Western blot with 8WG16, H14 and H5 antibodies, as indicated.

To alight the role of DYRK1A on the RNA polymerase II function, we investigated which residue was phosphorylated. For that, the CTD products of an *in vitro* kinase assay were analyzed by Western blot with 8WG16 antibody and specific antibodies against phosphorylated Ser^5 (H14) or Ser^2 (H5) (Figure 27C). Although it is difficult to quantify the phosphorylation levels by comparing the results of the detection with two different antibodies, it is clear that DYRK1A phosphorylated both type of residues *in vitro* (Figure 27D). Altogether the results show that DYRK1A is a CTD kinase.

5.6 Other DYRK family members interact with the CTD of RNA polymerase II

The mapping of the CTD interaction site within the catalytic domain of DYRK1A suggested that this function could be conserved in other members of the DYRK family of protein kinases. Therefore, pull-down assays were performed with the other human DYRKs: DYRK1B, DYRK2, DYRK3 and DYRK4 (Figure 28A). DYRK1B, which is the closest kinase to DYRK1A in the family, was able to bind the CTD (Figure 28B). As DYRK1B lacks the His tract, the result indirectly confirms that this region is not essential for the interaction. DYRK4 also interacted with the CTD, whereas DYRK2 and DYRK3 did not. Therefore, the ability to bind RNA polymerase II, and maybe of phosphorylating the CTD, is conserved by a subgroup of DYRK kinases.

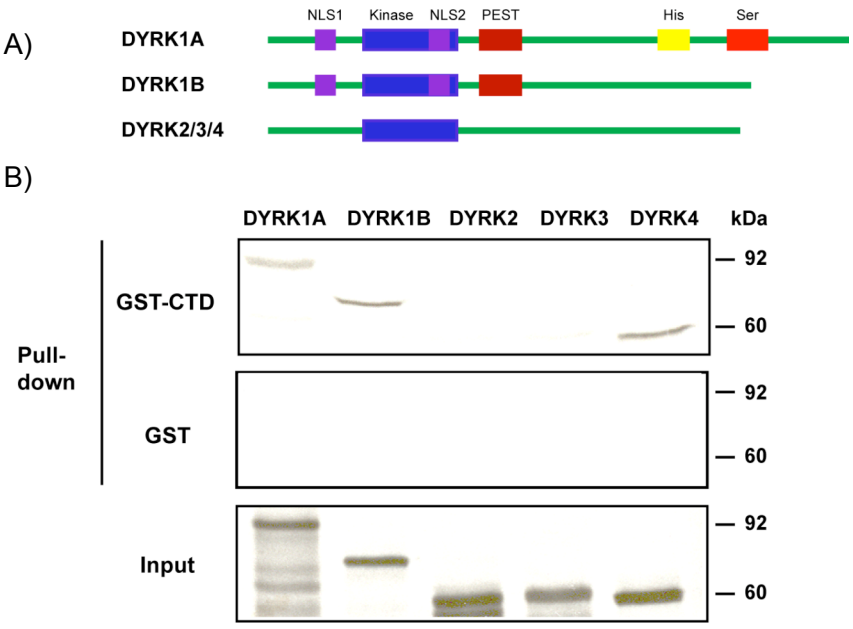


Figure 28: The interaction with the RNA polymerase II CTD is conserved in some members of the DYRK family. A) Schematic representation of human DYRK protein kinases. The main features are shown. B) Pull-down assays were performed with *in vitro* translated DYRK1A, DYRK1B, DYRK2, DYRK3 and DYRK4 and incubated with 5 μg of unfused GST (GST) or GST-CTD (CTD) for 3 h at 4ºC. Beads were washed and bound protein analyzed by SDS-PAGE and autoradiography. No binding was detected to the GST beads. Input represents 10% of the samples.

6. Proteins with His repeats that localize in nuclear speckles may form macromolecular protein complexes

6.1 DYRK1A interacts with polyHis proteins localized in speckles

Nuclear speckles, as other nuclear bodies, are proposed to be highly dynamic self-organizing entities (Misteli, 2001), and therefore are thought to be the result of protein-protein interactions among different proteins forming the speckle, as opposed to attachment to a preformed scaffolding surface (Sacco-Bubulya and Spector, 2002). Thus, it was plausible to think that some of the proteins that were identified in the speckles localization screen may interact. By establishing, at least, a single connection with another polyHis protein, through the His-repeat, accumulation in the speckle may be achieved.

To test this hypothesis, co-immunoprecipitation assays of SFC-positive proteins were performed. In a first assay, the kinase DYRK1A was chosen as the bait. Figure 29 shows a set of results in which the interaction of DYRK1A with several polyHis-containing proteins could be observed.

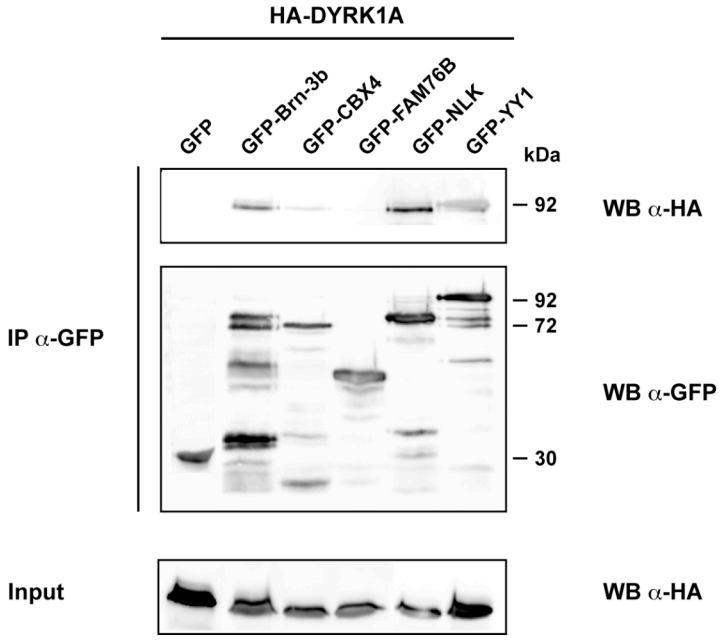


Figure 29: The protein kinase DYRK1A interacts with some of the polyHis containing proteins that accumulate in nuclear speckles. HEK-293T cells were transfected with plasmids encoding fusion proteins HA-DYRK1A, unfused GFP (-), GFP-Brn3b (649 amino acids), GFP-CBX4 (637 amino acids), GFP-FAM76B (579 amino acids), GFP-NLK (753 amino acids) and GFP-YY1 (652 amino acids). At 48 h after transfection soluble cell extracts were immunoprecipitated with anti-GFP antibody. Samples were analyzed by Western blot with anti-GFP and anti-HA antibodies. Input represents 10% of total extracts.

The results of the interaction analysis between DYRK1A and the whole set of proteins with His-repeats are summarized in Table 4, and indicate that 11 out of the 16 proteins analyzed behaved as positive according to the interaction assay (above 10% interaction).

Table 4: Summary of the results obtained in the interaction analysis of the whole set of polyHis-containing proteins with DYRK1A

Name	SFC[a]	Interaction[b]
Brn-3b	+	62%
CBX4	+	11%
Cyclin T1	+	2%
FAM76A	+	n.d.[c]
FOXG1B	+	15%
GSH2	+	32%
HAND1	+	8%
HOXA1	+	35%
HOXA9	+	n.d.[c]
LMO6	+	29%
MafA	+	89%
MEOX2	+	18%
NLK	+	87%
OTX1	+	100%
YY1	+	42%
MafB	-	n.d.[c]

a: accumulation in speckles; b: interaction with DYRK1A measured by co-immunoprecipitation was quantified by densitometry with the Multigauge Software and expressed as a percentage of the ratio between the bound fraction and the input; c: not detected

6.2 Proteins with His-tracts may form an interacting network in the speckle

To further extend the results on the interaction of polyHis-containing proteins, the transcription factor Brn-3b was used as bait in co-immunoprecipitation assays.

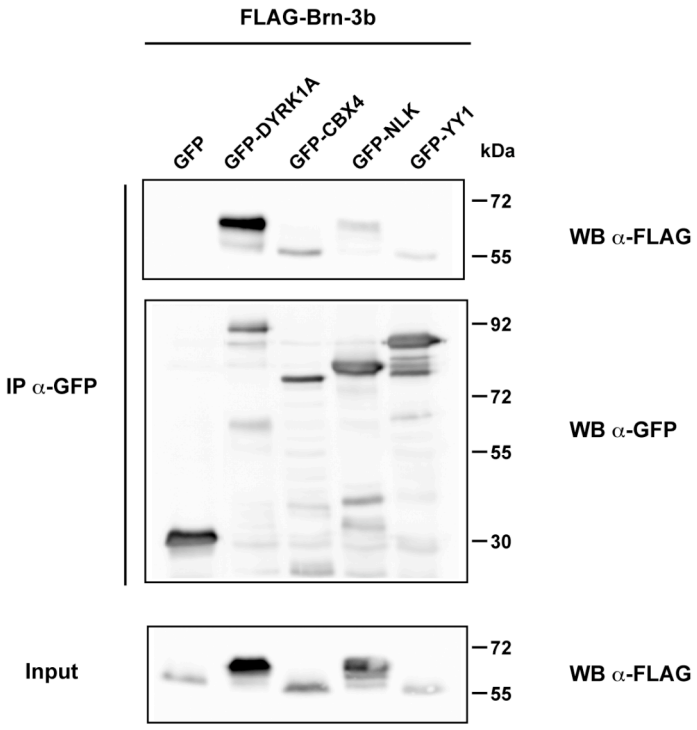


Figure 30: The transcription factor Brn-3b binds proteins with His-repeats. HEK-293T cells were transfected with expression vectors encoding Flag-Brn3b, GFP-DYRK1A, GFP-CBX4, GFP-NLK and GFP-YY1 fusion proteins. At 48 h after transfection, soluble cell extracts were immunoprecipitated with anti-GFP antibody and samples analyzed by Western blot with anti-Flag and anti-GFP antibodies. Input represents 10% of total extracts. Note that when Brn-3b is co-expressed with the protein kinases DYRK1A or NLK an electrophoretic mobility shift is detected.

As shown in Figure 30, Brn3b bound all the polyHis proteins tested, the two kinases DYRK1A and NLK, the chromatin remodeller CBX4 and the transcription factor YY1. This result confirms some polyHis proteins may establish a net of stable interactions among each other.

6.3 The interaction between DYRK1A and proteins with His repeats does not depend on the His tract

To test the involvement of the polyHis-segment in the interaction between polyHis proteins, co-immunopecipitation experiments were performed with the mutant protein DYRK1AΔHis. The interaction of this protein with most of the proteins tested was unaffected when compared with DYRK1A wild type (Figure 31). Only in the case of the transcription factor Brn-3b, the interaction was compromised by the deletion of the His-repeat (see next Section). This result would indicate that, at least in the case of the kinase DYRK1A, the His-repeat is not used as a general interaction domain with this subset of proteins.

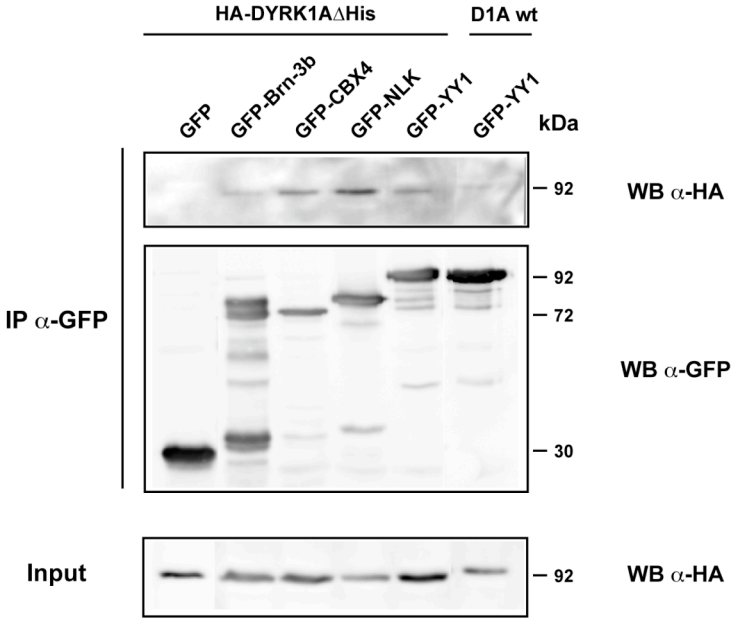


Figure 31: The interaction between DYRK1A and most of the polyHis-containing proteins is not mediated by the His-repeat. HEK-293T cells were transfected with plasmids encoding fusion proteins HA-DYRK1AΔHis, HA-DYRK1A wt, GFP-Brn-3b, GFP-CBX4, GFP-NLK and GFP-YY1. At 48 h after, soluble cell extracts were immunoprecipitated with anti-GFP antibody. Samples were analyzed by Western blot with anti-GFP and anti-HA antibodies. Input represents 10% of total extracts.

This result clearly demonstrates that, in most cases, the interaction with other polyHis-containing proteins is not mediated by the His repeat, and therefore could not be involved in nuclear speckle targeting. However, this binding may contribute to the accumulation of these proteins in the speckle.

7. Characterization of the interaction between Brn-3b and DYRK1A

7.1 The hypothesis

When co-expressing the kinase DYRK1A and the transcription factor Brn-3b, a change in the electrophoretic mobility of the latter was noticed (see Figure 29). The electrophoretic behaviour of the whole set of polyHis proteins was analyzed in the absence or presence of DYRK1A. Figure 32 shows the results obtained with a small subset of proteins as an example. Only Brn-3b showed a variation in its electrophoretic behaviour, with the appearance of lower mobility bands that could be the result of a phosphorylation event, suggesting that it may be a substrate of DYRK1A. In fact, the analysis of its primary sequence revealed several DYRK1A phosphorylation consensus sites (Himpel et al., 2000).

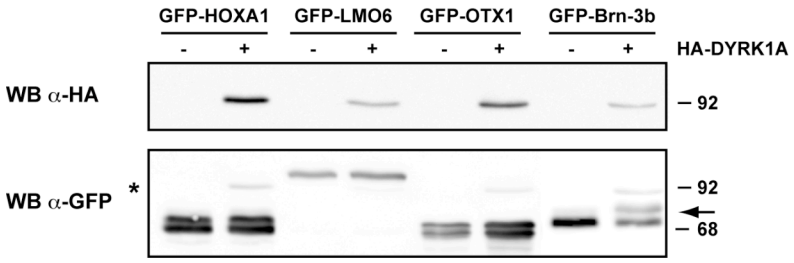


Figure 32: DYRK1A induces a change in the electrophoretic mobility of Brn-3b. HEK-293T cells were transfected with plasmids expressing HA-DYRK1A and GFP-HOXA1 (575 amino acids), GFP-LMO6 (855 amino acids), GFP-OTX1 (594 amino acids) and GFP-Brn-3b (649 amino acids). At 48 h after transfection, soluble extracts were prepared and analyzed by Western blot with anti-HA. The membrane was reblotted with anti-GFP; * indicates HA-DYRK1A signal. Note the appearance of a doublet of GFP-Brn3b in the presence of HA-DYRK1A (marked with an arrow).

Brn-3b (named as POU4F2 in Section 2) belongs to the small subfamily of POU domain transcription factors, characterized by a bipartite DNA binding domain, the POU domain. Brn-3b is essential for the differentiation and survival of retinal ganglion cells (RGC) (Gan et al., 1996), the output neurons in the retina that form the optic nerve and convey light signals detected by photoreceptors to the higher visual system (reviewed in Marquardt and Gruss, 2002).

There are several evidences supporting the existence of a functional link between Brn-3b and DYRK1A. First, the immunohistochemical analysis of retinas from DYRK1A heterozygous and transgenic animals showed specific defects in the generation of the correct number of ganglion cells where Brn-3b is expressed (Laguna, 2008). Second, the retinal phenotype of the heterozygous mice completely recapitulates the phenotype observed in Brn-3b null mice (Laguna, 2008). Third, a downregulation of several DYRK1A substrates, such Gli1 or tau has been found in a study reporting expression changes in Brn-3b null

retinas (Mu et al., 2004). Finally, a downregulation of DYRK1A has been reported in Math5-null retinal progenitor cells (RPC) (Mu et al., 2005). Math5 is an essential factor in the developing retina, which acts upstream of Brn-3b to establish RPC competence for ganglion cell fate.

Not much is known about the mechanisms that regulate the transcriptional activity of Brn-3b. Up to date, the only protein that has been described as a Brn-3b regulator is HIPK2, which belongs to the DYRK family of protein kinases. HIPK2 has been reported to interact with Brn-3b (Wiggins et al., 2004), and to bind to and phosphorylate Pax6, a master gene in eye development (Kim et al., 2006). This might point out to some degree of conservation in the function of DYRK proteins during retinogenesis.

Altogether, the existence of DYRK1A being a regulator of Brn-3b was hypothesized and thus their functional relationship was explored in this thesis work.

7.2 Brn-3b accumulates in nuclear speckles both *in vivo* and *in vitro*

The localization of Brn-3b in nuclear speckles, when overexpressed in mammalian cell lines, has been already shown (see Figure 7). To analyze the subcellular localization of the protein in a more physiological situation, the mouse retina was the tissue of election due to the high level of Brn-3b expression in ganglion cells. Figure 33 shows a perfect co-localization of Brn-3b and SC35 in all the cells of the ganglion cell layer, indicating that Brn-3b accumulates in the SFC in physiological conditions. Moreover, the result confirms once more that the nuclear speckle localization observed in HeLa cells is not an artefact of protein overexpression.

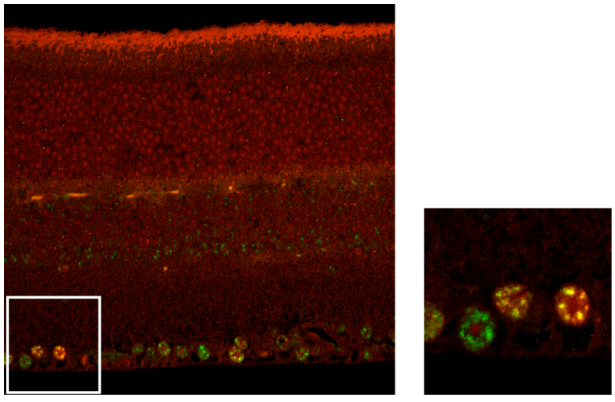


Figure 33: Brn-3b accumulates in nuclear speckles *in vivo.* Double immunofluorescence performed by A. Laguna (Dr. M. Arbonés' group at the CRG) onto 2 μm paraffin sections of the retina of 2-month old mice. SC35 is in green and Brn-3b in red, respectively. Immunofluorescence was done sequentially to eliminate antibody cross-reaction and images are confocal. Right image is a magnification of the area squared in white, which mostly corresponds to the ganglion cell layer of the retina.

7.3 Brn-3b and DYRK1A co-localize in the nucleus both *in vivo* and *in vitro*

Since both Brn-3b and DYRK1A accumulate in nuclear speckles, it was reasonable to postulate that the proteins co-localize, at least, in this subnuclear compartment. To address this question, the localization of the two proteins was analyzed by immunofluorescence staining both in a cell line and in sections of the mouse retina. As expected, exogenously expressed Brn-3b and DYRK1A accumulated and co-localized in some nuclear dots, very likely nuclear speckles (Figure 34A). Importantly, Brn-3b and DYRK1A are co-expressed in retinal ganglion cells, where they co-localize in nuclear dots (Figure 34B).

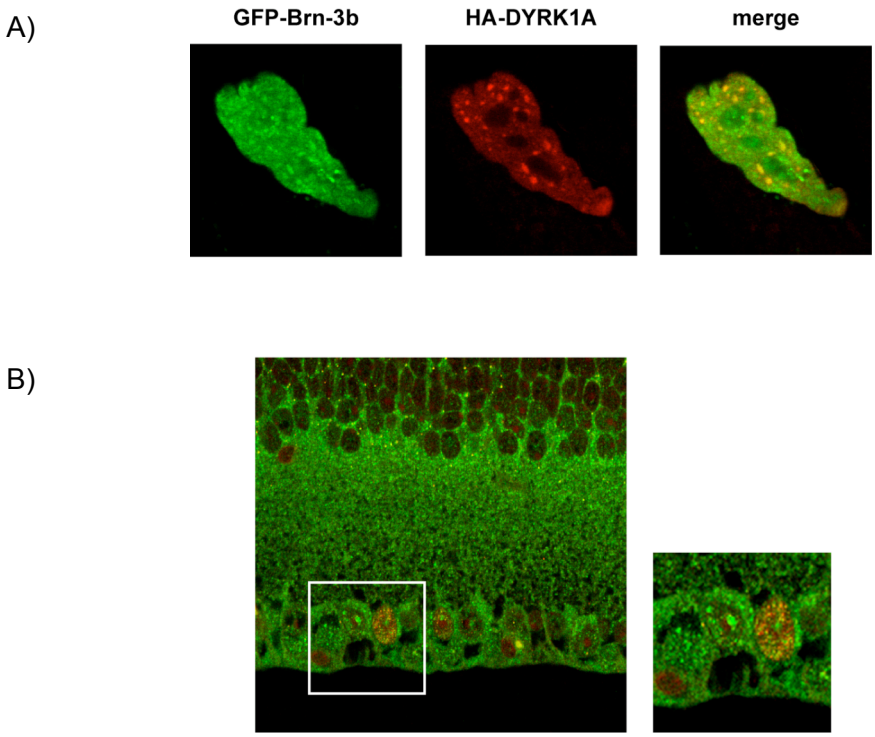


Figure 34: DYRK1A and Brn-3b co-localize in subnuclear structures. A) HeLa cells were transfected with GFP-Brn3b and HA-DYRK1A expression plasmids. At 48 h after transfection, cells were immunostained with anti-HA antibody and Alexa 647-conjugated goat anti-mouse. Note that the morphology of Brn-3b-positive speckles is slightly different to the observed in Section 3. This might be explained by the extremely low amount of GFP-Brn3b plasmid transfected in this experiment, due to the severe overexpression of the protein observed when DYRK1A is co-expressed. B) Double immunofluorescence performed by A. Laguna onto 2 μm paraffin sections of the retina of 2-month old mice. Primary antibodies were anti-DYRK1A (in green) and anti-Brn-3b (in red). The following secondary antibodies were used; Alexa 488-conjugated goat anti-rabbit (Molecular Probes, 1:400) and Alexa 555-conjugated donkey anti-goat (Molecular Probes, 1:400). Immunofluorescence was done sequentially to eliminate antibody cross-reaction. Right image is a magnification of the area squared in white, which mostly corresponds to the ganglion cell layer of the retina.

7.4 DYRK1A interacts with Brn-3b

As discussed in subsection 6 of this work, Brn-3b and DYRK1A were found in the same protein complexes (Figure 29). Pull-down assays with a GST-Brn3b fusion and *in vitro* translated DYRK1A indicated that the interaction between both proteins is direct (data not shown). Taking into account the enzymatic nature of DYRK1A, it would be possible that this

interaction depended on its kinase activity. Compared to the wild type form of DYRK1A, the kinase dead mutant DYRK1A K179R showed a decrease in the level of binding to Brn-3b (Figure 35B), indicating that the kinase activity of DYRK1A contributes to the interaction.

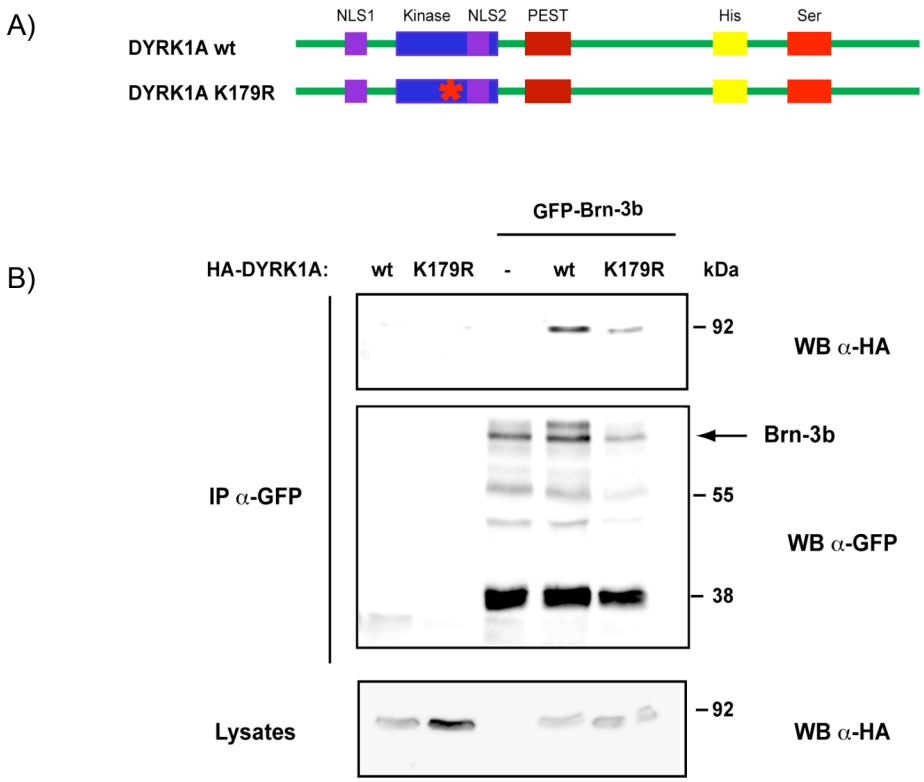


Figure 35: DYRK1A interacts with Brn-3b. A) Schematic picture of DYRK1A wt and the kinase mutant DYRK1A K179R. B) HEK-293T cells were transfected with plasmids expressing GFP-Brn3b, HA-DYRKA wt and the kinase-death mutant version (K179R). At 48 h after transfection, soluble cell extracts were immunoprecipitated with anti-GFP antibody. Samples were analyzed by Western blot with anti-GFP and anti-HA antibodies. Lysates represent 10% of the total extract.

In the previous section, the involvement of DYRK1A His repeat in the interaction with Brn-3b was suggested (see Figure 31), and this was confirmed by the results in Figure 36A showing that the binding of Brn-3b is abolished upon the deletion of the His tract in DYRK1A. This suggests that the His-repeat in DYRK1A acts as a surface of interaction between the two proteins.

To address whether the His run of Brn-3b was involved in the interaction with DYRK1A, a co-immunoprecipitation assay using either Brn-3b wild type or a mutant without the His repeat (Brn-3bΔHis) was performed. As shown in Figure 36B, DYRK1A only interacted with Brn-3b wild type, but not with the His repeat-deleted mutant. This result indicates that the His tract of Brn-3b is the region responsible for the binding to DYRK1A.

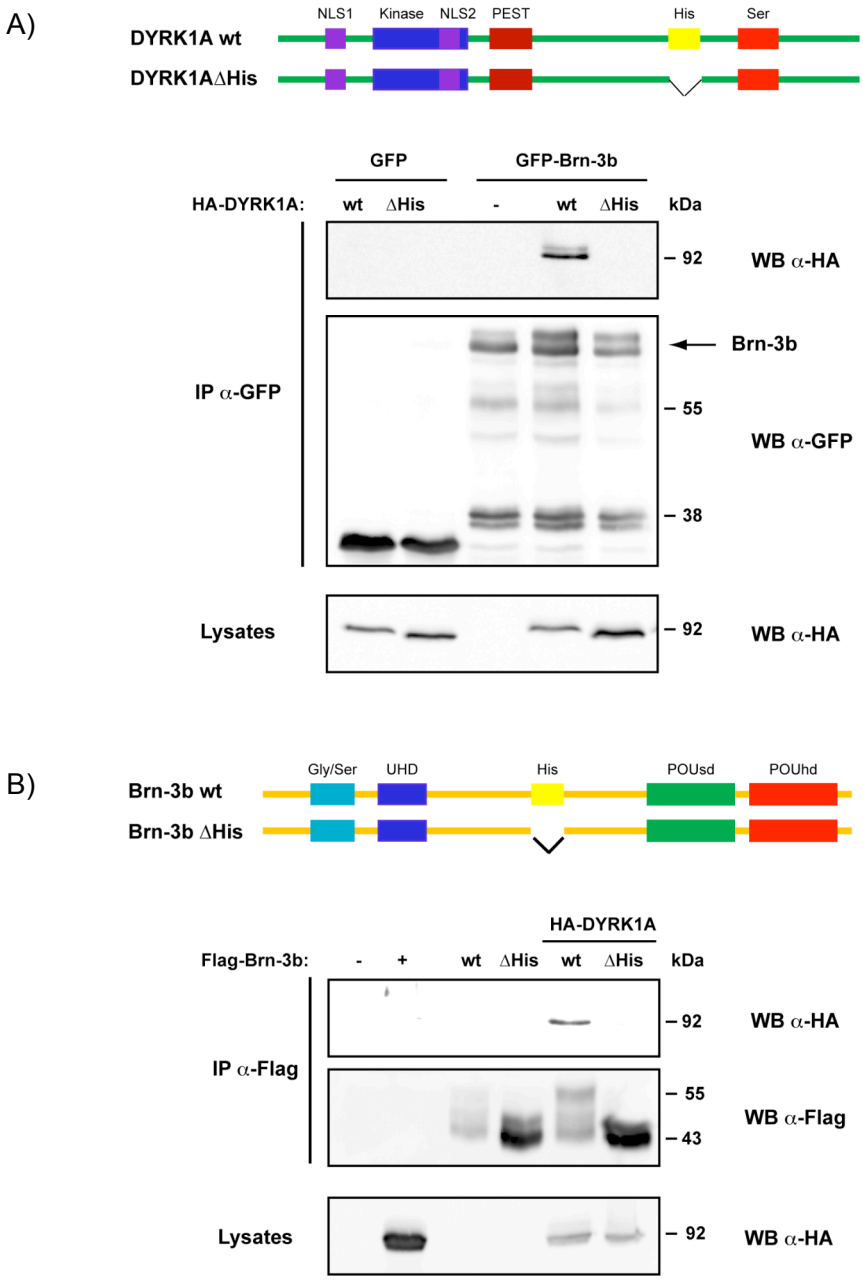


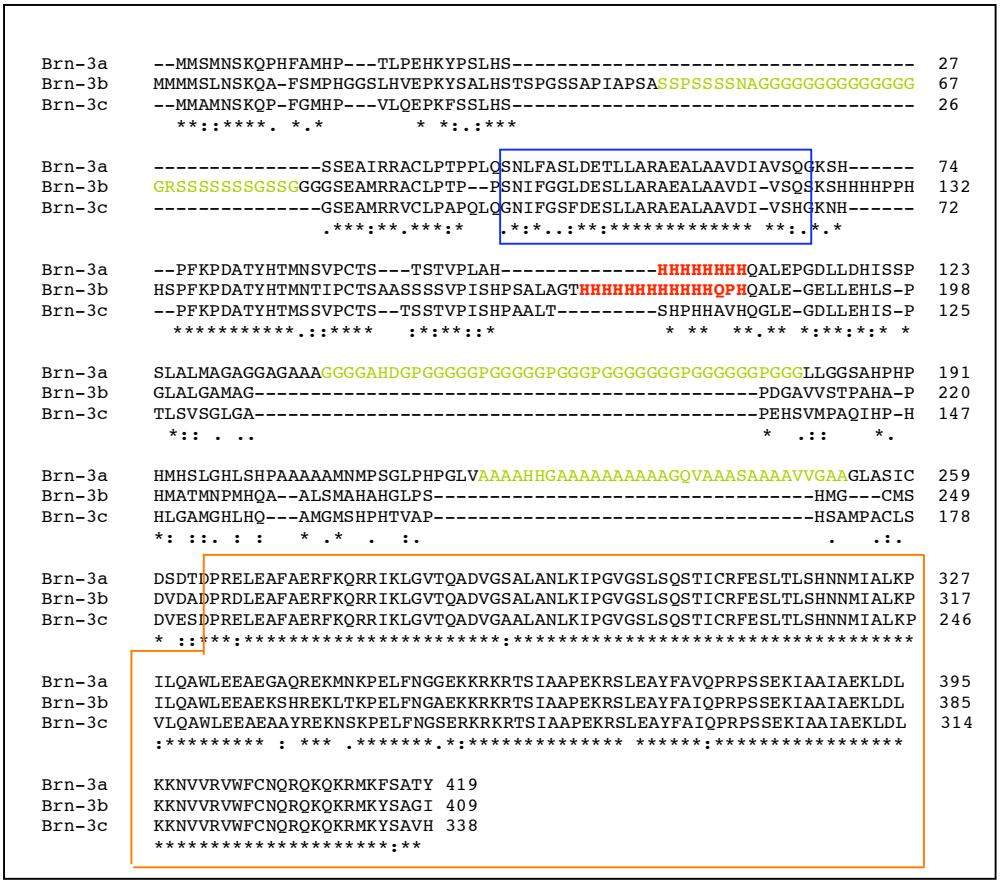
Figure 36: The interaction between Brn-3b and DYRK1A is mediated by the His repeat present in each protein. A) HEK-293T cells were transfected with expression plasmids for GFP-Brn3b, HA-DYRK1A wt and HA-DYRK1AΔHis. At 48 h post-transfection, soluble cell extracts were immunoprecipitated with anti-GFP antibody and the samples were analyzed by Western blot with anti-HA and anti-GFP antibodies. Input represents 10% of total extracts. Brn-3b is indicated with an arrowhead. B) HEK-293T cells were transfected with expression plasmids for HA-DYRK1A, Flag-Brn3b wt and Flag-Brn3bΔHis. At 48 h post-transfection, soluble cell extracts were immunoprecipitated with anti-Flag antibody and the samples were analyzed by Western blot with anti-HA and anti-Flag antibodies. Lysates represent 10% of total extracts.

7.5 DYRK1A does not interact with other members of the Brn-3 family

In humans, the Class IV of POU-domain transcription factors includes three members: Brn-3a (POU4F1), Brn-3b (POU4F2) and Brn-3c (POU4F3). They display 95% sequence identity in the DNA-binding POU domain and a homology of 70% in the rest of the protein (Figure 37A)

(Xiang et al., 1995). It is known that Brn-3a and Bnr-3c also participate in RGC differentiation and, in fact, a functional equivalence of the three Brn-3 genes in promoting RGC differentiation has been reported (Liu et al., 2000; Pan et al., 2005). However, the retinas of *Brn-3b/Brn-3c* and *Brn-3b/Brn-3a* double knock-out mice show more severe RGC defects than those of *Brn-3b* null mice (Wang et al., 2002), suggesting that redundancy is not complete. In addition, the expression of Brn-3a and Brn-3c is significantly down-regulated in retinas lacking Brn-3b (Gan et al., 1996; Wang et al., 2002).

A)



B)

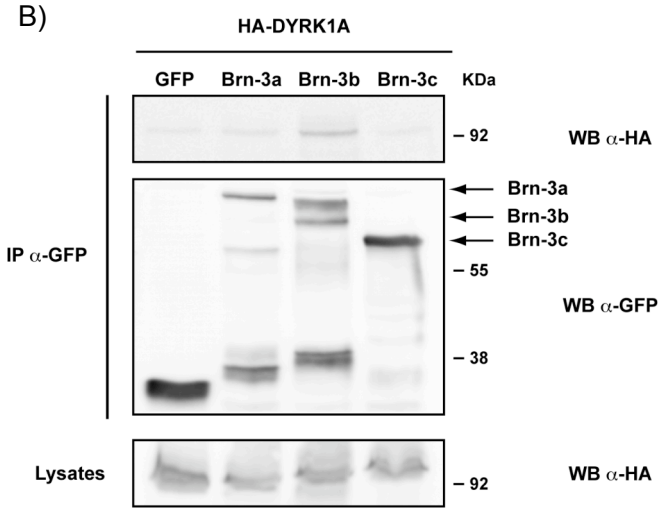


Figure 37: DYRK1A interacts differentially with the members of the Brn-3 family of proteins. A) Sequence alignment of the three Brn-3 proteins found in humans: Brn-3a (NP_006228), Brn-3b (NP_004566) and Brn-3c (NP_002691). His repeats appear highlighted in red. In green are depicted other amino acid rich regions. The blue box includes the residues of the upstream homology box (UHD box). The orange box corresponds to the POU domain. Note the high level of homoloy in the homeobox domain. B) HEK-293T cells were transfected with expression plasmids for GFP-Brn-3a, GFP-Brn-3b, GFP-Brn-3c and HA-DYRK1A. At 48 h after transfection, soluble cell extracts were immunoprecipitated with anti-GFP antibody and samples were analyzed by Western blot with anti-HA and anti-GFP antibodies. Lysates represent 10% of total extracts.

To test if DYRK1A could interact with the other Brn-3 proteins, co-immunoprecipitation experiments were performed. As shown in Figure 37B, Brn-3a interacted very weakly with DYRK1A, whereas Brn-3c binding was close to the background level. Moreover, no changes in the electrophoretic mobility of Brn-3a or Brn-3c were observed. Therefore, DYRK1A shows specificity towards Brn-3b.

7.6 DYRK1A phosphorylates Brn-3b

The appearance of the Brn-3b low electrophoretic mobility bands when co-expressed with DYRK1A was very likely a result of post-translational modifications, including phosphorylation. The absence of the upper band when Brn-3b is co-expressed with a kinase-dead mutant reinforces the idea of a phosphorylation event (Figure 38A). To check this, a phosphatase assay was performed. As Figure 36B shows, the Brn-3b low mobility band disappears when the cell extract is treated with alkaline phosphatase, confirming that phosphorylation is responsible of the changes in Brn-3b electrophoretic mobility in the presence of DYRK1A.

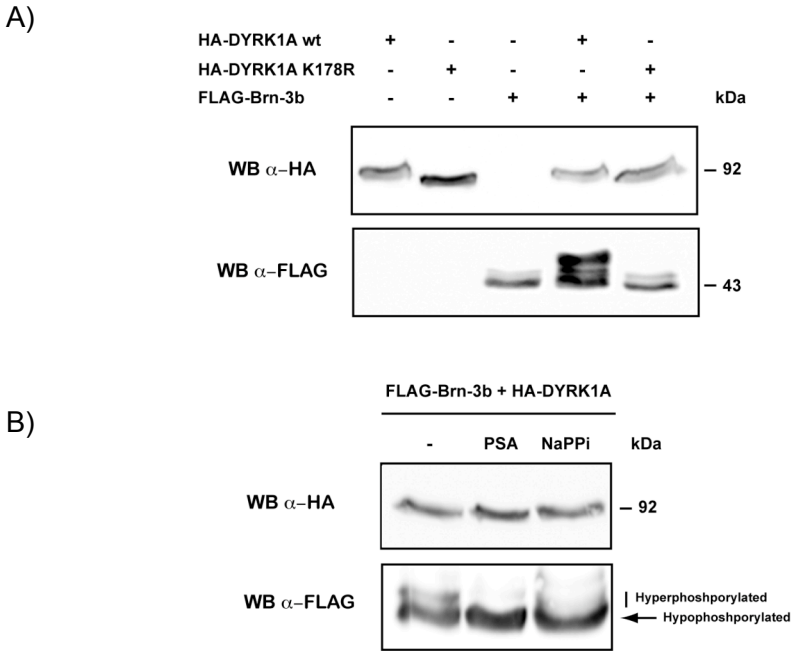


Figure 38: Brn-3b electrophoretic mobility changes are the result of protein phosphorylation. A) Total cell extracts from HEK-293T cells expressing the indicated proteins were analyzed by Western blot with anti-Flag and anti-HA antibodies. B) HEK-293T cells were transfected with expression plasmids pFlag-Brn3b and pHA-DYRK1A. Soluble extracts were incubated for 30 min at 30ºC with alkaline phosphatase (PSA, 400 u/ml), or with alkaline phosphatase plus sodium pyrophosphate (NaPPi, 200 mM) as a phosphatase inhibitor. Samples were analyzed by Western blot with anti-Flag and anti-HA antibodies. Control (-) means untreated extracts.

To further analyse if DYRK1A is the direct responsible of Brn-3b phosphorylation, an *in vitro* kinase assay with DYRK1A and Brn-3b expressed and purified from bacteria was

performed, showing that Brn-3b is phosphorylated by DYRK1A (Figure 39A). The same result was obtained when the *in vitro* kinase assay was performed with immunocomplexes from cell extracts containing DYRK1A and Brn-3b (Figure 39B). Both results confirm that the transcription factor Brn-3b is a novel substrate of the protein kinase DYRK1A.

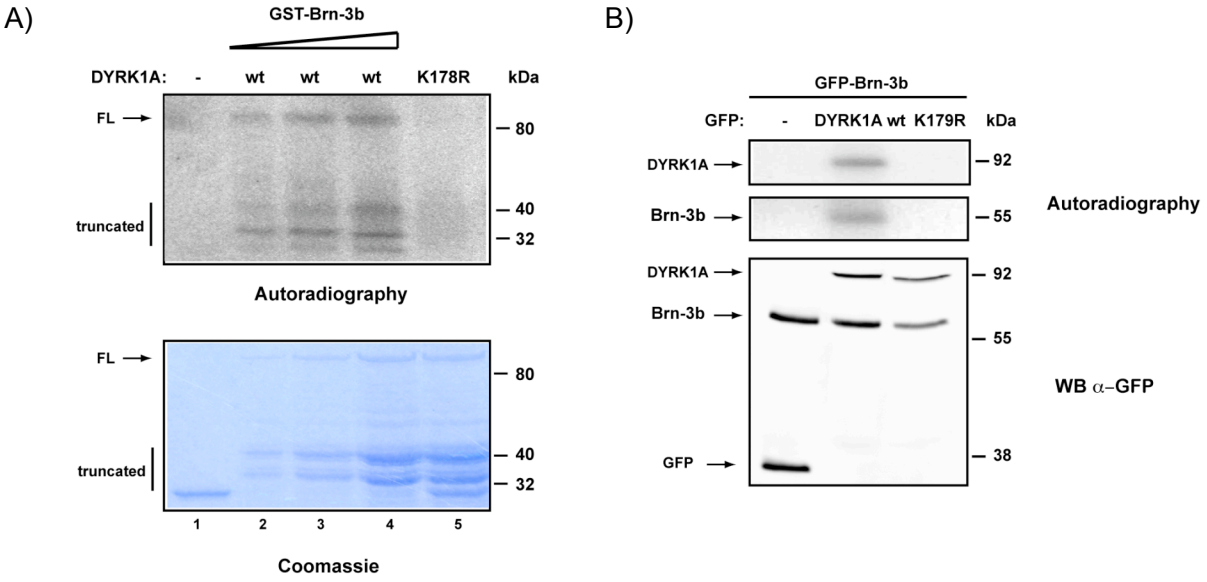
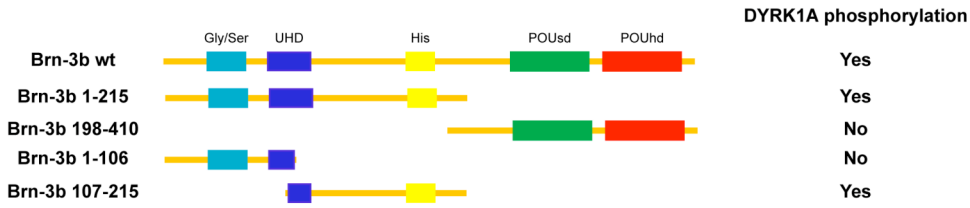


Figure 39: DYRK1A phosphorylates Brn-3b. A) *In vitro* kinase assay with increasing amounts of GST-Brn3b (100 ng, 250 ng and 500 ng) of incubated with [^{32}P]-γ-ATP in the presence of 50 ng of unfused GST (-) or 50 ng of GST-DYRK1A wt (wt) or the kinase-death mutant (K179R). The reaction was performed for 30 min at 30ºC, and samples were analyzed by SDS-PAGE and Coomassie staining and autoradiography. B) HEK-293T cells were transfected with plasmids expressing GFP, GFP-Brn3b, GFP-DYRK1A wt and K179R. At 48 h after transfection, soluble cell extracts were immunoprecipitated with anti-GFP antibody and beads were incubated with kinase buffer in the presence of [^{32}P]-γ-ATP for 30 min at 30ºC. Samples were loaded and analyzed by Western blot with anti-GFP antibody and by autoradiography.

Several Ser and Thr residues in Brn-3b follow the consensus site of DYRK1A phosphorylation (Himpel et al., 2000). As a first attempt to identify the specific Ser/Thr residues phosphorylated by DYRK1A, mass spectrometry analysis was used (UPF Proteomic Facility) with no success. Next, to initially narrow the phoshorylated region, GST fusions containing the two halves of Brn-3b (N-terminal: residues 1-215; C-terminal: residues 215-410) were generated and tested in an *in vitro* kinase assay with GST-DYRK1A. The result clearly indicated that only the N-terminal part of Brn-3b was phosphorylated by DYRK1A, which agrees with the fact that most of the SP and TP candidate phosphosites are distributed within the first half of the protein. Then, the fragment 1-215 was divided in two smaller regions (1-107 and 107-215) and tested again by an *in vitro* kinase assay. Only the region 107-215 was phosphorylated. The results obtained with the whole set of deletion mutants are summarized in Figure 40A.

A)



B)

```
Homo sapiens              SLLARAEALAAVDIVSQSKSHHHHPPHHSPFKPDATYHTMNTIPCTSAASSSSVP  162
Pan troglodytes           SLLARAEALAAVDIVSQSKSHHHHPPHHSPFKPDATYHTMNTIPCTSAASSSSVP  159
Macaca mulatta            SLLARAEALAAVDIVSQSKSHHHHPPHHSPFKPDATYHTMNTIPCTSAASSSSVP  161
Mus musculus              SLLARAEALAAVDIVSQSKSHHHHPPHHSPFKPDATYHTMNTIPCTSAASSSSVP  164
Rattus norvegicus         SLLARAEALAAVDIVSQSKSHHHHPPHHSPFKPDATYHTMNTIPCTSAASSSSVP  163
Bos taurus                SLLARAEALAAVDIVSQSKSHHHHPPHHSPFKPDATYHTMNTIPCTSAASSSSVP  161
Gallus gallus             ----------------------------FKPDATYHTMNTIPCTSAASSSSVP   27
Danio rerio               SLLARAEALAAVDIVSQTKS-HHHPPHHSPFKPDATYHTMNTLPCTSSSSSSSVP   40
                                                       ************:****::******

Homo sapiens              ISHPSALAGTHHHHHHHHHHHHQPHQALEGELLEHLSPGLALG-AMAGPDGAVVS  215
Pan troglodytes           ISHPSALAGTHHHHHHHHHHHHQPHQALEGELLEHLSPGLALG-AMAGPDGAVVS  212
Macaca mulatta            ISHPSALAGTHHHHHHHHHHHHQPHQALEGELLEHLSPGLALG-AMAGPDGAVVS  214
Mus musculus              ISHPSALAGTHHHHHHHHHHHHQPHQALEGELLEHLSPGLALG-AMAGPDGTVVS  217
Rattus norvegicus         ISHPSALAGTHHHHHHHHHHHHQPHQALEGELLEHLSPGLALG-AMAGPDGTVVS  216
Bos taurus                ISHPSALAGTHHHHHHHHHHHHQPHQALEGELLEHLSPGLALG-AMAGPDGAVVS  214
Gallus gallus             ISHPSALSGTHHHHHHHHHHHHQPHQALEGELLEHLTPGLALG-AMAAPDGAVVS   80
Danio rerio               ISHPSALAS---HHHHHHHHHHQPHQALEGDLLDHITPGLALAGAMAGPDGSVVS  193
                          *******:.   ******************:**:*::*****. ***.***:***
```

Figure 40: DYRK1A phosphorylates a Ser residue in the N-terminal region of Brn-3b. A) Schematic summary of the results obtained in DYRK1A *in vitro* kinase assays with bacterial purified deletion mutants of Brn-3b. B) Sequence alignment of the Brn-3b minimal region phosphorylated by DYRK1A in several vertebrate organisms. Putative DYRK1A-target residues that were subjected to site-directed mutagenesis (Ser to Ala) are highlighted in red.

Interestingly, the phosphorylated region is highly conserved when Brn-3b orthologue proteins from different species are compared (Figure 40B). The sequence includes 17 Ser or Thr residues conserved in all the species present in the alignment. Two of them are followed by a Pro residue (Ser^{136} and Ser^{198}), and another by a Val residue (Ser^{160}) (Figure 40B, residues in red). Considering that some DYRK1A substrates have a Pro or a Val at the +1 position, these three Ser residues were mutated to Ala in the context of the first N-terminal half of the protein. However, no loss of phosphorylation was observed in any of the mutants tested (data not shown), suggesting that the target residue of DYRK1A should be any of the other Ser or Thr residues present in the fragment.

7.7 The functional output of the interaction between DYRK1A and Brn-3b

The interaction and subsequent phosphorylation of Brn-3b by DYRK1A can have several functional consequences, including a change in Brn-3b subcellular localization or alterations in its transcriptional activity. Previous results showed that, both in physiological and over-expression conditions, DYRK1A and Brn-3b co-localize in nuclear speckles (see Figure 34). This co-localization does not depend on DYRK1A activity or interaction since neither the

DYRK1A kinase-dead mutant nor the ΔHis mutant altered Brn-3b accumulation in the SFC (Figure 41).

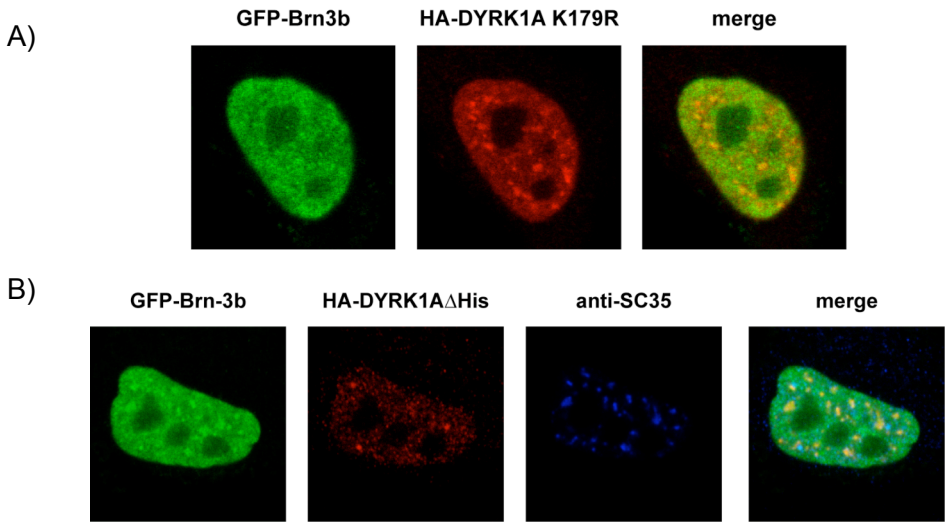


Figure 41: The expression of DYRK1A mutants does not modify the accumulation of Brn-3b in nuclear speckles. HeLa cells were transfected with GFP-Brn-3b and HA-DYRK1A K179R or ΔHis expression plasmids. At 48 h after transfection, cells were doubly immunostained with anti-HA and anti-SC35 antibodies. Alexa 555- conjugated goat anti-rabbit and Alexa 647-conjugated goat anti-mouse secondary antibodies were used, respectively.

In its role as a transcription factor, Brn-3b mainly modulates the transcription of genes implicated in retinogenesis (Mu et al., 2004; Plaza et al., 1999; Qiu et al., 2008) and the control of the cell cycle (Budhram-Mahadeo et al., 1999; Budhram-Mahadeo et al., 2008; Samady et al., 2004). To test whether the expression of DYRK1A affects Brn-3b-dependant transcription, a reporter assay was performed using a construct in which the luciferase gene is under the control of the Brn-3a promoter, which responds to Brn-3b transcription factors (Turner et al., 1997). Figure 42 shows the increase in Brn-3b transcriptional activity when DYRK1A is co-expressed, which is dependent on the amount of kinase over-expressed.

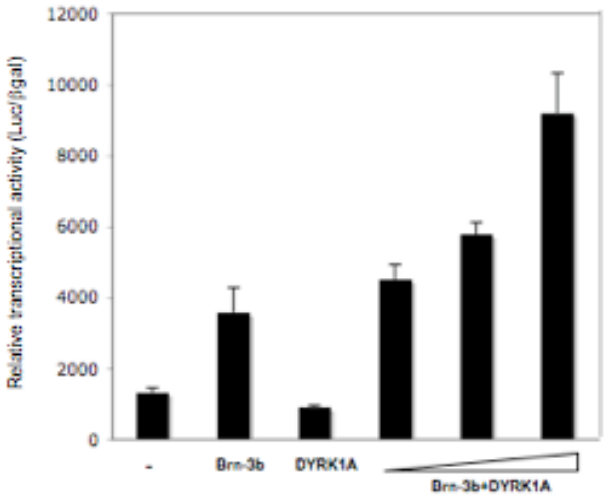


Figure 42: DYRK1A activates the transcriptional activity of Brn-3b. CV-1 cells were transfected with expression plasmids pCMV-β-galactosidase, pGL2-3xBrn-3a, Flag-Brn-3b and increasing amounts of HA-DYRK1A. At 48 h after transfection, luciferase activity was measured and corrected by β-galactosidase expression.

DYRK1A acts as a co-activator of several transcription factors, in a process that is completely independent on the kinase activity of the protein (Sitz et al., 2004; von Groote-Bidlingmaier et al., 2003). To test whether it also occurs in the case of Brn-3b co-activation, the same luciferase reporter assay has been carried out using the mutant of DYRK1A K179R. However, no concluding results have been obtained up to date.

In summary, DYRK1A interacts with and phosphorylates the transcription factor Brn-3b. Preliminary results suggest that this interaction results in an increase of the transcriptional activity of Brn-3b.

DISCUSSION

The aim of the work presented in this thesis has been to associate novel functions to the protein kinase DYRK1A in the nucleus. Given that the His repeat of DYRK1A had been previously defined as the protein region responsible for targeting to the nuclear speckles, the analysis of the mechanisms responsible for this targeting has been used as a way to understand the role of DYRK1A in the nucleus. Interestingly, studies performed with other polyHis-containing proteins have allowed defining His homopolymeic tracts as novel and general nuclear speckle-targeting signals. In addition, our results have contributed to enlarge the list of DYRK1A interactors and substrates with the RNA polymerase II and the transcription factor Brn-3b, among others. These findings leave the door opened to future studies directed to the characterization of the role of DYRK1A in the cell nucleus.

Previous work by M. Álvarez demonstrated that, when over-expressed in mammalian cells, DYRK1A localized in nuclear speckles through its polyHis domain (Alvarez et al., 2003). The first block of results presented in this report confirm His repeats as a *bona fide* SFC-targeting signal in DYRK1A, since they rule out the possibility that the localization of the kinase in nuclear speckles is an artefact due to protein over-expression or intranuclear protein aggregation, an output frequently associated to single amino acid repeats (Oma et al., 2004). In addition, they assure that the experimental approach was valid for the analysis of other polyHis-containing proteins.

His runs as a novel nuclear targeting signal

Nuclear compartments are fundamentally different from cytoplasmic organelles. In the cytoplasm, components of a metabolic pathway must be directed into a pre-existing membrane-delineated organelle in order to assemble a functional entity. In the nucleus, functional compartments such as repairosomes (Volker et al., 2001) or spliceosomes (Misteli and Spector, 1999) are generated “on the spot” by direct recruitment and assembly of catalytically active complexes with their substrates. In the recent years, a large number of nuclear compartments associated to different activities within the nucleus have been reported. One of them, the nuclear speckles, is the object of this study.

The mechanisms responsible for the formation of the nuclear speckles are not yet known, although it has been proposed that they are highly dynamic self-organizing entities in which protein-protein or protein-RNA interactions play a role (Misteli, 2001b). Up to date, only a few domains have been indeed described to direct proteins to nuclear speckles: the arginine/serine-rich (RS)-domain in SR proteins (Hedley et al., 1995) and in the kinases CrkRS (Ko et al., 2001) and PRP4 (Dellaire et al., 2002) and the RNA recognition motif (RRM) of the PTB-associated splicing factor and of the SF family of splicing factors (Caceres

et al., 1997; Dye and Patton, 2001). Moreover, other regions in specific proteins have also been reported to act as speckles-localizing sequences, like the threonine-proline repeats in $SF3b^{155}$ (Eilbracht and Schmidt-Zachmann, 2001), the Sm central domain of the protein U1 snRNP (Malatesta et al., 1999), the "Forkhead Associated" domain of the phosphatase NIPP1 (Jagiello et al., 2000), the ankyrin domain in IκBL (Semple et al., 2002), the FF domains in the transcription elongation factor CA150 (Sanchez-Alvarez et al., 2006) or the PDZ domains in syntenin-2 (Mortier et al., 2005).

Previous work from our lab had identified as a nuclear speckle-targeting signal in DYRK1A and cyclin T1 a segment of His residues (Alvarez et al., 2003). Given that His repeats are relatively rare and show a restricted distribution (Faux et al., 2005), it was plausible to think that the role as a subnuclear-directing signal could be more generally found in other proteins containing His runs. A bioinformatics screen found only 86 proteins with polyHis segments in the human proteome. Their analysis indicated that His tracts may not be neutral and are very likely associated to a functional role. First, most of the 86 proteins are nuclear and have functions related to nucleic acids metabolism, that is, synthesis and processing of RNA. Second, there is an overrepresentation of essential developmental factors, as well as nervous system and neurological class of proteins. This may be linked to the necessity for a finer regulation of the function of this kind of proteins, which might be facilitated by the particular structural and functional characteristics of His repeats. Third, the longest His tract present in the databases only contained 15 residues, in contrast to other amino acid repeats that can reach up to 60 residues. Moreover, the length of the repeat seems to be quite conserved among species. Both observations are signs of the presence of a strong selection against the expansion of His tracts. In fact, (CAC)n microsatellites present low levels of heterozygosity and, therefore, a significant expansion of such triplet repeats might not be expected (Sertedaki and Lindsay, 1996). All these features may be an indication of functional restrictions associated to the His runs.

Our subcellular localization analysis of a representative set of polyHis-containing proteins demonstrated that His runs act as SFC-directing sequences. Both the fusion of a His run to a heterologous protein and the deletion of the His repeat in several polyHis proteins has allowed to define His runs as both necessary and sufficient to act as speckle-localizing signals. Two types of repeats were observed in the set of His-rich proteins: pure and extended repeats. This difference could have been a clue to determine which proteins are targeted to the SFC and which are not. However, neither the repeat length nor its interrupting amino acid composition seemed to determine a pattern of positive or negative accumulation, which would indicate a high degree of flexibility in the functional signal. Therefore, it has not been possible to elucidate the internal code underlying targeting to nuclear speckles

mediated by His repeats. Despite this, a minimum of six His residues has been defined as the threshold for such localization. This finding should be considered in the light of His-tag protein purification, since the 6xHis tag could be modifying the subcellular localization of the fusion protein.

It is widely accepted that amino acid repeats function as facilitators of evolution, providing abundant and robust variation and thus enabling extremely rapid evolution of new forms. For instance, the expansion and contraction of repeats within transcription factors are linked to major morphological changes in dogs (Fondon and Garner, 2004). An interesting observation reported in this thesis work is that the presence of a His repeat may confer a novel function to the host protein. According to the results obtained by our collaborator M. Albà, 78% of the polyHis-containing proteins belong to protein families with paralogous that do not contain any His repeat. This is in contrast wit the results reported for other SARs, since only 22% of proteins with repeats, as an average, have paralogous proteins (Siwach et al., 2006). As demonstrated in this work, the presence of a His tract causes a change in the subnuclear localization of the protein, associating of a novel function to the paralogous with the His repeat. Moreover, the His run may induce alterations in putative interactions that finally result in overall changes in the gene expression regulatory networks depending on these transcription factors. Considering that a high proportion of polyHis-containing proteins have roles in developmental processes, His runs may have played a role in evolutionary diversification. This has been recently demonstrated for a polyAla tract in the transcription factor HOXD13. In this case, the specific loss of the repeat causes visible morphological changes, proving that taxon-specific homopolymeric tracts are involved in phenotypic diversification at the organism level (Anan et al., 2007).

A notorious feature of SARs is the possibility of suffering expansions and deletions that are associated to developmental and neurodegenerative diseases (Albrecht and Mundlos, 2005; Gatchel and Zoghbi, 2005; Karlin et al., 2002). No pathological expansions/deletions have been reported for His runs. However, Paraguison and co-workers identified novel variants of HOXA1 gene encoding various polyHis repeat lengths, both deletions and expansions, in the Japanese population. The ectopic expression of the expanded forms (11 and 12 His) revealed enhanced intranuclear aggregation and increased cell death in a time- and repeat length-dependent manner, suggesting that polyHis may have an important role for correct folding of this transcription factor (Paraguison et al., 2005). In addition, overexpression of HOXA1 expanded and deleted polyHis variants inhibited neurite outgrowth during retinoic acid-induced neuronal differentiation in neuroblastoma cells (Paraguison et al., 2007). Moreover, the expansion of the His tract resulted in impaired cooperative binding with PBX1, a known HOX cofactor, which in turn caused a decrease in

HOXA1 transcriptional activity (Paraguison et al., 2007). This, together with the length restrictions discussed above and the fact that some of the polyHis-containing proteins have critical roles in essential developmental processes, may be an indication that expansions could be linked to disease. In this context, a possible association between a His tract polymorphism in the transcription factor ZIC2 (from 9 His to 10 His) and neural tube defects has been reported (Brown et al., 2002).

Transit through nuclear speckles is a dynamic property of polyHis-containing proteins

Another conclusion extracted from this work is that proteins with His-rich domains have a dynamic behaviour. This would be consistent with the theory of the self-assembly process, in which proteins interact with each other to form morphological structures such as the nuclear speckle. Low-affinity interactions, such as those provided by the homopolymeric His runs, would allow a rapid interchange of molecules between compartments. For instance, it has been calculated that for the splicing factor SF2/ASF the residence time in SFCs is not more than 45 seconds and that at least 10,000 molecules of SF2/ASF are lost from SFCs per second per cell (Phair and Misteli, 2000).

Most of the poly-His proteins that did not accumulate in the SFC were transcription factors which contain domains that may mediate their localization to specific compartments within the nucleus, such as DNA binding regions or protein/protein interaction domains. In fact, polyHis-containing proteins with a nucleoplasmic staining pattern re-localized to nuclear speckles after loss of DNA binding activity. A similar behaviour has been recently described for the transcription factor GATA-4, although the subnuclear compartment was not identified (Philips et al., 2007). The deletion of the protein-protein interaction domains in CBX4 that target this protein to PML nuclear bodies also caused the re-localization to the SFC. Our results support, therefore, the existence of some kind of competition between His-repeats and other protein regions as a way to regulate intranuclear distribution of His-rich proteins. Treatment with an inhibitor of RNA polymerase II-dependent transcription, which would induce loss of DNA binding for some transcription factors, caused the accumulation in nuclear speckles for most of the proteins analyzed that showed a dispersed nuclear staining in the steady state. These results allow us to propose a dynamic model in which the intranuclear localization of some transcription factors with His-repeats is the net result of competition for binding to different recruiting sites, DNA, nuclear speckles or other nuclear bodies within the nuclei. Moreover, the dynamic behaviour could give reasons for the fact that only OTX1 appeared in a proteomic analysis of enriched preparations of interchromatin

granule clusters (Saitoh et al., 2004); the proteomic analysis would not consider proteins present in low amounts and/or proteins present transiently.

It is broadly accepted that RNA processing occurs co-transcriptionally, and thus, co-localization of factors related to RNA biogenesis, such as transcription and splicing factors occurs (Maniatis and Reed, 2002). When needed, transcription factors are recruited to specific promoters in active transcription sites, whereas splicing factors are assembled into the spliceosome. During periods of inactive transcription, the splicing factors re-locate to the speckle domains, and some transcription factors could also behave similarly. Transit through the speckles may offer the occasion for the transcription factor to come across RNA processing factors and assemble into complexes working on the same gene. In the case of DYRK1A, localization in the speckle may have a role in its ability to regulate splicing events, since it has been recently described to be a modulator of the ASF/SF2-regulated alternative splicing of tau transcripts (Shi et al., 2008). The re-localization may also represent targeting of transcription factors that are not longer able to bind DNA to other compartments for degradation or other processing activities (von Mikecz, 2006). In addition, compartmentalization of transcription-related proteins within distinct nuclear bodies has been suggested to act as an important mechanism for regulating gene expression. For instance, the inactivation of the transcription factor HAND1 by nucleoli retention has been implicated in trophoblast stem cell proliferation and renewal (Martindill et al., 2007), and estrogen receptor-enhanced transcription requires interchromosomal interactions occurring at the nuclear speckles (Nunez et al., 2008). Therefore, the existence of different binding domains to distinct nuclear organelles within a protein, being one of them His runs, implies that a variety of mechanisms should exist to modulate binding affinities and, consequently, to regulate its dynamics within the nucleus.

The bioinformatics screen showed that His repeats are present in several members of different protein families, such as HOX, ZIC, POU, DLX, FOX, MAF, NKD and ONECUT. The independent appearance of polyHis domains in these proteins could point to a process of convergent evolution, by which proteins sharing a functional role would tend to develop favourable common features. Actually, this has been demonstrated with exhaustive phylogenetic analyses of polyAla-containing protein families (Lavoie et al., 2003). The presence of the His-repeats to direct a subset of proteins to the nuclear speckles may confer functional advantages. First, it may represent a way to concentrate proteins that are functionally related, perhaps facilitating their physical interaction. The ability of several of the His-containing proteins to interact among them, as shown in this work, would indeed support this hypothesis. Second, it may be at the basis of a common mechanism to regulate these proteins: given that most of the polyHis-containing proteins are involved in developmental

processes, it may represent a way of to keep transcription factors away from promoters when not required.

Molecular mechanisms of His runs as SFC-targeting signals

As previously discussed, nuclear compartments are assumed to be the reflection of the association/dissociation rates of their residents between the nucleoplasmic space and transient and low-affinity interactions with locally immobilized binding sites (Carmo-Fonseca, 2002; Misteli, 2001a). Not much is known about how proteins are recruited to the SFC, even though for those of them for which a targeting signal has been characterized. However, it has been suggested that some speckle resident proteins play structural roles in facilitating other interactions as in the case of Par-6 (Cline and Nelson, 2007) or syntenin-2 (Mortier et al., 2005). This could be also true for the set of polyHis-containing proteins: those that accumulate in the speckle at the steady state would be "resident" proteins, whereas proteins by which the SFC-targeting has to be forced would be "circulating" factors.

Whatever the case is, the presence of a His-rich surface would be an advantage to establish different low-affinity interactions. His is a versatile amino acid that can adopt flexible roles in conformation, in metal coordination and in various enzymatic activities (Karlin and Burge, 1996). The concentration of His residues in an extended conformation, such as a β-strand, displays a positive charge or charge gradient that could mediate protein-protein, protein-DNA/RNA or protein-substrate interactions, or orient protein subunits via eletrostatic interactions. His is also an excellent ligand for coordinating metal ions like Zn^{2+} or Cu^{2+} (Karlin et al., 2003). Moreover, His features make His runs susceptible of different levels of regulation. First, His is a phosphorylatable residue. Several His kinases and phosphatases have been described in eukaryotes and, although it is a relatively recent field, it has been suggested that His phosphorylation may be involved in cell signalling and cancer (reviewed in (Besant and Attwood, 2005; Steeg et al., 2003). Second, due to their chemical properties, His repeats mediate protein-protein interaction in a pH-dependent fashion. As the pKa of a His side chain is in the range of physiological pH values, a small shift in intracellular pH has a dramatic effect on the binding capacity of His repeats. By this mechanism, extracellular inductors like growth factors, which can increase the intracellular pH by 0.1-0.3 units (Moolenaar, 1986), might influence the association state of an entire network of transcription factors.

Therefore, His runs seem to be a good platform for establishing interactions. A substantial amount of effort during the thesis work has been directed to try to understand the mechanism involved in the His repeat-mediated targeting to the nuclear speckle. Since the

interest of our group is focused in DYRK1A, we have used this kinase as a model system to approach the problem.

Regarding the hypothesis of the His repeat as a nucleic acid-interacting surface, our results indicate that, although DYRK1A binds ssDNA and RNA, this is not mediated by the His run, at least in the conditions used. The initial characterization shows that binding to single-stranded nucleic acids is conserved in the homologous kinase DYRK1B, and depends on a sequence located in the N-terminal region of both proteins. An overlap between NLSs and DNA-binding regions exists in 90% of the proteins for which both domains are known (Cokol et al., 2000). In the case of DYRK1A, our results permit to conclude that the bipartite NLS located in the N-terminus is not responsible for binding. More experiments are required to finely map the binding site and find a functional role for this uncovered interaction.

To identify protein targets specific for His runs, a gene discovering strategy based on TAP-tag purification in mammalian cells (Tandem Affinity Purification; Rigaut et al., 1999) was initially followed. This approach, which has not been included in this thesis report, presented a variety of technical problems and was abandoned. Alternatively, a candidate strategy was also followed. According to the reasons exposed in the hypothesis paragraph of Section 5 (Results), the RNA polymerase II turned out to be a good candidate. However, our results showed that binding to RNA polymerase II was not mediated by the His repeat and, therefore, the polymerase cannot be the recruiting element in the speckle. Nevertheless, the biological importance of the finding justified a further characterization of the interaction (discussed in the next Section).

If the SFC accumulation of DYRK1A is not mediated by resident molecules such as RNA or the RNA polymerase II, which is then the mechanism of recruitment? It is known that many nuclear compartments, such as nuclear speckles, are formed through a process of self-assembly, by which transient molecular interactions are likely at the basis of their own morphogenesis. In fact, a striking feature of many proteins that reside in these compartments is the presence of self-interacting domains as in the case of the SR protein splicing factors or many nucleolar proteins that are highly charged, a fact that may facilitate their self-assembly (Misteli, 2001b). Moreover, association into macromolecular complexes implies the interaction not only between two proteins but also among a large number of them, establishing an interacting network. Proteins highly connected in a network are known as hub proteins (Dosztanyi et al., 2006). These hub proteins are molecules characterized by having disordered, low complexity regions that provide them with the necessary flexibility to establish low-affinity interactions with a great variety of proteins (Dunker et al., 2005). The primary sequence analysis of several polyHis-containing proteins with the IUPRED software

confirmed that His domains correlated with the more disordered regions, suggesting that polyHis-containing proteins could be suitable nodes for interacting networks (Figure D2).

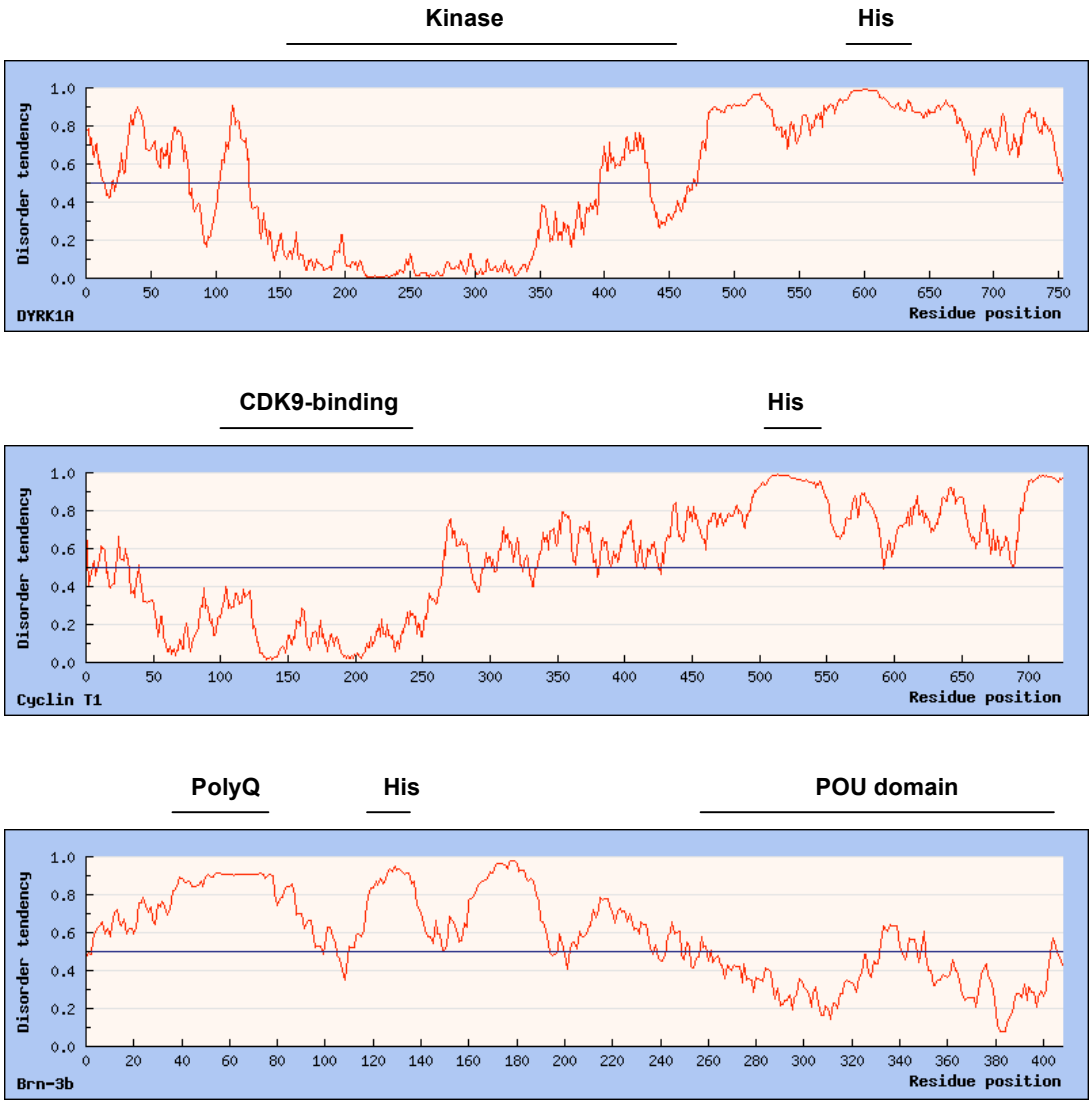


Figure D2: His-stretches are regions with high tendency to be disordered. The software IUPRED (http://iupred.enzim.hu/index.html) is a prediction method for recognizing ordered and disordered regions in proteins, based on estimating the capacity of polypeptides to form stabilizing contacts. The analysis was performed with the amino acidic sequences of DYRK1A (NP_01387), cyclin T1 (NP_001231) and Brn-3b (NP_004566). Note that the regions with highest disorder tendency correspond to the residues comprising the His-repeat: DYRK1A (590-616), cyclin T1 (513-530) and Brn-3b (171-182).

Moreover, it is known that unstructured proteins with repeats frequently interact with other structures built of repeated elements (reviewed in Kajava, 2001). This introduces a new possibility, by which proteins with His repeats could interact with each other. The results showing the interaction of DYRK1A and Brn-3b with several other polyHis proteins that reside in the speckle do support this hypothesis. Interestingly, several His-containing proteins have

been reported to participate in common molecular pathways. For instance, the bromodomain protein Brd4 interacts with cyclin T1 and CDK9 in nuclear speckles, causing an increased P-TEFb-dependent phosphorylation of the CTD and stimulation of transcription from promoters in vivo (Jang et al., 2005). Although physical interaction has not been reported, Dlx2 and Brn-3b are co-expressed in the ganglion cells of the retina. Both genes participate in the inner retinal development and in the terminal differentiation and/or maintenance of ganglion cells in the adult organism (de Melo et al., 2003). It is very plausible that other direct and indirect interactions exist among this subgroup of proteins. In fact, a large fraction of polyHis proteins also contained Ala, Gly, Ser, Pro or Gln tracts, which may provide interaction surfaces for different partners. An example of a network built with polyHis proteins is given in Figure D3.

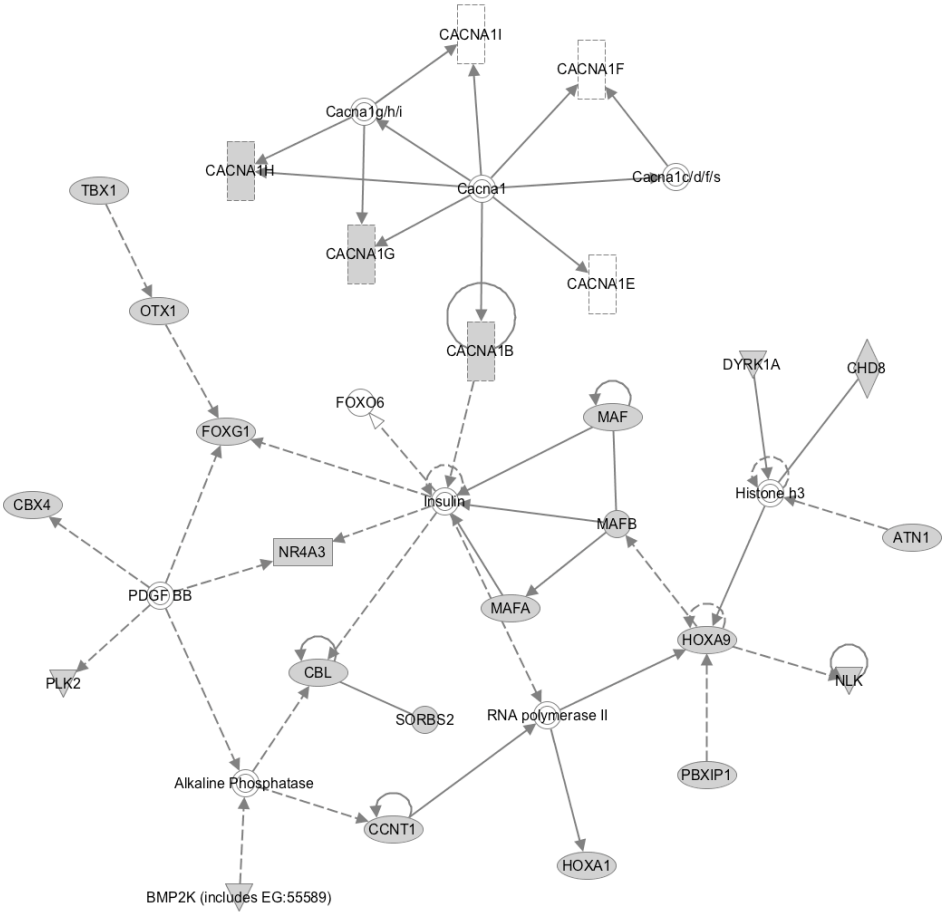


Figure D3: Functional analysis of the polyHis proteins dataset with the Ingenuity Pathway Analisys software. PolyHis proteins were distributed in 7 functional networks, from which the one included here had the highest score (representing the lowest probability to occur by chance). This network includes proteins with a role in organ development, auditory and vestibular system development and cardiovascular disease. The Network Generation Algorithm chose the intermediate proteins (in white) that show a relevant relationship with the proteins of interest (in grey) and allow the connection between them. Direct interactions are represented with a continuous line; indirect interactions with the discontinuous one. Shapes represent different molecular types.

The ability of the kinase DYRK1A to interact with His-proteins involved in different signalling pathways suggests it as one of the hubs in the speckle network. Moreover, this putative role is reinforced by the high conservation of the protein, its low evolution rate and by its complete requirement for the survival of the organism, other features of hub proteins (He and Zhang, 2006). In this sense, DYRK1A is one of the few kinases able to disassemble speckles upon overexpression (Alvarez et al., 2003), which would support an important role in speckle formation.

Unexpectedly, the interaction screen among His-containing proteins has also allowed uncovering novel DYRK1A partners. This is the case of Brn-3b, further studied within this thesis work, and NLK, an atypical MAPK that functions as a negative regulator of the Wnt signalling pathway (Ishitani et al., 1999). K. Arató is currently characterizing the functional outcome of the interaction between DYRK1A and NLK, which has lead to position DYRK1A within canonical Wnt signalling (K. Arato and S. de la Luna, unpublished results). Although DYRK1A phosphorylation has been analyzed only in these two proteins, other SFC-accumulating polyHis proteins interacting with DYRK1A have been found phosphorylated *in vivo* at residues that lie within DYRK1A consensus sequences of phosphorylation (Table D1) and, therefore, could be potential substrates of the kinase.

Table D1: residues of polyHis proteins susceptible of phosphorylation by DYRK1A

Protein	Phosphorylation site	Reference
CBX4	RKL**S**^{347}P KPE**T**^{495}P	Dephoure et al., 2008 Roscic et al., 2002
LMO6/PRICKLE1	RPR**S**^{502}P	Olsen et al., 2006 Beausoleil et al., 2004
NLK	YFP**S**^{96}P MPP**S**^{522}P	K. Arató and S. de la Luna, unpublished results
YY1	GEN**S**^{247}P	Dephoure et al., 2008 Beausoleil et al., 2004

DYRK1A is a CTD kinase

Our results show that DYRK1A directly interacts with the CTD of the RNA polymerase II and participates in protein complexes with the polymerase. Both findings strongly suggest that DYRK1A and RNA polymerase II interact. The data obtained indicates that the kinase activity of DYRK1A is not necessary for CTD binding and, furthermore, that the interaction seems to be mediated by the kinase domain. However, since the deletion of amino acids 377 to 474 in the DYRK1A mutant unable to bind eliminates kinase subdomains X and XI, it cannot be ruled out that the lack of interaction is caused by a conformational change in the mutant

protein. The results do not allow concluding whether there is a specific binding to the hypo- or hyperphosphorylated form of the RNA polymerase II. The lack of success in the initial attempt to co-immunoprecipitate DYRK1A with the 8WG16 antibody (data not shown), which recognizes the hypophosphorylated form of the CTD, could be due to specific binding of the kinase to the hyperphosphorylated fraction. In fact, preliminary results indicate that DYRK1A interacts with the phospho-Ser^2 fraction of the RNA polymerase II in nuclear extracts (data not shown). Alternatively, competition between DYRK1A and the antibody for binding to the same CTD region could exist. Indeed, this has been reported in the analysis of other CTD-binding proteins (Baillat et al., 2005).

The results have clearly demonstrated that DYRK1A is a new CTD-kinase. The study of the specific target residue suggested that DYRK1A preferentially phosphorylates Ser^2. However, cross-reaction of the two antibodies used (H5 and H14) has been reported at high antigen concentrations, and their detected bands should be taken as the sum of Ser^2+Ser^5 phosphorylation (Chapman et al., 2007; Jones et al., 2004). CDK9 phosphorylates both residues *in vitro*, but it is known to generate only phospho-Ser^2 *in vivo* (Ramanathan et al., 2001). Nevertheless, the similarities between DYRK1A and CDK9 would point to a participation of DYRK1A in the elongation process. We have found no differences in the CTD phosphorylation pattern when DYRK1A was over-expressed in cell lines (data not shown). Then, as an alternative to distinguish the specific effect of DYRK1A on RNA polymerase II phosphorylation, the CDK9-inhibitor KM05382 could be used (Medlin et al., 2005). Unfortunately, the compound also inhibited DYRK1A activity at low concentrations (data not shown), precluding this type of experiments. It could be therefore possible that DYRK1A contributes to the phosphorylation of only a fraction of the polymerase, unlike other CTD-kinases such as CDK9 that affect general CTD phosphorylation (Marshall et al., 1996). In this context, it can be hypothesized that DYRK1A would be recruited to specific promoters by binding to transcription factors as a way to be in close proximity to the RNA polymerase II. Examples of DYRK1A binding to transcription factors are found in the literature (see Table I2, Introduction) and also in this report (see Table 4, Results).

Interestingly, the observed differences among DYRK proteins in the binding to the RNA polymerase II coincide with their phylogenetical grouping in vertebrates (Figure D1). The ancestors of the DYRK subfamily of proteins appeared in yeast (Yak1p in *S. cerevisiae* and Pom1p In *S. pombe*) (Bahler and Pringle, 1998; Garrett et al., 1991), and gave rise to the two different lineages of DYRK proteins found in vertebrates: those related with Yak1p (DYRK1A and DYRK1B) and those arising from Pom1p (DYRK2, DYRK3 and DYRK4). Despite all these proteins may have functionally diverged during evolution, it is plausible that still share common features. This may be the case of DYRK1A, DYRK1B and DYRK4

interacting with RNA polymerase II. Even with these communalities, it is possible that they still behave as specific kinases. This specificity could be provided by the tissue of expression, the promoter context, by the co-activator/co-repressor molecules that are needed or by the time-window when they act. On the other hand, although both DYRK2 and DYRK3 have shown to be implicated in the regulation of transcription through the phosphorylation of several transcription factors (Gwack et al., 2006; Li et al., 2002; Taira et al., 2007), our results indicate that this function would be carried out without direct interaction with RNA polymerase II.

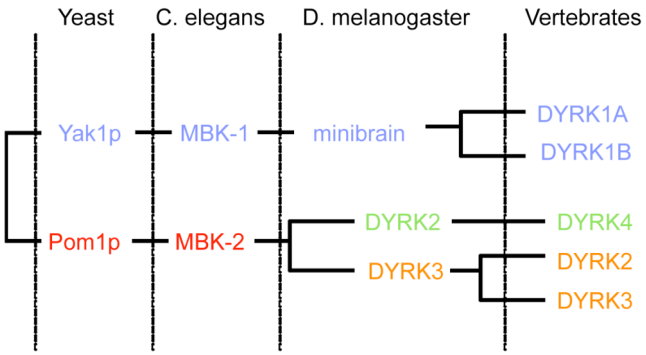


Figure D1: Phylogenetic tree of the DYRK subfamily of proteins.

In summary, the results obtained in this thesis work are highly consistent with DYRK1A having a direct role in the transcription process. Through its ability to bind nucleic acids, DYRK1A could be directly interacting with the nascent RNA. Moreover, it could be also bound to the single-stranded DNA while the double helix remains opened. DYRK1A could also bind to the promoter of a target gene when tethered by a transcription factor, and either regulate its transcriptional activity or act as a modulator of the transcription machinery. In fact, results recently obtained by C. di Vona (C. di Vona and S. de la Luna, unpublished results) using one-hybrid assays show that DYRK1A is indeed able to induce transcription in a kinase activity-dependent manner, suggesting that the phosphorylation of a component of the basal transcription machinery is responsible of the effect. This result resembles that described for other transcription-related kinases such as CDK9 (Taube et al., 2002). The idea of DYRK1A as a CTD-kinase, together with the ability to bind single-stranded nucleic acids and to interact and phosphorylate splicing factors (de Graaf et al., 2006; de Graaf et al., 2004; Shi et al., 2008), suggests that DYRK1A may be an important molecule in the coupling of transcription and RNA processing.

The protein kinase DYRK1A and the transcription factor Brn-3b interact

Brn-3b is a transcription factor required for the development of a set of RGCs (Gan et al., 1996). Although it is expressed at very early stages in development, it is not required for the initial commitment of RGC fate or the migration of these cells to the ganglion cell layer. However, Brn-3b is essential for the normal differentiation of these cells and in its absence, cells undergo apoptosis (Gan et al., 1999). It is known that Brn-3b exerts its function by regulating the activity of a set of genes whose products are required for the organization of the cytoskeleton and for normal cell polarity and axon outgrowth (Gan et al., 1999; Mu et al., 2004; Wang et al., 2000) and accordingly, absence of Brn-3b produces alterations in axon growth and pathfinding, which result in an inability to respond appropriately to extracellular cues (Erkman et al., 2000). These defects are likely to be the cause of the enhanced apoptosis of RGCs between E15.5 and birth in *Brn-3b* knock-out mice (Wang et al., 2000).

The main evidence pointing to a functional relationship between DYRK1A and Brn-3b comes from the data obtained in the analysis of the retina phenotype in *Dyrk1A*$^{+/-}$ and tg*Dyrk1A* mouse models (Laguna, 2008). Alterations in *Dyrk1A* dosage result in variations in the number of Brn-3b-expressing RGCs, as well as an impairment in the morphology of their axons (Laguna, 2008). The fact that the heterozygous model completely recapitulates the phenotype observed in the *Brn-3b* knock-out (Gan et al., 1999), strongly suggests that both proteins act in the same pathway.

Not much is known about the mechanisms that regulate Brn-3b transcriptional activity. The results presented in this thesis work indicate that the kinase DYRK1A is a novel regulator of Brn-3b. The two proteins directly interact and the binding depends both on the kinase activity of DYRK1A and on the presence of the His repeats in both proteins. Brn-3b has been shown to interact with p53 and the estrogen receptor via its POU domain (Budhram-Mahadeo et al., 1998; Budhram-Mahadeo et al., 2006); it is thought that the POU domain is also involved in the interaction with the DYRK kinase HIPK2 (Wiggins et al., 2004). Therefore, DYRK1A interaction would represent the first functional association to the His repeat in Brn-3b. Unlike HIPK2, DYRK1A only interacts with Brn-3b and not with the other two members of the Brn-3 family. The fact that Brn-3a has a shorter His tract and Brn-3c does not would explain the lack of binding. DYRK1A and Brn-3b co-localize in the same subnuclear structure, the nuclear speckles. Whether if they interact in the nuclear speckle or in the nucleus remains to be defined.

Phosphorylation is a key post-translational modification to regulate the activity of transcription factors. The addition of phosphate groups by cellular kinases may result in i) alteration of the half-life, ii) nuclear translocation or export, iii) increase/decrease of DNA binding capabilities, iv) modulation of the interaction with repressors or activators and iv)

release from resting compartments to active transcription sites. In the case of DYRK1A, it has been described that phosphorylation induces nuclear import of Gli1 (Mao et al., 2002) and nuclear export of NFAT transcription factors (Arron et al., 2006; Gwack et al., 2006); DYRK1A-mediated phosphorylation of CREB at Ser^{133} would increase association with the activator p300 (Yang et al., 2001). DYRK1A phosphorylates Brn-3b in one or several Ser residues comprised within amino acids 107 and 215. Although the exact residue could not been found in the frame of this work, it is worth to know that this region is included in the N-terminal transactivation domain of Brn-3b (amino acids 100-239) (Martin et al., 2005), suggesting that phosphorylation by DYRK1A would be directly modulating the transcriptional activity of the protein. In fact, preliminary results with Brn-3b-dependent reporters do support this proposal. In addition, no doublets appear in Brn-3a or Brn-3c when co-expressed with the kinase, suggesting that they are not DYRK1A substrates, in agreement with the lack of interaction observed.

Several works report the existence of a hierarchical gene regulatory network for the specification and differentiation of the RGC lineage, with transcription factors regulating the early events positioned at the top of the hierarchy and those for the late events at the bottom (reviewed in Mu and Klein, 2004) (Figure D4). Pax6, a homeobox-containing transcription factor expressed in progenitor cells and essential for eye field specification during early stages in eye development, is located at the top. It is followed by Math5, which promotes the exit of the cell cycle and the acquisition of RGC competence in a subpopulation of retinal progenitor cells. Math5 activates the expression of Brn-3b in migrating RGC precursors as well as in postmitotic and mature RGCs, where the factor is essential for RGC differentiation and survival (Gan et al., 1999). Brn-3b promotes both by activating the transcription of pro-differentiation genes and by suppressing non-RGC differentiation programs (Qiu et al., 2008). The data reported here indicate that DYRK1A would act as a pro-survival transcriptional cofactor by positively acting on the Brn-3b-dependent gene expression program in RGCs. Figure D4 shows a hypothetical model based on the results exposed in this thesis report and on experimental evidences reported by other groups during last years. The model has been simplified in order to facilitate the understanding of the genetic cascade discussed.

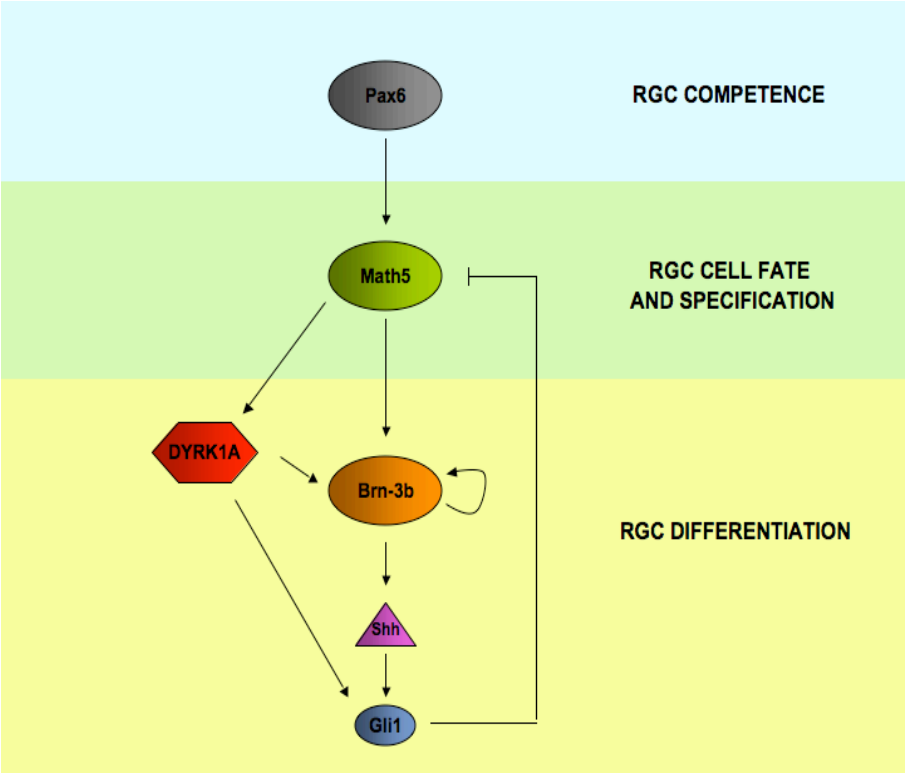


Figure D4: DYRK1A may be involved in the regulation of RGC differentiation or survival. The kinase DYRK1A increases the transcriptional activity of Brn-3b by phosphorylation of the N-terminal transactivation domain. Several other evidences that position DYRK1A in this scenario are i) DYRK1A is one of the genes downregulated in Math5$^{-/-}$ retinas (Mu et al., 2005); ii) DYRK1A activates the transcriptional activity of Gli1 and synergically acts with Sonic hedgehog (Shh) to promote Gli1 activity and cell differentiation in embryonic fibroblasts (Mao et al., 2002). Brn-3b activates the expression of Shh to inhibit RGC proliferation and retina patterning via Gli1, which establishes a negative feedback to maintain the number of RGC (Zhang and Yang, 2001). Therefore, it could be possible that DYRK1A cooperates with this pathway by directly activating Gli1 in RGCs. This scheme is very simplified and only represents a small fraction of the whole gene network of RGC development (a complete overview can be found in Mu et al., 2005).

The data discussed here regarding the functional implications of the interaction between DYRK1A and Brn-3b could be extrapolated to a conserved and more general mechanism involving DYRK family of kinases and Brn-3 family of transcription factors in neuronal survival. Evidences in the literature include the interaction between the DYRK protein HIPK2 and Brn-3a in sensory neurons (Wiggins et al., 2004). Although the binding promotes Brn-3a attachment to DNA, Brn-3a-dependent transcription of pro-survival genes is abolished. These results define HIPK2 as a pro-apoptotic co-factor. On the contrary, DYRK1A seems to exert a protective role in the retina according to the increased apoptosis detected in *Dyrk1A* heterozygous mices (Laguna, 2008). This could be accomplished both by promoting Brn-3b activity as shown here as well as by an inhibitory phosphorylation in caspase 9, a key effector of developmental programmed cell death (Laguna, 2008). Therefore, the functions of DYRK1A in the retina agree with the general and more and more accepted role of DYRK proteins in the regulation of cell fate and apoptosis in vertebrates (reviewed in Yoshida, 2008).

CONCLUSIONS

I. PolyHis repeats are general nuclear speckles-directing sequences, based in the subcellular localization analysis of a subset of polyHis-containing proteins.

II. Paralogous proteins without His repeats do not localize in nuclear speckles, suggesting that the presence of His tracts is a mechanism to acquire novel and probably more complex functions during evolution of proteins.

III. Transit through the speckle for most polyHis transcription factors is a dynamic process that depends on other interaction domains present in the proteins and on the transcriptional requirements of the cell at a specific time point.

IV. DYRK1A interacts with single-stranded nucleic acids, both DNA and RNA. The binding is direct and does not depend on the His tract, but is mediated by a region comprising the first 167 amino acids of the kinase. This activity is conserved in the homologous kinase DYRK1B. The fact that the His repeat is dispensable for binding to nucleic acids suggests that targeting of DYRK1A to nuclear speckles is not mediated by RNA interactions.

V. DYRK1A interacts with the CTD of the RNA polymerase II. The binding does not depend on DYRK1A kinase activity and seems to be mediated by a region whithin the kinase domain. This interaction is conserved in other human DYRK homologs such as DYRK1B and DYRK4.

VI. The CTD of the RNA polymerase II is a novel substrate of DYRK1A. Phosphorylation occurs at Ser^2 and Ser^5 residues of the heptapeptide YS_2PTS_5PS, suggesting that DYRK1A may have a role both in the initiation and the elongation steps during the transcription process.

VII. DYRK1A is present in protein complexes with RNA polymerase II and co-localize in the same subnuclear structures.

VIII. DYRK1A interacts with a significant proportion of the polyHis proteins that accumulate in nuclear speckles, mostly in a His-repeat independent manner. At their turn, other polyHis proteins interact among them, participating of a putative interaction network in the speckle.

IX. DYRK1A directly interacts with the polyHis-containing transcription factor Brn-3b. The kinase activity of DYRK1A is necessary but not sufficient for this binding. The interaction is mediated by the His repeats in each protein and is further support by the fact that both

proteins co-localize. The binding is specific of Brn-3b, since it is not observed with the homologs Brn-3a and Brn-3c.

X. Brn-3b is a novel substrate of DYRK1A. Phosphorylation occurs in a residue comprised between amino acids 107 and 215, which include a transcriptional activation domain. Preliminary results show that DYRK1A would activate the transcriptional activity of Brn-3b over a Brn-3 specific promoter.

REFERENCES

Adayev, T., Chen-Hwang, M. C., Murakami, N., Wang, R., and Hwang, Y. W. (2006). MNB/DYRK1A phosphorylation regulates the interactions of synaptojanin 1 with endocytic accessory proteins. *Biochem Biophys Res Commun,* **351**: 1060-1065.

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.

Alba, M. M., and Guigo, R. (2004). Comparative analysis of amino acid repeats in rodents and humans. *Genome Res,* **14**: 549-554.

Alba, M. M., Santibanez-Koref, M. F., and Hancock, J. M. (1999). Conservation of polyglutamine tract size between mice and humans depends on codon interruption. *Mol Biol Evol,* **16**: 1641-1644.

Albrecht, A., and Mundlos, S. (2005). The other trinucleotide repeat: polyalanine expansion disorders. *Curr Opin Genet Dev,* **15**: 285-293.

Altafaj, X., Dierssen, M., Baamonde, C., Marti, E., Visa, J., Guimera, J., Oset, M., Gonzalez, J. R., Florez, J., Fillat, C., and Estivill, X. (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. (2004). Localización subcelular de la proteína quinasa DYRK1A: compartimentos, señales y regulación. Universitat de Barcelona, Barcelona.

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*.

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.

Amiel, J., Trochet, D., Clement-Ziza, M., Munnich, A., and Lyonnet, S. (2004). Polyalanine expansions in human. *Hum Mol Genet,* **13 Spec No 2**: R235-243.

Anan, K., Yoshida, N., Kataoka, Y., Sato, M., Ichise, H., Nasu, M., and Ueda, S. (2007). Morphological change caused by loss of the taxon-specific polyalanine tract in Hoxd-13. *Mol Biol Evol,* **24**: 281-287.

Andersen, J. S., Lyon, C. E., Fox, A. H., Leung, A. K., Lam, Y. W., Steen, H., Mann, M., and Lamond, A. I. (2002). Directed proteomic analysis of the human nucleolus. *Curr Biol,* **12**: 1-11.

Arai, S., Matsushita, A., Du, K., Yagi, K., Okazaki, Y., and Kurokawa, R. (2007). Novel homeodomain-interacting protein kinase family member, HIPK4, phosphorylates human p53 at serine 9. *FEBS Lett,* **581**: 5649-5657.

Aranda, S., Alvarez, M., Turro, S., Laguna, A., and de la Luna, S. (2008). Sprouty2 inhibitory activity on FGF-signaling is modulated by the protein kinase DYRK1A. *Mol Cell Biol*.

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.

Bahler, J., and Nurse, P. (2001). Fission yeast Pom1p kinase activity is cell cycle regulated and essential for cellular symmetry during growth and division. *Embo J,* **20**: 1064-1073.

Bahler, J., and Pringle, J. R. (1998). Pom1p, a fission yeast protein kinase that provides positional information for both polarized growth and cytokinesis. *Genes Dev,* **12**: 1356-1370.

Baillat, D., Hakimi, M. A., Naar, A. M., Shilatifard, A., Cooch, N., and Shiekhattar, R. (2005). Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell,* **123**: 265-276.

Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villen, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004). Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A,* **101**: 12130-12135.

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.

Bernardi, R., and Pandolfi, P. P. (2007). Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol,* **8**: 1006-1016.

Besant, P. G., and Attwood, P. V. (2005). Mammalian histidine kinases. *Biochim Biophys Acta,* **1754**: 281-290.

Bescond, M., and Rahmani, Z. (2005). Dual-specificity tyrosine-phosphorylated and regulated kinase 1A (DYRK1A) interacts with the phytanoyl-CoA alpha-hydroxylase associated protein 1 (PAHX-AP1), a brain specific protein. *Int J Biochem Cell Biol,* **37**: 775-783.

Bleoo, S., Sun, X., Hendzel, M. J., Rowe, J. M., Packer, M., and Godbout, R. (2001). Association of human DEAD box protein DDX1 with a cleavage stimulation factor involved in 3'-end processing of pre-MRNA. *Mol Biol Cell,* **12**: 3046-3059.

Boisvert, F. M., van Koningsbruggen, S., Navascues, J., and Lamond, A. I. (2007). The multifunctional nucleolus. *Nat Rev Mol Cell Biol,* **8**: 574-585.

Bonnet, F., Vigneron, M., Bensaude, O., and Dubois, M. F. (1999). Transcription-independent phosphorylation of the RNA polymerase II C-terminal domain (CTD) involves ERK kinases (MEK1/2). *Nucleic Acids Res,* **27**: 4399-4404.

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.

Bres, V., Yoh, S. M., and Jones, K. A. (2008). The multi-tasking P-TEFb complex. *Curr Opin Cell Biol,* **20**: 334-340.

Brown, L., Paraso, M., Arkell, R., and Brown, S. (2005). In vitro analysis of partial loss-of-function ZIC2 mutations in holoprosencephaly: alanine tract expansion modulates DNA binding and transactivation. *Hum Mol Genet,* **14**: 411-420.

Brown, L. Y., and Brown, S. A. (2004). Alanine tracts: the expanding story of human illness and trinucleotide repeats. *Trends Genet,* **20**: 51-58.

Brown, L. Y., Hodge, S. E., Johnson, W. G., Guy, S. G., Nye, J. S., and Brown, S. (2002). Possible association of NTDs with a polyhistidine tract polymorphism in the ZIC2 gene. *Am J Med Genet,* **108**: 128-131.

Buchanan, G., Yang, M., Cheong, A., Harris, J. M., Irvine, R. A., Lambert, P. F., Moore, N. L., Raynor, M., Neufing, P. J., Coetzee, G. A., and Tilley, W. D. (2004). Structural and functional consequences of glutamine tract variation in the androgen receptor. *Hum Mol Genet,* **13**: 1677-1692.

Budhram-Mahadeo, V., Ndisang, D., Ward, T., Weber, B. L., and Latchman, D. S. (1999). The Brn-3b POU family transcription factor represses expression of the BRCA-1 anti-oncogene in breast cancer cells. *Oncogene,* **18**: 6684-6691.

Budhram-Mahadeo, V., Parker, M., and Latchman, D. S. (1998). POU transcription factors Brn-3a and Brn-3b interact with the estrogen receptor and differentially regulate transcriptional activity via an estrogen response element. *Mol Cell Biol,* **18**: 1029-1041.

Budhram-Mahadeo, V. S., Bowen, S., Lee, S., Perez-Sanchez, C., Ensor, E., Morris, P. J., and Latchman, D. S. (2006). Brn-3b enhances the pro-apoptotic effects of p53 but not its induction of cell cycle arrest by cooperating in trans-activation of bax expression. *Nucleic Acids Res,* **34**: 6640-6652.

Budhram-Mahadeo, V. S., Irshad, S., Bowen, S., Lee, S. A., Samady, L., Tonini, G. P., and Latchman, D. S. (2008). Proliferation-associated Brn-3b transcription factor can activate cyclin D1 expression in neuroblastoma and breast cancer cells. *Oncogene,* **27**: 145-154.

Bunce, M. W., Boronenkov, I. V., and Anderson, R. A. (2008). Coordinated activation of the nuclear ubiquitin ligase Cul3-SPOP by the generation of phosphatidylinositol 5-phosphate. *J Biol Chem,* **283**: 8678-8686.

Caceres, J. F., Misteli, T., Screaton, G. R., Spector, D. L., and Krainer, A. R. (1997). Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol,* **138**: 225-238.

Campbell, L. E., and Proud, C. G. (2002). Differing substrate specificities of members of the DYRK family of arginine-directed protein kinases. *FEBS Lett,* **510**: 31-36.

Carmo-Fonseca, M. (2002). The contribution of nuclear compartmentalization to gene regulation. *Cell,* **108**: 513-521.

Celton-Morizur, S., Racine, V., Sibarita, J. B., and Paoletti, A. (2006). Pom1 kinase links division plane position to cell polarity by regulating Mid1p cortical distribution. *J Cell Sci,* **119**: 4710-4718.

Chabert, C., Jamon, M., Cherfouh, A., Duquenne, V., Smith, D. J., Rubin, E., and Roubertoux, P. L. (2004). Functional analysis of genes implicated in Down syndrome: 1. Cognitive abilities in mice transpolygenic for Down Syndrome Chromosomal Region-1 (DCR-1). *Behav Genet,* **34**: 559-569.

Chang, H. S., Lin, C. H., Yang, C. H., Yen, M. S., Lai, C. R., Chen, Y. R., Liang, Y. J., and Yu, W. C. (2007). Increased expression of Dyrk1a in HPV16 immortalized Keratinocytes enable evasion of apoptosis. *Int J Cancer*.

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, T., Boisvert, F. M., Bazett-Jones, D. P., and Richard, S. (1999). A role for the GSG domain in localizing Sam68 to novel nuclear structures in cancer cell lines. *Mol Biol Cell,* **10**: 3015-3033.

Chen-Hwang, M. C., Chen, H. R., Elzinga, M., and Hwang, Y. W. (2002). Dynamin is a minibrain kinase/dual specificity Yak1-related kinase 1A substrate. *J Biol Chem,* **277**: 17597-17604.

Cho, E. J. (2007). RNA polymerase II carboxy-terminal domain with multiple connections. *Exp Mol Med,* **39**: 247-254.

Choong, M. L., Luo, B., and Lodish, H. F. (2004). Microenvironment-driven changes in the expression profile of hematopoietic cobblestone area-forming cells. *Ann Hematol,* **83**: 160-169.

Cioce, M., and Lamond, A. I. (2005). Cajal bodies: a long history of discovery. *Annu Rev Cell Dev Biol,* **21**: 105-131.

Cline, E. G., and Nelson, W. J. (2007). Characterization of mammalian Par 6 as a dual-location protein. *Mol Cell Biol,* **27**: 4431-4443.

Cokol, M., Nair, R., and Rost, B. (2000). Finding nuclear localization signals. *EMBO Rep,* **1**: 411-415.

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.

Corden, J. L., Cadena, D. L., Ahearn, J. M., Jr., and Dahmus, M. E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A,* **82**: 7934-7938.

Coute, Y., Burgess, J. A., Diaz, J. J., Chichester, C., Lisacek, F., Greco, A., and Sanchez, J. C. (2006). Deciphering the human nucleolar proteome. *Mass Spectrom Rev,* **25**: 215-234.

Cremer, T., and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet,* **2**: 292-301.

D'Orazi, G., Cecchinelli, B., Bruno, T., Manni, I., Higashimoto, Y., Saito, S., Gostissa, M., Coen, S., Marchetti, A., Del Sal, G., *et al.* (2002). Homeodomain-interacting protein kinase-2 phosphorylates p53 at Ser 46 and mediates apoptosis. *Nat Cell Biol,* **4**: 11-19.

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 Melo, J., Du, G., Fonseca, M., Gillespie, L. A., Turk, W. J., Rubenstein, J. L., and Eisenstat, D. D. (2005). Dlx1 and Dlx2 function is necessary for terminal differentiation and survival of late-born retinal ganglion cells in the developing mouse retina. *Development,* **132**: 311-322.

de Melo, J., Qiu, X., Du, G., Cristante, L., and Eisenstat, D. D. (2003). Dlx1, Dlx2, Pax6, Brn3b, and Chx10 homeobox gene expression defines the retinal ganglion and inner nuclear layers of the developing and adult mouse retina. *J Comp Neurol,* **461**: 187-204.

de Wit, N. J., Burtscher, H. J., Weidle, U. H., Ruiter, D. J., and van Muijen, G. N. (2002). Differentially expressed genes identified in human melanoma cell lines with different metastatic behaviour using high density oligonucleotide arrays. *Melanoma Res,* **12**: 57-69.

Dellaire, G., Makarov, E. M., Cowger, J. J., Longman, D., Sutherland, H. G., Luhrmann, R., Torchia, J., and Bickmore, W. A. (2002). Mammalian PRP4 kinase copurifies and interacts with components of both the U5 snRNP and the N-CoR deacetylase complexes. *Mol Cell Biol,* **22**: 5141-5156.

Delot, E., King, L. M., Briggs, M. D., Wilcox, W. R., and Cohn, D. H. (1999). Trinucleotide expansion mutations in the cartilage oligomeric matrix protein (COMP) gene. *Hum Mol Genet,* **8**: 123-128.

Deng, X., Ewton, D. Z., Li, S., Naqvi, A., Mercer, S. E., Landas, S., and Friedman, E. (2006). The kinase Mirk/Dyrk1B mediates cell survival in pancreatic ductal adenocarcinoma. *Cancer Res,* **66**: 4149-4158.

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.

Deng, X., Ewton, D. Z., Pawlikowski, B., Maimone, M., and Friedman, E. (2003). Mirk/dyrk1B is a Rho-induced kinase active in skeletal muscle differentiation. *J Biol Chem,* **278**: 41347-41354.

Deng, X., Mercer, S. E., Shah, S., Ewton, D. Z., and Friedman, E. (2004). The cyclin-dependent kinase inhibitor p27Kip1 is stabilized in G(0) by Mirk/dyrk1B kinase. *J Biol Chem,* **279**: 22498-22504.

Dephoure, N., Zhou, C., Villen, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J., and Gygi, S. P. (2008). A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A,* **105**: 10762-10767.

Dosztanyi, Z., Chen, J., Dunker, A. K., Simon, I., and Tompa, P. (2006). Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res,* **5**: 2985-2995.

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.

Dunah, A. W., Jeong, H., Griffin, A., Kim, Y. M., Standaert, D. G., Hersch, S. M., Mouradian, M. M., Young, A. B., Tanese, N., and Krainc, D. (2002). Sp1 and TAFII130 transcriptional activity disrupted in early Huntington's disease. *Science,* **296**: 2238-2243.

Dundr, M., and Misteli, T. (2001). Functional architecture in the cell nucleus. *Biochem J,* **356**: 297-310.

Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002). Intrinsic disorder and protein function. *Biochemistry,* **41**: 6573-6582.

Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005). Flexible nets. The roles of intrinsic disorder in protein interaction networks. *Febs J,* **272**: 5129-5148.

Dye, B. T., and Patton, J. G. (2001). An RNA recognition motif (RRM) is required for the localization of PTB-associated splicing factor (PSF) to subnuclear speckles. *Exp Cell Res,* **263**: 131-144.

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.

Eilbracht, J., and Schmidt-Zachmann, M. S. (2001). Identification of a sequence element directing a protein to nuclear speckles. *Proc Natl Acad Sci U S A,* **98**: 3849-3854.

Epstein, C. J. (1990). The consequences of chromosome imbalance. *Am J Med Genet Suppl,* **7**: 31-37.

Erkman, L., Yates, P. A., McLaughlin, T., McEvilly, R. J., Whisenhunt, T., O'Connell, S. M., Krones, A. I., Kirby, M. A., Rapaport, D. H., Bermingham, J. R.*, et al.* (2000). A POU domain transcription factor-dependent program regulates axon pathfinding in the vertebrate visual system. *Neuron,* **28**: 779-792.

Ewton, D. Z., Lee, K., Deng, X., Lim, S., and Friedman, E. (2003). Rapid turnover of cell-cycle regulators found in Mirk/dyrk1B transfectants. *Int J Cancer,* **103**: 21-28.

Faux, N. G., Bottomley, S. P., Lesk, A. M., Irving, J. A., Morrison, J. R., de la Banda, M. G., and Whisstock, J. C. (2005). Functional insights from the distribution and role of homopeptide repeat-containing proteins. *Genome Res,* **15**: 537-551.

Faux, N. G., Huttley, G. A., Mahmood, K., Webb, G. I., de la Banda, M. G., and Whisstock, J. C. (2007). RCPdb: An evolutionary classification and codon usage database for repeat-containing proteins. *Genome Res,* **17**: 1118-1127.

Fedtsova, N., Quina, L. A., Wang, S., and Turner, E. E. (2008). Regulation of the development of tectal neurons and their projections by transcription factors Brn3a and Pax7. *Dev Biol,* **316**: 6-20.

Fiol, C. J., Haseman, J. H., Wang, Y. H., Roach, P. J., Roeske, R. W., Kowalczuk, M., and DePaoli-Roach, A. A. (1988). Phosphoserine as a recognition determinant for glycogen synthase kinase-3: phosphorylation of a synthetic peptide based on the G-component of protein phosphatase-1. *Arch Biochem Biophys,* **267**: 797-802.

Fischer, U., Liu, Q., and Dreyfuss, G. (1997). The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell,* **90**: 1023-1029.

Flicek, P., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T.*, et al.* (2008). Ensembl 2008. *Nucleic Acids Res,* **36**: D707-714.

Fondon, J. W., 3rd, and Garner, H. R. (2004). Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A,* **101**: 18058-18063.

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.

Fotaki, V., Martinez De Lagran, M., Estivill, X., Arbones, M., and Dierssen, M. (2004). Haploinsufficiency of Dyrk1A in mice leads to specific alterations in the development and regulation of motor activity. *Behav Neurosci,* **118**: 815-821.

Fox, A. H., Lam, Y. W., Leung, A. K., Lyon, C. E., Andersen, J., Mann, M., and Lamond, A. I. (2002). Paraspeckles: a novel nuclear domain. *Curr Biol,* **12**: 13-25.

Friedman, E. (2007). Mirk/Dyrk1B in cancer. *J Cell Biochem,* **102**: 274-279.

Friedman, M. J., Shah, A. G., Fang, Z. H., Ward, E. G., Warren, S. T., Li, S., and Li, X. J. (2007). Polyglutamine domain modulates the TBP-TFIIB interaction: implications for its normal function and neurodegeneration. *Nat Neurosci,* **10**: 1519-1528.

Frigerio, G., Burri, M., Bopp, D., Baumgartner, S., and Noll, M. (1986). Structure of the segmentation gene paired and the Drosophila PRD gene set as part of a gene network. *Cell,* **47**: 735-746.

Fu, X. D., and Maniatis, T. (1990). Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature,* **343**: 437-441.

Galant, R., and Carroll, S. B. (2002). Evolution of a transcriptional repression domain in an insect Hox protein. *Nature,* **415**: 910-913.

Gall, J. G. (2000). Cajal bodies: the first 100 years. *Annu Rev Cell Dev Biol,* **16**: 273-300.

Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M., and Mackay, J. P. (2007). Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci,* **32**: 63-70.

Gan, L., Wang, S. W., Huang, Z., and Klein, W. H. (1999). POU domain factor Brn-3b is essential for retinal ganglion cell differentiation and survival but not for initial cell fate specification. *Dev Biol,* **210**: 469-480.

Gan, L., Xiang, M., Zhou, L., Wagner, D. S., Klein, W. H., and Nathans, J. (1996). POU domain factor Brn-3b is required for the development of a large set of retinal ganglion cells. *Proc Natl Acad Sci U S A,* **93**: 3920-3925.

Garrett, S., Menold, M. M., and Broach, J. R. (1991). The Saccharomyces cerevisiae YAK1 gene encodes a protein kinase that is induced by arrest early in the cell cycle. *Mol Cell Biol,* **11**: 4045-4052.

Gatchel, J. R., and Zoghbi, H. Y. (2005). Diseases of unstable repeat expansion: mechanisms and common principles. *Nat Rev Genet,* **6**: 743-755.

Geiger, J. N., Knudsen, G. T., Panek, L., Pandit, A. K., Yoder, M. D., Lord, K. A., Creasy, C. L., Burns, B. M., Gaines, P., Dillon, S. B., and Wojchowski, D. M. (2001). mDYRK3 kinase is expressed selectively in late erythroid progenitor cells and attenuates colony-forming unit-erythroid development. *Blood,* **97**: 901-910.

Gerber, H. P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S., and Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science,* **263**: 808-811.

Gerhard, D. S., Wagner, L., Feingold, E. A., Shenmen, C. M., Grouse, L. H., Schuler, G., Klein, S. L., Old, S., Rasooly, R., Good, P.*, et al.* (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res,* **14**: 2121-2127.

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.

Green, H., and Wang, N. (1994). Codon reiteration and the evolution of proteins. *Proc Natl Acad Sci U S A,* **91**: 4298-4302.

Gruber, C. A., Rhee, J. M., Gleiberman, A., and Turner, E. E. (1997). POU domain factors of the Brn-3 class recognize functional DNA elements which are distinctive, symmetrical, and highly conserved in evolution. *Mol Cell Biol,* **17**: 2391-2400.

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., Pritchard, M., Nadal, M., and Estivill, X. (1997a). Minibrain (MNBH) is a single copy gene mapping to human chromosome 21q22.2. *Cytogenet Cell Genet,* **77**: 182-184.

Guimera, J., Pucharcos, C., Domenech, A., Casas, C., Solans, A., Gallardo, T., Ashley, J., Lovett, M., Estivill, X., and Pritchard, M. (1997b). 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, B., Odgren, P. R., van Wijnen, A. J., Last, T. J., Nickerson, J., Penman, S., Lian, J. B., Stein, J. L., and Stein, G. S. (1995). The nuclear matrix protein NMP-1 is the transcription factor YY1. *Proc Natl Acad Sci U S A,* **92**: 10526-10530.

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

Haberman, Y., Amariglio, N., Rechavi, G., and Eisenberg, E. (2008). Trinucleotide repeats are prevalent among cancer-related genes. *Trends Genet,* **24**: 14-18.

Hammerle, B., Carnicero, A., Elizalde, C., Ceron, J., Martinez, S., and Tejedor, F. J. (2003a). 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., Elizalde, C., Galceran, J., Becker, W., and Tejedor, F. J. (2003b). The MNB/DYRK1A protein kinase: neurobiological functions and Down syndrome implications. *J Neural Transm Suppl*: 129-137.

Hammerle, B., Elizalde, C., and Tejedor, F. J. (2008). The spatio-temporal and subcellular expression of the candidate Down syndrome gene Mnb/Dyrk1A in the developing mouse brain suggests distinct sequential roles in neuronal development. *Eur J Neurosci,* **27**: 1061-1074.

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.

Hancock, J. M., and Simon, M. (2005). Simple sequence repeats in proteins and their significance for network evolution. *Gene,* **345**: 113-118.

Handwerger, K. E., and Gall, J. G. (2006). Subnuclear organelles: new insights into form and function. *Trends Cell Biol,* **16**: 19-26.

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C.*, et al.* (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res,* **32**: D258-261.

He, X., and Zhang, J. (2006). Why do hubs tend to be essential in protein networks? *PLoS Genet,* **2**: e88.

Hecht, J. T., Nelson, L. D., Crowder, E., Wang, Y., Elder, F. F., Harrison, W. R., Francomano, C. A., Prange, C. K., Lennon, G. G., Deere, M., and et al. (1995). Mutations in exon 17B of cartilage oligomeric matrix protein (COMP) cause pseudoachondroplasia. *Nat Genet,* **10**: 325-329.

Hedley, M. L., Amrein, H., and Maniatis, T. (1995). An amino acid sequence motif sufficient for subnuclear localization of an arginine/serine-rich splicing factor. *Proc Natl Acad Sci U S A,* **92**: 11524-11528.

Herrmann, C. H., and Mancini, M. A. (2001). The Cdk9 and cyclin T subunits of TAK/P-TEFb localize to splicing factor-rich nuclear speckle regions. *J Cell Sci,* **114**: 1491-1503.

Heun, P. (2007). SUMOrganization of the nucleus. *Curr Opin Cell Biol,* **19**: 350-355.

Himpel, S., Panzer, P., Eirmbter, K., Czajkowska, H., Sayed, M., Packman, L. C., Blundell, T., Kentrup, H., Grotzinger, J., Joost, H. G., and Becker, W. (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.

Hofmann, T. G., Moller, A., Sirma, H., Zentgraf, H., Taya, Y., Droge, W., Will, H., and Schmitz, M. L. (2002). Regulation of p53 activity by its interaction with homeodomain-interacting protein kinase-2. *Nat Cell Biol,* **4**: 1-10.

Hoque, M., Young, T. M., Lee, C. G., Serrero, G., Mathews, M. B., and Pe'ery, T. (2003). The growth factor granulin interacts with cyclin T1 and modulates P-TEFb-dependent transcription. *Mol Cell Biol,* **23**: 1688-1702.

Huang, Y., Chen-Hwang, M. C., Dolios, G., Murakami, N., Padovan, J. C., Wang, R., and Hwang, Y. W. (2004). Mnb/Dyrk1A phosphorylation regulates the interaction of dynamin 1 with SH3 domain-containing proteins. *Biochemistry,* **43**: 10173-10185.

Huang, Y., Chew, T. G., Ge, W., and Balasubramanian, M. K. (2007). Polarity determinants Tea1p, Tea4p, and Pom1p inhibit division-septum assembly at cell ends in fission yeast. *Dev Cell,* **12**: 987-996.

Huntley, M. A., and Golding, G. B. (2002). Simple sequences are rare in the Protein Data Bank. *Proteins,* **48**: 134-140.

Huntley, M. A., Mahmood, S., and Golding, G. B. (2005). Simple sequence in brain and nervous system specific proteins. *Genome,* **48**: 291-301.

Ishitani, T., Ninomiya-Tsuji, J., Nagai, S., Nishita, M., Meneghini, M., Barker, N., Waterman, M., Bowerman, B., Clevers, H., Shibuya, H., and Matsumoto, K. (1999). The TAK1-NLK-MAPK-related pathway antagonizes signalling between beta-catenin and transcription factor TCF. *Nature,* **399**: 798-802.

Isono, K., Nemoto, K., Li, Y., Takada, Y., Suzuki, R., Katsuki, M., Nakagawara, A., and Koseki, H. (2006). Overlapping roles for homeodomain-interacting protein kinases hipk1 and hipk2 in the mediation of cell growth in response to morphogenetic and genotoxic signals. *Mol Cell Biol,* **26**: 2758-2771.

Jagiello, I., Van Eynde, A., Vulsteke, V., Beullens, M., Boudrez, A., Keppens, S., Stalmans, W., and Bollen, M. (2000). Nuclear and subnuclear targeting sequences of the protein phosphatase-1 regulator NIPP1. *J Cell Sci,* **113 Pt 21**: 3761-3768.

Jang, M. K., Mochizuki, K., Zhou, M., Jeong, H. S., Brady, J. N., and Ozato, K. (2005). The bromodomain protein Brd4 is a positive regulatory component of P-TEFb and stimulates RNA polymerase II-dependent transcription. *Mol Cell,* **19**: 523-534.

Janicki, S. M., Tsukamoto, T., Salghetti, S. E., Tansey, W. P., Sachidanandam, R., Prasanth, K. V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R. H., and Spector, D. L. (2004). From silencing to gene expression: real-time analysis in single cells. *Cell,* **116**: 683-698.

Jin, K., Park, S., Ewton, D. Z., and Friedman, E. (2007). The survival kinase Mirk/Dyrk1B is a downstream effector of oncogenic K-ras in pancreatic cancer. *Cancer Res,* **67**: 7247-7255.

Jones, J. C., Phatnani, H. P., Haystead, T. A., MacDonald, J. A., Alam, S. M., and Greenleaf, A. L. (2004). C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats. *J Biol Chem,* **279**: 24957-24964.

Kagey, M. H., Melhuish, T. A., Powers, S. E., and Wotton, D. (2005). Multiple activities contribute to Pc2 E3 function. *Embo J,* **24**: 108-119.

Kajava, A. V. (2001). Review: proteins with repeated sequence--structural prediction and modeling. *J Struct Biol,* **134**: 132-144.

Kalita, M. K., Ramasamy, G., Duraisamy, S., Chauhan, V. S., and Gupta, D. (2006). ProtRepeatsDB: a database of amino acid repeats in genomes. *BMC Bioinformatics,* **7**: 336.

Kanei-Ishii, C., Ninomiya-Tsuji, J., Tanikawa, J., Nomura, T., Ishitani, T., Kishida, S., Kokura, K., Kurahashi, T., Ichikawa-Iwata, E., Kim, Y.*, et al.* (2004). Wnt-1 signal induces phosphorylation and degradation of c-Myb protein via TAK1, HIPK2, and NLK. *Genes Dev,* **18**: 816-829.

Karlin, S. (1995). Statistical significance of sequence patterns in proteins. *Curr Opin Struct Biol,* **5**: 360-371.

Karlin, S., Brocchieri, L., Bergman, A., Mrazek, J., and Gentles, A. J. (2002). Amino acid runs in eukaryotic proteomes and disease associations. *Proc Natl Acad Sci U S A,* **99**: 333-338.

Karlin, S., and Burge, C. (1996). Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. *Proc Natl Acad Sci U S A,* **93**: 1560-1565.

Karlin, S., Mrazek, J., and Gentles, A. J. (2003). Genome comparisons and analysis. *Curr Opin Struct Biol,* **13**: 344-352.

Kashi, Y., and King, D. G. (2006). Simple sequence repeats as advantageous mutators in evolution. *Trends Genet,* **22**: 253-259.

Kelly, P. A., and Rahmani, Z. (2005). DYRK1A enhances the mitogen-activated protein kinase cascade in PC12 cells by forming a complex with Ras, B-Raf, and MEK1. *Mol Biol Cell,* **16**: 3562-3573.

Kentrup, H., Becker, W., Heukelbach, J., Wilmes, A., Schurmann, A., Huppertz, C., Kainulainen, H., and Joost, H. G. (1996). Dyrk, a dual specificity protein kinase with unique structural features whose activity is dependent on tyrosine residues between subdomains VII and VIII. *J Biol Chem,* **271**: 3488-3495.

Kim, E. A., Noh, Y. T., Ryu, M. J., Kim, H. T., Lee, S. E., Kim, C. H., Lee, C., Kim, Y. H., and Choi, C. Y. (2006a). Phosphorylation and transactivation of Pax6 by homeodomain-interacting protein kinase 2. *J Biol Chem,* **281**: 7489-7497.

Kim, E. J., Sung, J. Y., Lee, H. J., Rhim, H., Hasegawa, M., Iwatsubo, T., Min do, S., Kim, J., Paik, S. R., and Chung, K. C. (2006b). Dyrk1A phosphorylates alpha-synuclein and enhances intracellular inclusion formation. *J Biol Chem,* **281**: 33250-33257.

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. (2006c). 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, Y. H., Choi, C. Y., Lee, S. J., Conti, M. A., and Kim, Y. (1998). Homeodomain-interacting protein kinases, a novel family of co-repressors for homeodomain transcription factors. *J Biol Chem,* **273**: 25875-25879.

Kinstrie, R., Lochhead, P. A., Sibbet, G., Morrice, N., and Cleghon, V. (2006). dDYRK2 and Minibrain interact with the chromatin remodelling factors SNR1 and TRX. *Biochem J*.

Kizawa, H., Kou, I., Iida, A., Sudo, A., Miyamoto, Y., Fukuda, A., Mabuchi, A., Kotani, A., Kawakami, A., Yamamoto, S*., et al.* (2005). An aspartic acid repeat polymorphism in asporin inhibits chondrogenesis and increases susceptibility to osteoarthritis. *Nat Genet,* **37**: 138-144.

Ko, T. K., Kelly, E., and Pines, J. (2001). CrkRS: a novel conserved Cdc2-related protein kinase that colocalises with SC35 speckles. *J Cell Sci,* **114**: 2591-2603.

Kondo, S., Lu, Y., Debbas, M., Lin, A. W., Sarosi, I., Itie, A., Wakeham, A., Tuan, J., Saris, C., Elliott, G*., et al.* (2003). Characterization of cells and gene-targeted mice deficient for the p53-binding kinase homeodomain-interacting protein kinase 1 (HIPK1). *Proc Natl Acad Sci U S A,* **100**: 5431-5436.

Kopp, K., and Huang, S. (2005). Perinucleolar compartment and transformation. *J Cell Biochem,* **95**: 217-225.

Kruhlak, M. J., Lever, M. A., Fischle, W., Verdin, E., Bazett-Jones, D. P., and Hendzel, M. J. (2000). Reduced mobility of the alternate splicing factor (ASF) through the nucleoplasm and steady state speckle compartments. *J Cell Biol,* **150**: 41-51.

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.

La Carbona, S., and Le Goff, X. (2006). Spatial regulation of cytokinesis by the Kin1 and Pom1 kinases in fission yeast. *Curr Genet,* **50**: 377-391.

Lafarga, M., Berciano, M. T., Pena, E., Mayo, I., Castano, J. G., Bohmann, D., Rodrigues, J. P., Tavanez, J. P., and Carmo-Fonseca, M. (2002). Clastosome: a subtype of nuclear body enriched in 19S and 20S proteasomes, ubiquitin, and protein substrates of proteasome. *Mol Biol Cell,* **13**: 2771-2782.

Laguna, A. (2008). Study of DYRK1A kinase in central nervous system development: implication in mouse retina development. Universitat Pompeu Fabra, Barcelona.

Laity, J. H., Lee, B. M., and Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol,* **11**: 39-46.

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

Lanz, R. B., Wieland, S., Hug, M., and Rusconi, S. (1995). A transcriptional repressor obtained by alternative translation of a trinucleotide repeat. *Nucleic Acids Res,* **23**: 138-145.

Lavoie, H., Debeane, F., Trinh, Q. D., Turcotte, J. F., Corbeil-Girard, L. P., Dicaire, M. J., Saint-Denis, A., Page, M., Rouleau, G. A., and Brais, B. (2003). Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. *Hum Mol Genet,* **12**: 2967-2979.

Leder, S., Weber, Y., Altafaj, X., Estivill, X., Joost, H. G., and Becker, W. (1999). Cloning and characterization of DYRK1B, a novel member of the DYRK family of protein kinases. *Biochem Biophys Res Commun,* **254**: 474-479.

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.

Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996). The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics,* **33**: 151-152.

Leypoldt, F., Lewerenz, J., and Methner, A. (2001). Identification of genes up-regulated by retinoic-acid-induced differentiation of the human neuronal precursor cell line NTERA-2 cl.D1. *J Neurochem,* **76**: 806-814.

Li, C., Hao, M., Cao, Z., Ding, W., Graves-Deal, R., Hu, J., Piston, D. W., and Coffey, R. J. (2007). Naked2 acts as a cargo recognition and targeting protein to ensure proper delivery and fusion of TGF-alpha containing exocytic vesicles at the lower lateral membrane of polarized MDCK cells. *Mol Biol Cell,* **18**: 3081-3093.

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.

Li, X., Zhang, R., Luo, D., Park, S. J., Wang, Q., Kim, Y., and Min, W. (2005). Tumor necrosis factor alpha-induced desumoylation and cytoplasmic translocation of homeodomain-interacting protein kinase 1 are critical for apoptosis signal-regulating kinase 1-JNK/p38 activation. *J Biol Chem,* **280**: 15061-15070.

Liu, F., Li, B., Tung, E. J., Grundke-Iqbal, I., Iqbal, K., and Gong, C. X. (2007). Site-specific effects of tau phosphorylation on its microtubule assembly activity and self-aggregation. *Eur J Neurosci,* **26**: 3429-3436.

Liu, Q., and Dreyfuss, G. (1996). A novel nuclear structure containing the survival of motor neurons protein. *Embo J,* **15**: 3555-3565.

Liu, W., Khare, S. L., Liang, X., Peters, M. A., Liu, X., Cepko, C. L., and Xiang, M. (2000). All Brn3 genes can promote retinal ganglion cell differentiation in the chick. *Development,* **127**: 3237-3247.

Lochhead, P. A., Sibbet, G., Kinstrie, R., Cleghon, T., Rylatt, M., Morrison, D. K., and Cleghon, V. (2003). dDYRK2: a novel dual-specificity tyrosine-phosphorylation-regulated kinase in Drosophila. *Biochem J,* **374**: 381-391.

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.

MacKeigan, J. P., Murphy, L. O., and Blenis, J. (2005). Sensitized RNAi screen of human kinases and phosphatases identifies new regulators of apoptosis and chemoresistance. *Nat Cell Biol,* **7**: 591-600.

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.

Malatesta, M., Fakan, S., and Fischer, U. (1999). The Sm core domain mediates targeting of U1 snRNP to subnuclear compartments involved in transcription and splicing. *Exp Cell Res,* **249**: 189-198.

Manfiolli, A. O., Maragno, A. L., Baqui, M. M., Yokoo, S., Teixeira, F. R., Oliveira, E. B., and Gomes, M. D. (2008). FBXO25-associated Nuclear Domains: A Novel Subnuclear Structure. *Mol Biol Cell,* **19**: 1848-1861.

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.

Marquardt, T., and Gruss, P. (2002). Generating neuronal diversity in the retina: one for nearly all. *Trends Neurosci,* **25**: 32-38.

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.

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.

Martin, D. E., Soulard, A., and Hall, M. N. (2004). TOR regulates ribosomal protein gene expression via PKA and the Forkhead transcription factor FHL1. *Cell,* **119**: 969-979.

Martin, S. E., Mu, X., and Klein, W. H. (2005). Identification of an N-terminal transcriptional activation domain within Brn3b/POU4f2. *Differentiation,* **73**: 18-27.

Martindill, D. M., Risebro, C. A., Smart, N., Franco-Viseras Mdel, M., Rosario, C. O., Swallow, C. J., Dennis, J. W., and Riley, P. R. (2007). Nucleolar release of Hand1 acts as a molecular switch to determine cell fate. *Nat Cell Biol,* **9**: 1131-1141.

Matera, A. G. (1999). Nuclear bodies: multifaceted subdomains of the interchromatin space. *Trends Cell Biol,* **9**: 302-309.

Matsumoto, N., Ohashi, H., Tsukahara, M., Kim, K. C., Soeda, E., and Niikawa, N. (1997). Possible narrowed assignment of the loci of monosomy 21-associated microcephaly and intrauterine growth retardation to a 1.2-Mb segment at 21q22.2. *Am J Hum Genet,* **60**: 997-999.

Matsuo, R., Ochiai, W., Nakashima, K., and Taga, T. (2001). A new expression cloning strategy for isolation of substrate-specific kinases by using phosphorylation site-specific antibody. *J Immunol Methods,* **247**: 141-151.

Medlin, J., Scurry, A., Taylor, A., Zhang, F., Peterlin, B. M., and Murphy, S. (2005). P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *Embo J,* **24**: 4154-4165.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. *Genes Dev,* **19**: 1401-1415.

Menon, R. P., and Pastore, A. (2006). Expansion of amino acid homo-sequences in proteins: insights into the role of amino acid homo-polymers and of the protein context in aggregation. *Cell Mol Life Sci,* **63**: 1677-1685.

Mercer, S. E., Ewton, D. Z., Deng, X., Lim, S., Mazur, T. R., and Friedman, E. (2005). Mirk/Dyrk1B mediates survival during the differentiation of C2C12 myoblasts. *J Biol Chem,* **280**: 25788-25801.

Mercer, S. E., Ewton, D. Z., Shah, S., Naqvi, A., and Friedman, E. (2006). Mirk/Dyrk1b mediates cell survival in rhabdomyosarcomas. *Cancer Res,* **66**: 5143-5150.

Mercer, S. E., and Friedman, E. (2006). Mirk/Dyrk1B: a multifunctional dual-specificity kinase involved in growth arrest, differentiation, and cell survival. *Cell Biochem Biophys,* **45**: 303-315.

Mermoud, J. E., Cohen, P. T., and Lamond, A. I. (1994). Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *Embo J,* **13**: 5679-5688.

Miller, C. T., Aggarwal, S., Lin, T. K., Dagenais, S. L., Contreras, J. I., Orringer, M. B., Glover, T. W., Beer, D. G., and Lin, L. (2003). Amplification and overexpression of the dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2 (DYRK2) gene in esophageal and lung adenocarcinomas. *Cancer Res,* **63**: 4136-4143.

Mintz, P. J., Patterson, S. D., Neuwald, A. F., Spahr, C. S., and Spector, D. L. (1999). Purification and biochemical characterization of interchromatin granule clusters. *Embo J,* **18**: 4308-4320.

Misteli, T. (2000). Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function. *J Cell Sci,* **113 (Pt 11)**: 1841-1849.

Misteli, T. (2001a). Protein dynamics: implications for nuclear architecture and gene expression. *Science,* **291**: 843-847.

Misteli, T. (2001b). The concept of self-organization in cellular architecture. *J Cell Biol,* **155**: 181-185.

Misteli, T., Caceres, J. F., Clement, J. Q., Krainer, A. R., Wilkinson, M. F., and Spector, D. L. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J Cell Biol,* **143**: 297-307.

Misteli, T., Caceres, J. F., and Spector, D. L. (1997). The dynamics of a pre-mRNA splicing factor in living cells. *Nature,* **387**: 523-527.

Misteli, T., and Spector, D. L. (1996). Serine/threonine phosphatase 1 modulates the subnuclear distribution of pre-mRNA splicing factors. *Mol Biol Cell,* **7**: 1559-1572.

Misteli, T., and Spector, D. L. (1997). Protein phosphorylation and the nuclear organization of pre-mRNA splicing. *Trends Cell Biol,* **7**: 135-138.

Misteli, T., and Spector, D. L. (1999). RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell,* **3**: 697-705.

Miyata, Y., and Nishida, E. (1999). Distantly related cousins of MAP kinase: biochemical properties and possible physiological functions. *Biochem Biophys Res Commun,* **266**: 291-295.

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.

Moolenaar, W. H. (1986). Effects of growth factors on intracellular pH regulation. *Annu Rev Physiol,* **48**: 363-376.

Moriya, H., Shimizu-Yoshida, Y., Omori, A., Iwashita, S., Katoh, M., and Sakai, A. (2001). Yak1p, a DYRK family kinase, translocates to the nucleus and phosphorylates yeast Pop2p in response to a glucose signal. *Genes Dev,* **15**: 1217-1228.

Mortier, E., Wuytens, G., Leenaerts, I., Hannes, F., Heung, M. Y., Degeest, G., David, G., and Zimmermann, P. (2005). Nuclear speckles and nucleoli targeting by PIP2-PDZ domain interactions. *Embo J,* **24**: 2556-2565.

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.

Morton, S., Davis, R. J., McLaren, A., and Cohen, P. (2003). A reinvestigation of the multisite phosphorylation of the transcription factor c-Jun. *Embo J,* **22**: 3876-3886.

Mu, X., Beremand, P. D., Zhao, S., Pershad, R., Sun, H., Scarpa, A., Liang, S., Thomas, T. L., and Klein, W. H. (2004). Discrete gene sets depend on POU domain transcription factor Brn3b/Brn-3.2/POU4f2 for their expression in the mouse embryonic retina. *Development,* **131**: 1197-1210.

Mu, X., Fu, X., Sun, H., Beremand, P. D., Thomas, T. L., and Klein, W. H. (2005). A gene network downstream of transcription factor Math5 regulates retinal progenitor cell competence and ganglion cell fate. *Dev Biol,* **280**: 467-481.

Mu, X., and Klein, W. H. (2004). A gene regulatory hierarchy for retinal ganglion cell specification and differentiation. *Semin Cell Dev Biol,* **15**: 115-123.

Mularoni, L., Guigo, R., and Alba, M. M. (2006). Mutation patterns of amino acid tandem repeats in the human proteome. *Genome Biol,* **7**: R33.

Mularoni, L., Veitia, R. A., and Alba, M. M. (2007). Highly constrained proteins contain an unexpectedly large number of amino acid tandem repeats. *Genomics,* **89**: 316-325.

Murakami, N., Xie, W., Lu, R. C., Chen-Hwang, M. C., Wieraszko, A., and Hwang, Y. W. (2006). Phosphorylation of amphiphysin I by minibrain kinase/dual-specificity tyrosine phosphorylation-regulated kinase, a kinase implicated in Down syndrome. *J Biol Chem,* **281**: 23712-23724.

Ngo, J. C., Chakrabarti, S., Ding, J. H., Velazquez-Dones, A., Nolen, B., Aubol, B. E., Adams, J. A., Fu, X. D., and Ghosh, G. (2005). Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Mol Cell,* **20**: 77-89.

Nishi, Y., and Lin, R. (2005). DYRK2 and GSK-3 phosphorylate and promote the timely degradation of OMA-1, a key regulator of the oocyte-to-embryo transition in C. elegans. *Dev Biol,* **288**: 139-149.

Nishi, Y., Rogers, E., Robertson, S. M., and Lin, R. (2008). Polo kinases regulate C. elegans embryonic polarity via binding to DYRK2-primed MEX-5 and MEX-6. *Development,* **135**: 687-697.

Nousiainen, M., Sillje, H. H., Sauer, G., Nigg, E. A., and Korner, R. (2006). Phosphoproteome analysis of the human mitotic spindle. *Proc Natl Acad Sci U S A,* **103**: 5391-5396.

Nunez, E., Kwon, Y. S., Hutt, K. R., Hu, Q., Cardamone, M. D., Ohgi, K. A., Garcia-Bassets, I., Rose, D. W., Glass, C. K., Rosenfeld, M. G., and Fu, X. D. (2008). Nuclear receptor-enhanced transcription requires motor- and LSD1-dependent gene networking in interchromatin granules. *Cell,* **132**: 996-1010.

Okui, M., Ide, T., Morita, K., Funakoshi, E., Ito, F., Ogita, K., Yoneda, Y., Kudoh, J., and Shimizu, N. (1999). High-level expression of the Mnb/Dyrk1A gene in brain and heart during rat early development. *Genomics,* **62**: 165-171.

Olsen, J. V., Blagoev, B., Gnad, F., Macek, B., Kumar, C., Mortensen, P., and Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell,* **127**: 635-648.

Oma, Y., Kino, Y., Sasagawa, N., and Ishiura, S. (2004). Intracellular localization of homopolymeric amino acid-containing proteins expressed in mammalian cells. *J Biol Chem,* **279**: 21217-21222.

Oma, Y., Kino, Y., Sasagawa, N., and Ishiura, S. (2005). Comparative analysis of the cytotoxicity of homopolymeric amino acids. *Biochim Biophys Acta,* **1748**: 174-179.

Palancade, B., and Bensaude, O. (2003). Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem,* **270**: 3859-3870.

Pan, L., Yang, Z., Feng, L., and Gan, L. (2005). Functional equivalence of Brn3 POU-domain transcription factors in mouse retinal neurogenesis. *Development,* **132**: 703-712.

Pang, K. M., Ishidate, T., Nakamura, K., Shirayama, M., Trzepacz, C., Schubert, C. M., Priess, J. R., and Mello, C. C. (2004). The minibrain kinase homolog, mbk-2, is required for spindle positioning and asymmetric cell division in early C. elegans embryos. *Dev Biol,* **265**: 127-139.

Paraguison, R. C., Higaki, K., Sakamoto, Y., Hashimoto, O., Miyake, N., Matsumoto, H., Yamamoto, K., Sasaki, T., Kato, N., and Nanba, E. (2005). Polyhistidine tract expansions in HOXA1 result in intranuclear aggregation and increased cell death. *Biochem Biophys Res Commun,* **336**: 1033-1039.

Paraguison, R. C., Higaki, K., Yamamoto, K., Matsumoto, H., Sasaki, T., Kato, N., and Nanba, E. (2007). Enhanced autophagic cell death in expanded polyhistidine variants of HOXA1 reduces PBX1-coupled transcriptional activity and inhibits neuronal differentiation. *J Neurosci Res,* **85**: 479-487.

Pearson, C. E., Nichol Edamura, K., and Cleary, J. D. (2005). Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet,* **6**: 729-742.

Pellettieri, J., Reinke, V., Kim, S. K., and Seydoux, G. (2003). Coordinate activation of maternal protein degradation during the egg-to-embryo transition in C. elegans. *Dev Cell,* **5**: 451-462.

Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994). Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc Natl Acad Sci U S A,* **91**: 5355-5358.

Phair, R. D., and Misteli, T. (2000). High mobility of proteins in the mammalian cell nucleus. *Nature,* **404**: 604-609.

Philips, A. S., Kwok, J. C., and Chong, B. H. (2007). Analysis of the signals and mechanisms mediating nuclear trafficking of GATA-4. Loss of DNA binding is associated with localization in intranuclear speckles. *J Biol Chem,* **282**: 24915-24927.

Plaza, S., Hennemann, H., Moroy, T., Saule, S., and Dozier, C. (1999). Evidence that POU factor Brn-3B regulates expression of Pax-6 in neuroretina cells. *J Neurobiol,* **41**: 349-358.

Prasanth, K. V., Rajendra, T. K., Lal, A. K., and Lakhotia, S. C. (2000). Omega speckles - a novel class of nuclear speckles containing hnRNPs associated with noncoding hsr-omega RNA in Drosophila. *J Cell Sci,* **113 Pt 19**: 3485-3497.

Prasanth, K. V., Sacco-Bubulya, P. A., Prasanth, S. G., and Spector, D. L. (2003). Sequential entry of components of the gene expression machinery into daughter nuclei. *Mol Biol Cell,* **14**: 1043-1057.

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.

Qiu, F., Jiang, H., and Xiang, M. (2008). A comprehensive negative regulatory program controlled by Brn3b to ensure ganglion cell specification from multipotential retinal precursors. *J Neurosci,* **28**: 3392-3403.

Quintin, S., Mains, P. E., Zinke, A., and Hyman, A. A. (2003). The mbk-2 kinase is required for inactivation of MEI-1/katanin in the one-cell Caenorhabditis elegans embryo. *EMBO Rep,* **4**: 1175-1181.

Rachidi, M., Lopes, C., Vayssettes, C., Smith, D. J., Rubin, E. M., and Delabar, J. M. (2007). New cerebellar phenotypes in YAC transgenic mouse in vivo library of human Down syndrome critical region-1. *Biochem Biophys Res Commun,* **364**: 488-494.

Rahmani, Z., Lopes, C., Rachidi, M., and Delabar, J. M. (1998). Expression of the mnb (dyrk) protein in adult and embryonic mouse tissues. *Biochem Biophys Res Commun,* **253**: 514-518.

Raich, W. B., Moorman, C., Lacefield, C. O., Lehrer, J., Bartsch, D., Plasterk, R. H., Kandel, E. R., and Hobert, O. (2003). Characterization of Caenorhabditis elegans homologs of the Down syndrome candidate gene DYRK1A. *Genetics,* **163**: 571-580.

Raman, M., Chen, W., and Cobb, M. H. (2007). Differential regulation and properties of MAPKs. *Oncogene,* **26**: 3100-3112.

Ramanathan, Y., Rajpara, S. M., Reza, S. M., Lees, E., Shuman, S., Mathews, M. B., and Pe'ery, T. (2001). Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. *J Biol Chem,* **276**: 10913-10920.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol,* **17**: 1030-1032.

Rinaldo, C., Prodosmo, A., Siepi, F., and Soddu, S. (2007). HIPK2: a multitalented partner for transcription factors in DNA damage response and development. *Biochem Cell Biol,* **85**: 411-418.

Rochat-Steiner, V., Becker, K., Micheau, O., Schneider, P., Burns, K., and Tschopp, J. (2000). FIST/HIPK3: a Fas/FADD-interacting serine/threonine kinase that induces FADD phosphorylation and inhibits fas-mediated Jun NH(2)-terminal kinase activation. *J Exp Med,* **192**: 1165-1174.

Rolfsmeier, M. L., and Lahue, R. S. (2000). Stabilizing effects of interruptions on trinucleotide repeat expansions in Saccharomyces cerevisiae. *Mol Cell Biol,* **20**: 173-180.

Roscic, A., Moller, A., Calzado, M. A., Renner, F., Wimmer, V. C., Gresko, E., Ludi, K. S., and Schmitz, M. L. (2006). Phosphorylation-dependent control of Pc2 SUMO E3 ligase activity by its substrate protein HIPK2. *Mol Cell,* **24**: 77-89.

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., and Song, W. J. (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.

Ryoo, S. R., Jeong, H. K., Radnaabazar, C., Yoo, J. J., Cho, H. J., Lee, H. W., Kim, I. S., Cheon, Y. H., Ahn, Y. S., Chung, S. H., and Song, W. J. (2007). DYRK1A-mediated Hyperphosphorylation of Tau: A FUNCTIONAL LINK BETWEEN DOWN SYNDROME AND ALZHEIMER DISEASE. *J Biol Chem,* **282**: 34850-34857.

Sacco-Bubulya, P., and Spector, D. L. (2002). Disassembly of interchromatin granule clusters alters the coordination of transcription and pre-mRNA splicing. *J Cell Biol,* **156**: 425-436.

Sacher, F., Moller, C., Bone, W., Gottwald, U., and Fritsch, M. (2007). The expression of the testis-specific Dyrk4 kinase is highly restricted to step 8 spermatids but is not required for male fertility in mice. *Mol Cell Endocrinol,* **267**: 80-88.

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.

Samady, L., Dennis, J., Budhram-Mahadeo, V., and Latchman, D. S. (2004). Activation of CDK4 gene expression in human breast cancer cells by the Brn-3b POU family transcription factor. *Cancer Biol Ther,* **3**: 317-323.

Sanchez-Alvarez, M., Goldstrohm, A. C., Garcia-Blanco, M. A., and Sune, C. (2006). Human transcription elongation factor CA150 localizes to splicing factor-rich nuclear speckles and assembles transcription and splicing components into complexes through its amino and carboxyl regions. *Mol Cell Biol,* **26**: 4998-5014.

Sargent, L. M., Ensell, M. X., Ostvold, A. C., Baldwin, K. T., Kashon, M. L., Lowry, D. T., Senft, J. R., Jefferson, A. M., Johnson, R. C., Li, Z*., et al.* (2008). Chromosomal changes in high- and low-invasive mouse lung adenocarcinoma cell strains derived from early passage mouse lung adenocarcinoma cell strains. *Toxicol Appl Pharmacol*.

Saurin, A. J., Shiels, C., Williamson, J., Satijn, D. P., Otte, A. P., Sheer, D., and Freemont, P. S. (1998). The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain. *J Cell Biol,* **142**: 887-898.

Schreiber, E., Matthias, P., Muller, M. M., and Schaffner, W. (1989). Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells. *Nucleic Acids Res,* **17**: 6419.

Schul, W., Groenhout, B., Koberna, K., Takagaki, Y., Jenny, A., Manders, E. M., Raska, I., van Driel, R., and de Jong, L. (1996). The RNA 3' cleavage factors CstF 64 kDa and CPSF 100 kDa are concentrated in nuclear domains closely associated with coiled bodies and newly synthesized RNA. *Embo J,* **15**: 2883-2892.

Semple, J. I., Brown, S. E., Sanderson, C. M., and Campbell, R. D. (2002). A distinct bipartite motif is required for the localization of inhibitory kappaB-like (IkappaBL) protein to nuclear speckles. *Biochem J,* **361**: 489-496.

Sertedaki, A., and Lindsay, S. (1996). CAC--the neglected repeat. *Bioessays,* **18**: 237-242.

Shao, J., and Diamond, M. I. (2007). Polyglutamine diseases: emerging concepts in pathogenesis and therapy. *Hum Mol Genet,* **16 Spec No. 2**: R115-123.

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 ASF-regulated alternative splicing of tau in down syndrome. *J Biol Chem*.

Shimohata, T., Nakajima, T., Yamada, M., Uchida, C., Onodera, O., Naruse, S., Kimura, T., Koide, R., Nozaki, K., Sano, Y*., et al.* (2000). Expanded polyglutamine stretches interact with TAFII130, interfering with CREB-dependent transcription. *Nat Genet,* **26**: 29-36.

Shindoh, N., Kudoh, J., Maeda, H., Yamaki, A., Minoshima, S., Shimizu, Y., and Shimizu, N. (1996). Cloning of a human homolog of the Drosophila minibrain/rat Dyrk gene from "the Down syndrome critical region" of chromosome 21. *Biochem Biophys Res Commun,* **225**: 92-99.

Shirayama, M., Soto, M. C., Ishidate, T., Kim, S., Nakamura, K., Bei, Y., van den Heuvel, S., and Mello, C. C. (2006). The Conserved Kinases CDK-1, GSK-3, KIN-19, and MBK-2 Promote OMA-1 Destruction to Regulate the Oocyte-to-Embryo Transition in C. elegans. *Curr Biol,* **16**: 47-55.

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.

Siwach, P., Pophaly, S. D., and Ganesh, S. (2006). Genomic and evolutionary insights into genes encoding proteins with single amino acid repeats. *Mol Biol Evol,* **23**: 1357-1369.

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.

Smith, D. J., Stevens, M. E., Sudanagunta, S. P., Bronson, R. T., Makhinson, M., Watabe, A. M., O'Dell, T. J., Fung, J., Weier, H. U., Cheng, J. F., and Rubin, E. M. (1997). Functional screening of 2 Mb of human chromosome 21q22.2 in transgenic mice implicates minibrain in learning defects associated with Down syndrome. *Nat Genet,* **16**: 28-36.

Sone, M., Hayashi, T., Tarui, H., Agata, K., Takeichi, M., and Nakagawa, S. (2007). The mRNA-like noncoding RNA Gomafu constitutes a novel nuclear domain in a subset of neurons. *J Cell Sci,* **120**: 2498-2506.

Song, W. J., Chung, S. H., and Kurnit, D. M. (1997). The murine Dyrk protein maps to chromosome 16, localizes to the nucleus, and can form multimers. *Biochem Biophys Res Commun,* **231**: 640-644.

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., and Kurnit, D. M. (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.

Songyang, Z., Lu, K. P., Kwon, Y. T., Tsai, L. H., Filhol, O., Cochet, C., Brickey, D. A., Soderling, T. R., Bartleson, C., Graves, D. J.*, et al.* (1996). A structural basis for substrate specificities of protein Ser/Thr kinases: primary sequence preference of casein kinases I and II, NIMA, phosphorylase kinase, calmodulin-dependent kinase II, CDK5, and Erk1. *Mol Cell Biol,* **16**: 6486-6493.

Souza, G. M., Lu, S., and Kuspa, A. (1998). YakA, a protein kinase required for the transition from growth to development in Dictyostelium. *Development,* **125**: 2291-2302.

Spector, D. L. (1993). Macromolecular domains within the cell nucleus. *Annu Rev Cell Biol,* **9**: 265-315.

Spector, D. L. (2001). Nuclear domains. *J Cell Sci,* **114**: 2891-2893.

Spector, D. L., Fu, X. D., and Maniatis, T. (1991). Associations between distinct pre-mRNA splicing components and the cell nucleus. *Embo J,* **10**: 3467-3481.

Steeg, P. S., Palmieri, D., Ouatas, T., and Salerno, M. (2003). Histidine kinases and histidine phosphorylated proteins in mammalian cell biology, signal transduction and cancer. *Cancer Lett,* **190**: 1-12.

Stitzel, M. L., Pellettieri, J., and Seydoux, G. (2006). The C. elegans DYRK Kinase MBK-2 Marks Oocyte Proteins for Degradation in Response to Meiotic Maturation. *Curr Biol,* **16**: 56-62.

Subrahmanyam, Y. V., Yamaga, S., Prashar, Y., Lee, H. H., Hoe, N. P., Kluger, Y., Gerstein, M., Goguen, J. D., Newburger, P. E., and Weissman, S. M. (2001). RNA expression patterns change dramatically in human neutrophils exposed to bacteria. *Blood,* **97**: 2457-2468.

Taira, N., Nihira, K., Yamaguchi, T., Miki, Y., and Yoshida, K. (2007). DYRK2 is targeted to the nucleus and controls p53 via Ser46 phosphorylation in the apoptotic response to DNA damage. *Mol Cell,* **25**: 725-738.

Takahashi-Yanaga, F., Mori, J., Matsuzaki, E., Watanabe, Y., Hirata, M., Miwa, Y., Morimoto, S., and Sasaguri, T. (2006). Involvement of GSK-3beta and DYRK1B in differentiation-inducing factor-3-induced phosphorylation of cyclin D1 in HeLa cells. *J Biol Chem*.

Taminato, A., Bagattini, R., Gorjao, R., Chen, G., Kuspa, A., and Souza, G. M. (2002). Role for YakA, cAMP, and protein kinase A in regulation of stress responses of Dictyostelium discoideum cells. *Mol Biol Cell,* **13**: 2266-2275.

Tatebe, H., Nakano, K., Maximo, R., and Shiozaki, K. (2008). Pom1 DYRK regulates localization of the Rga4 GAP to ensure bipolar activation of Cdc42 in fission yeast. *Curr Biol,* **18**: 322-330.

Taube, R., Lin, X., Irwin, D., Fujinaga, K., and Peterlin, B. M. (2002). Interaction between P-TEFb and the C-terminal domain of RNA polymerase II activates transcriptional elongation from sites upstream or downstream of target genes. *Mol Cell Biol,* **22**: 321-331.

Teague, T. K., Hildeman, D., Kedl, R. M., Mitchell, T., Rees, W., Schaefer, B. C., Bender, J., Kappler, J., and Marrack, P. (1999). Activation changes the spectrum but not the diversity of genes expressed by T cells. *Proc Natl Acad Sci U S A,* **96**: 12691-12696.

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.

Tewari, D., Monk, B. J., Al-Ghazi, M. S., Parker, R., Heck, J. D., Burger, R. A., and Fruehauf, J. P. (2005). Gene expression profiling of in vitro radiation resistance in cervical carcinoma: a feasibility study. *Gynecol Oncol,* **99**: 84-91.

Turner, E. E., Rhee, J. M., and Feldman, L. T. (1997). The POU-domain factor Brn-3.0 recognizes characteristic sites in the herpes simplex virus genome. *Nucleic Acids Res,* **25**: 2589-2594.

van Es, S., Weening, K. E., and Devreotes, P. N. (2001). The protein kinase YakA regulates g-protein-linked signaling responses during growth and development of Dictyostelium. *J Biol Chem,* **276**: 30761-30765.

Varjosalo, M., Bjorklund, M., Cheng, F., Syvanen, H., Kivioja, T., Kilpinen, S., Sun, Z., Kallioniemi, O., Stunnenberg, H. G., He, W. W.*, et al.* (2008). Application of active and kinase-deficient kinome collection for identification of kinases regulating hedgehog signaling. *Cell,* **133**: 537-548.

Veitia, R. A. (2004). Amino acids runs and genomic compositional biases in vertebrates. *Genomics,* **83**: 502-507.

Volker, M., Mone, M. J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J. H., van Driel, R., van Zeeland, A. A., and Mullenders, L. H. (2001). Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell,* **8**: 213-224.

von Groote-Bidlingmaier, F., Schmoll, D., Orth, H. M., Joost, H. G., Becker, W., and Barthel, A. (2003). DYRK1 is a co-activator of FKHR (FOXO1a)-dependent glucose-6-phosphatase gene expression. *Biochem Biophys Res Commun,* **300**: 764-769.

von Mikecz, A. (2006). The nuclear ubiquitin-proteasome system. *J Cell Sci,* **119**: 1977-1984.

Wang, H. Y., Lin, W., Dyck, J. A., Yeakley, J. M., Songyang, Z., Cantley, L. C., and Fu, X. D. (1998a). 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., Kudoh, J., Shintani, A., Minoshima, S., and Shimizu, N. (1998b). Identification of two novel 5' noncoding exons in human MNB/DYRK gene and alternatively spliced transcripts. *Biochem Biophys Res Commun,* **250**: 704-710.

Wang, S. W., Gan, L., Martin, S. E., and Klein, W. H. (2000). Abnormal polarization and axon outgrowth in retinal ganglion cells lacking the POU-domain transcription factor Brn-3b. *Mol Cell Neurosci,* **16**: 141-156.

Wang, S. W., Mu, X., Bowers, W. J., Kim, D. S., Plas, D. J., Crair, M. C., Federoff, H. J., Gan, L., and Klein, W. H. (2002). Brn3b/Brn3c double knockout mice reveal an unsuspected role for Brn3c in retinal ganglion cell axon outgrowth. *Development,* **129**: 467-477.

Wegiel, J., Dowjat, K., Kaczmarski, W., Kuchna, I., Nowicki, K., Frackowiak, J., Mazur Kolecka, B., Wegiel, J., Silverman, W. P., Reisberg, B.*, et al.* (2008). The role of overexpressed DYRK1A protein in the early onset of neurofibrillary degeneration in Down syndrome. *Acta Neuropathol,* **116**: 391-407.

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, X., Somanathan, S., Samarabandu, J., and Berezney, R. (1999). Three-dimensional visualization of transcription sites and their association with splicing factor-rich nuclear speckles. *J Cell Biol,* **146**: 543-558.

Wiggins, A. K., Wei, G., Doxakis, E., Wong, C., Tang, A. A., Zang, K., Luo, E. J., Neve, R. L., Reichardt, L. F., and Huang, E. J. (2004). Interaction of Brn3a and HIPK2 mediates transcriptional repression of sensory neuron survival. *J Cell Biol,* **167**: 257-267.

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.

Xiang, M., Zhou, L., Macke, J. P., Yoshioka, T., Hendry, S. H., Eddy, R. L., Shows, T. B., and Nathans, J. (1995). The Brn-3 family of POU-domain factors: primary structure, binding specificity, and expression in subsets of retinal ganglion cells and somatosensory neurons. *J Neurosci,* **15**: 4762-4785.

Yamanaka, R., Arao, T., Yajima, N., Tsuchiya, N., Homma, J., Tanaka, R., Sano, M., Oide, A., Sekijima, M., and Nishio, K. (2006). Identification of expressed genes characterizing long-term survival in malignant glioma patients. *Oncogene,* **25**: 5994-6002.

Yang, E. J., Ahn, Y. S., and Chung, K. C. (2001). Protein kinase Dyrk1 activates cAMP response element-binding protein during neuronal differentiation in hippocampal progenitor cells. *J Biol Chem,* **276**: 39819-39824.

Yoshida, K. (2008). Role for DYRK family kinases on regulation of apoptosis. *Biochem Pharmacol*.

Young, T. M., Wang, Q., Pe'ery, T., and Mathews, M. B. (2003). The human I-mfa domain-containing protein, HIC, interacts with cyclin T1 and modulates P-TEFb-dependent transcription. *Mol Cell Biol,* **23**: 6373-6384.

Yuan, Y., Li, D. M., and Sun, H. (1998). PIR1, a novel phosphatase that exhibits high affinity to RNA . ribonucleoprotein complexes. *J Biol Chem,* **273**: 20347-20353.

Zeng, C., Kim, E., Warren, S. L., and Berget, S. M. (1997). Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *Embo J,* **16**: 1401-1412.

Zhang, J., Pho, V., Bonasera, S. J., Holtzman, J., Tang, A. T., Hellmuth, J., Tang, S., Janak, P. H., Tecott, L. H., and Huang, E. J. (2007). Essential function of HIPK2 in TGFbeta-dependent survival of midbrain dopamine neurons. *Nat Neurosci,* **10**: 77-86.

Zhang, X. M., and Yang, X. J. (2001). Regulation of retinal ganglion cell production by Sonic hedgehog. *Development,* **128**: 943-957.

Zhong, S., Salomoni, P., and Pandolfi, P. P. (2000). The transcriptional role of PML and the nuclear body. *Nat Cell Biol,* **2**: E85-90.

Zou, Y., Ewton, D. Z., Deng, X., Mercer, S. E., and Friedman, E. (2004). Mirk/dyrk1B kinase destabilizes cyclin D1 by phosphorylation at threonine 288. *J Biol Chem,* **279**: 27790-27798.