SCUOLA NORMALE SUPERIORE DI PISA

PH. D. THESIS

Molecular Biology

Dynamics and interactions of an oncogenic homeotic protein within human replicative complexes

Laura Marchetti

ADVISOR
Prof. Arturo Falaschi

2010

to my family,
to Andrea

TABLE OF CONTENTS

List of Figures and Tables v

List of Abbreviations ix

List of Publications xi

ABSTRACT 1

1. INTRODUCTION 3

1.1 Regulatory mechanisms of eukaryotic DNA replication 3

1.1.1 Eukaryotic DNA replication: an overview 3

1.1.2 Multiple levels of DNA replication regulation within cell-cycle progression 5

1.1.3 From initiation to elongation in eukaryotic DNA replication 7

1.1.4 Recurrence of the ring-shape geometry in RC protein structures 16

1.1.5 The riddle of origin specification in metazoa 19

1.2 The family of HOX proteins 22

1.2.1 *Hox* genes encode HOX proteins regulating development 23

1.2.2 The homeodomain and other homeobox signatures 27

1.2.3 The “HOX paradox”: transcriptional and functional specificity 29

1.2.3.1 HOX cofactors and collaborators *30*

1.2.3.2 Role of DNA shape in HOX specific DNA recognition *32*

1.2.4 HOX proteins and cancer 33

1.2.5 HOX proteins vs. DNA replication and cell-cycle regulation 34

1.3 Biophysical approaches for the *in vivo* study of nuclear structure and function 36

1.3.1 FRAP and FRET applied to the study of the nucleus 38

1.3.2 Study of DNA replication regulation in living cells 40

1.4 Aim of the work 42

2. RESULTS 43

2.1 Expression and localization of endogenous HOXC13 in human cells 43

2.2 Expression of GFP fusions of HOXC13 protein in human cells 46

2.3 Co-localization of fluorolabelled HOXC13 with replication foci 49

2.3.1 Pulsed BrdUrd immunofluorescence 50

2.3.2 Replication factories visualized in real-time in living cells 53

2.3.3 Role of the homeodomain in the targeting of early replicating chromatin 55

2.4 HOXC13 nuclear dynamics 57

2.4.1 Half-FRAP experiments using the EGFP-HP1α protein as a standard 58

2.4.2 Choice of a fluorophore suitable for the half-FRAP procedure 59

2.4.3 Half-FRAP analysis of the nuclear dynamics of wt mCherry-HOXC13 60

2.4.4 Quantitative analysis of wt mCherry-HOXC13 nuclear dynamics 65
2.4.5 Concluding remarks about HOXC13 nuclear dynamics 67
2.5 Cell-cycle related changes of sub-cellular localization of RC proteins in living cells 68
2.5.1 ORC2 69
2.5.2 Cdc6 72
2.5.3 ORC1 73
2.5.4 MCM3 74
2.5.5 Concluding remarks about RC proteins vs. HOXC13 75
2.6 *In vivo* detection of HOXC13 interaction with RC proteins 79
2.6.1 FLIM detection of HOXC13-RC interactions in living cells 80
2.6.2 Spatial definition of HOXC13-RC interactions in the nucleus 82
2.6.3 Temporal definition of HOXC13-RC interactions along cell-cycle progression 84
2.6.5 Biochemical test of HOXC13-RC proteins affinity 85

3. DISCUSSION 87
3.1 HOXC13 is a stable chromatin component 87
3.2 HOXC13 is present at early S replication foci due to its homeodomain 88
3.3 HOXC13 contributes to origin specification in human DNA replication 90

4. MATERIALS AND METHODS 95
4.1 Expression constructs 95
4.1.1 HOXC13 protein constructs 95
4.1.2 Control constructs 96
4.1.3 RC proteins constructs 96
4.1.4 Other nuclear proteins constructs 96
4.2 Cell culture, transient and stable transfection, synchronization 96
4.3 Antibodies 97
4.4 Immunofluorescence 97
4.5 Biochemical cell extraction, fractionation and co-immunoprecipitation 98
4.6 Co-localization analysis 99
4.6.1 Image acquisition 99
4.6.2 Image data analysis 100
4.7 FRAP 101
4.7.1 Image acquisition 101
4.7.2 Image data analysis 101
4.8 FLIM 102
4.8.1 Image acquisition 102
4.8.2 Image data analysis 102
4.9 Time lapse imaging 103
4.9.1 Image acquisition 103
4.9.2 Time-lapse imaging presentation 104
4.10 Image presentation 104

APPENDIX A 105
  A.1 Available models to analyze the nuclear dynamics of chromatin-binding proteins 105
  A.2 Equations of the diffusion-reaction model 106
    A.2.1 Half-nuclear geometry 107
    A.2.2 2nd/5th nuclear slices geometry 110
    A.2.3 Selection criteria of FRAP experiments for the fitting with the diffusion-reaction model 110
APPENDIX B 113
  B.1 Definition of FRET 113
  B.2 FRET detection by FLIM: the ideally suited E^{0}GFP-mCherry FRET pair 115
  B.3 Distance calculation by FLIM 117
Acknowledgments 121
References 123

List of Figures and Tables

Figure 1.1 Regulation of DNA replication by origin usage. 4

Figure 1.2 Model of the regulation of DNA replication. 8

Figure 1.3 Schematic pictures of ORC1-6 and Cdc6 proteins from *S. cerevisiae*. 9

Figure 1.4 Proposed model for the evolution of multiple pre-RC regulatory pathways..12

Figure 1.5 Helicase activation and replisome loading at the onset of S phase. 16

Figure 1.6 Models for MCM helicase activity. 17

Figure 1.7 3D reconstructions of the RC proteins structure imaged by electron microscopy. 18

Figure 1.8 AT-rich and CpG-rich islands at DNA replication origins...........................20

Figure 1.9 Model describing the “opportunistic” origin specification during G1 phase...21

Figure 1.10 *Hox* Genes: evolutionary conservation of genomic organization and expression patterns from *Drosophila* to mammals. 26

Figure 1.11 The homeodomain tertiary structure...28

Figure 1.12 Homeodomain-DNA interactions between HOX proteins and the canonical TAAT core. 28

Figure 1.13 Model of the two-step mechanism of HOX specificity.. 31

Figure 1.14 Model of the two levels of HOX-DNA binding specificity.......................32

Figure 1.15 Model of Geminin interplay between proliferation and differentiation. 36

Figure 1.16 Nuclear FRAP procedures. 38

Figure 1.17 FRET to study mammalian cell nuclei. 39

Figure 1.18 Model of ORC dynamics. 41

Figure 2.1 Expression and subcellular localization of endogenous HOXC13 in U2OS and T98G cells. 43

Figure 2.2 Nuclear distribution of endogenous chromatin-bound HOXC13 in U2OS and T98G cells. 44

Figure 2.3 Western Blot detection of endogenous HOXC13 in comparison with HP1β in biochemically fractionated asynchronous U2OS cells. 45

Figure 2.4 Fluorolabelled HOXC13: design of fusion constructs and expression in human cells.......46

Figure 2.5 Check of correct expression and functionality of recombinant fluorolabelled HOXC13.......47

Figure 2.6 Impact of HOXC13 over-expression on HeLa cell-cycle progression.......48

Figure 2.7 Definition of replication foci (RF) nuclear patterns in U2OS cells.......50

Figure 2.8 Co-localization analysis of fluorolabelled HOXC13 with RF throughout S phase progression.......54

Figure 2.9 Cell-cycle modulated presence of HOXC13 at RF.......54

Figure 2.10 Time-lapse imaging of S phase progression in NIH3T3 cells expressing EGFP-HOXC13 and RFP-PCNA.......55

Figure 2.11 Fluorolabelled deletion mutant of HOXC13.......56

Figure 2.12 Half-FRAP experiments performed with the EGFP-HP1α protein in U2OS cells.......58

Figure 2.13 FRAP curves for EGFP- and mCherry- HOXC13 fusion constructs.......60

Figure 2.14 Dependence of wt HOXC13 nuclear dynamics on expression levels.......61

Figure 2.15 Half-FRAP analysis of wt mCherry-HOXC13.......63

Figure 2.16 FRAP analysis of three homeodomain (HBX) mutants of mCherry-HOXC13.......64

Figure 2.17 Quantitative analysis of wt mCherry-HOXC13 nuclear dynamics.......67

Figure 2.18 Fluorolabelled RC proteins: design of fusion constructs and expression in human cells.......69

Figure 2.19 Subnuclear distributions of E^0GFP-ORC2 in human living cells.......70

Figure 2.20 Heterochromatic nature of E^0GFP-ORC2 nuclear aggregation.......70

Figure 2.21 Dynamics of E^0GFP-ORC2 localization at the G1/S transition.......71

Figure 2.22 Subcellular distributions of E^0GFP-Cdc6 in human living cells.......72

Figure 2.23 Dynamics of E^0GFP-Cdc6 localization at the G1/S transition.......72

Figure 2.24 Subnuclear distributions of E^0GFP-ORC1 in human living cells.......73

Figure 2.25 Subnuclear distribution of E^0GFP-MCM3 in human living cells.......74

Figure 2.26 ORC2/HOXC13 co-localization in human living cells.......76

Figure 2.27 Cdc6/HOXC13 co-localization in human living cells.......77

Figure 2.28 ORC1/HOXC13 co-localization in human living cells.......78

Figure 2.29 MCM3/HOXC13 co-localization in human living cells. 78

Figure 2.30 FLIM analysis of HOXC13 interaction with RC proteins. 81

Figure 2.31 Analysis of the donor lifetime maps obtained by FLIM measurements. 83

Figure 2.32 FLIM analysis of ORC2/HOXC13 interaction resolved within cell-cycle progression. 85

Figure 2.33 *In vitro* HOXC13 affinity for Cdc6 and ORC2 detected by co-immunoprecipitation. 86

Figure 3.1 Model for the spatio-temporal determination of RF by pre-RC sites assembled during the G1 phase of the cell-cycle. 90

Figure A.1 Fluorescence re-distribution in a half-bleached nucleus. 108

Figure A.2 Maps of Log_{10} (χ^2) obtained in the fitting procedure. 109

Figure B.1 Spectral properties of the E^0GFP/mCherry FRET pair. 114

Figure B.2 Fluorescence intensity decays for E^0GFP in the presence of mCherry. 116

Figure B.3 Image analysis for the calculation of *E* for the ORC2-HOXC13 and Cdc6-HOXC13 interaction. 118

Table 2.1 Co-localization analysis of fluorolabelled HOXC13 with RF throughout S phase progression. 53

Table 2.2 Mean lifetime values obtained with FLIM analysis.. 82

Table B.1 *E* and distance calculation by FLIM. 119

List of Abbreviations

ATP = Adenosine triphosphate

BrdUrd = Bromo deoxy-Uridine

cDNA = complementary DNA

CDK =cyclin-dependent-kinase

co-IP = co-immunoprecipitation

DDK = Dbf4-dependent-kinase

Dnmt1 = DNA-methyl-transferase 1

dsDNA = double-stranded DNA

DTT = Dithiothreitol

EDTA = Ethylenediamine tetra-acetic acid

EtOH = Ethanol

FBS = Fetal bovine serum

FP(s) = fluorescent protein(s)

FLIM = Fluorescence Lifetime Imaging Microscopy

FRAP = Fluorescence Recovery After Photobleaching

FRET = Fluorescence Resonance Energy Transfer

GFP = green fluorescent protein

h = hours

IF = immunofluorescence

IP = immunoprecipitation

min = minutes

MW = molecular weight

NLS = Nuclear Localization Signal

NMR = Nuclear Magnetic Resonance

MCM = Mini-Chromosome-Maintenance proteins

NTP = Nucleoside triphosphate

ORC = Origin Recognition Complex

PBS = Phosphate buffered saline

PCNA = Proliferating Cell Nuclear Antigen

PCR = Polymerase chain reaction

POL = DNA polymerase

pre-IC = pre-initiative complex

pre-RC = pre-replicative complex

RC = replicative complex

RF = replication foci

rpm = revolutions per minute

s, sec = seconds

SSB = single-stranded binding protein

ssDNA = single-stranded DNA

t = time

TALE = Three Aminoacid Loop Extension

UV = Ultra-Violet

WB = Western Blot

List of Publications

L. Marchetti, L. Comelli, B. D'Innocenzo, L. Puzzi, S. Luin, D. Arosio, M. Calvello, R. Mendoza-Maldonado, F. Peverali, F. Trovato, S. Riva, G. Biamonti, G. Abdurashidova, F. Beltram, A. Falaschi, "Homeotic proteins participate in the function of human DNA replication origins", *submitted*.

L. Comelli, **L. Marchetti**, D. Arosio, S. Riva, G. Abdurashidova, F. Beltram and A. Falaschi, "The homeotic protein HOXC13 is a member of human DNA replication complexes", *Cell Cycle* 2009 (8): 454-9.

L. Albertazzi, D. Arosio, **L. Marchetti**, F. Ricci, F. Beltram "Quantitative FRET analysis with the E0GFP-mCherry fluorescent protein pair", *Photochem. Photobiol.* 2009 (85): 287-97.

Publications not included in this thesis:

D. Arosio, F. Ricci, **L. Marchetti**, R. Gualdani, L. Albertazzi and F. Beltram, "Simultaneous Intracellular Chloride and pH measurements using a GFP-based sensor", *submitted*.

M. Hafner, E. Vianini, B. Albertoni, **L. Marchetti**, I. Grüne, C. Glöckner, M. Famulok "Displacement of protein-bound aptamers with small molecules screened by fluorescence polarization", *Nat. Protoc.* 2008 (3): 579-587.

D. Arosio, G. Garau, F. Ricci, **L. Marchetti**, R. Bizzarri, R. Nifosì, F. Beltram "Spectroscopic and structural study of proton and halide ion cooperative binding to gfp", *Biophys J.* 2007 (93): 232-44.

ABSTRACT

The regulation of human DNA replication operates via the time-programmed activation and deactivation of approximately 30,000 replication origins distributed along the genome. A multi-protein replicative complex recognizes and assembles onto each origin; this determines the local unwinding of the origin DNA and the start of two oppositely moving replicative forks. The mechanism that governs the selection of a specific DNA sequence as human (and, more generally, metazoan) origin, in the course of G1 phase of the cell-cycle, is still poorly understood. The lack of DNA-sequence consensus among well-characterized replication origins, together with the little binding-specificity displayed by the Origin Recognition Complex, suggest that origin selection might rather be determined by local chromatin structures and/or accessory targeting proteins. With regard to the latter possibility, it was interesting to find out that three homeotic proteins, namely HOXC13, HOXC10, and HOXA13 display a specific affinity for a DNA fragment corresponding to the sequence covered by the Replicative Complex of the human Lamin B2 replication origin.

In the study conducted during this Ph.D. program, the possible role of homeotic proteins in origin function was explored by investigating the involvement of a selected homeotic protein, namely HOXC13, within the replicative complexes in living human cells. To this purpose, recent advances in biophysical microscopy technologies were exploited to study *in vivo* the localization, dynamics, and interactions of HOXC13 protein in the context of DNA replication regulation. The data reported in this thesis demonstrate that HOXC13 indeed participates in origin function. The protein is a stable component of early replicating chromatin, as it displays stable chromatin binding in correspondence to the nuclear areas where replication foci of early S phase are collected. This peculiar behavior is driven by the homeodomain and relies mainly on the conserved homeodomain arginine-5 anchoring to the DNA minor groove. Furthermore, HOXC13 displays unambiguous affinity for origin sequences and for selected replicative-complex proteins. The close proximity of HOXC13 to both Cdc6 and ORC2 proteins measured in living cells proves that the homeotic protein is involved in direct protein-protein interactions within the replicative-complex; not unexpectedly, such interactions are modulated in a cell-cycle dependent fashion that is consistent with origin function.

These observations are not restricted to a single origin, but rather appear to have a general significance in the nuclear architecture of DNA replication; nor are they restricted to a single homeotic protein, as the HOXC13 exerts its function via highly conserved homeodomain residues. Hence, this dissertation argues that the homeoproteins functionally contribute in a general manner, dependent on their chromatin-binding properties, to the specification of origins, likely the early replicating ones. In this view, HOX proteins, probably in the context of a multi-protein homeotic effector, contribute to recruit and stabilize the replicative complexes onto early replicating origins, in presence of specific chromatin and topological configurations. Considering that HOXC13, involved in development and differentiation, is also an oncoprotein, the data presented in this thesis, besides offering an indication for the basis of origin selection, hint at the homeotic proteins as actors in the cross-talk between development and DNA replication regulation.

1. INTRODUCTION

The work reported in this dissertation explores the possible connection between two traditionally separated fields of biology, namely the regulation of DNA replication and the function of homeotic proteins. Therefore, in order to provide the conceptual frame of my experimental work, the two following paragraphs will be focused respectively on a description of the mechanisms of DNA replication (**paragraph 1.1**) and of the structure and function of homeotic proteins (**paragraph 1.2**). In particular, I shall underline what is still missing for a satisfactory understanding of DNA replication regulation in metazoan organisms, and to what extent could the homeotic proteins be involved in principle in the process. Moreover, in this thesis I describe novel technological approaches for the investigation of the dynamics of DNA replication regulation in living cells. They result from recent advances in biophysical and microscopy technologies, providing improved strategies and tools for approaching problems in modern proteomics, namely high-resolution imaging of biological processes and *in vivo* analysis of protein dynamics and protein-protein interactions. Accordingly, in **paragraph 1.3** I shall review the main advantages offered by biophysical approaches and the results achieved through them so far in the study of nuclear dynamics and of the subnuclear details of DNA replication mechanisms. Finally, **paragraph 1.4** will clearly indicate the scope of this thesis and how it has oriented the choice of the experimental approach.

1.1 Regulatory mechanisms of eukaryotic DNA replication

1.1.1 Eukaryotic DNA replication: an overview

DNA replication is a tightly regulated and complex process whereby the exact duplication of a genome is achieved, only once per cell division, in every proliferating cell. The regulation of DNA replication in eukaryotic genomes appears to be far more complex than in bacteria, where replication occurs starting from a single origin (Figure 1.1a). The latter case is suitably described by the original replicon model proposed in 1963 by Jacob and Brenner, who postulated the existence of two important elements required for replication initiation: the replicator and the initiator [1]. The replicator is the cis-acting sequence within the genome from which replication starts; the initiator is a positive trans-acting factor able to recognize specifically the sequence of the replicator within the genome. In response to the appropriate cellular signals, the initiator directs the local unwinding of the replicator sequence and recruits additional factors to initiate the process of DNA replication. Eukaryotic genomes are very large and their replication rate is slow, when compared to the prokaryotic replicon model [2]. Nevertheless, the process of DNA replication is restricted to a relatively narrow window of the eukaryotic cell-cycle, namely the S phase. This is made possible by the start of DNA replication at thousands of different chromosomal locations that are specifically selected (Figure 1.1b). These sites are referred to as origins of DNA replication, and the process of DNA synthesis relies on a spatio-temporal coordinated cycle of activation and deactivation of the origins. The advantage of this mechanism, besides reducing the overall time required to duplicate the entire genome, is that the generation of single-stranded DNA (ssDNA) is much more localized and transient, helping preserving the genome integrity [3, 4]. Actually, the activation of all origins dislocated in eukaryotic genomes leads to the formation of tandemly arranged replication units, and each one can be putatively considered as an analog of the bacterial replicon [5].

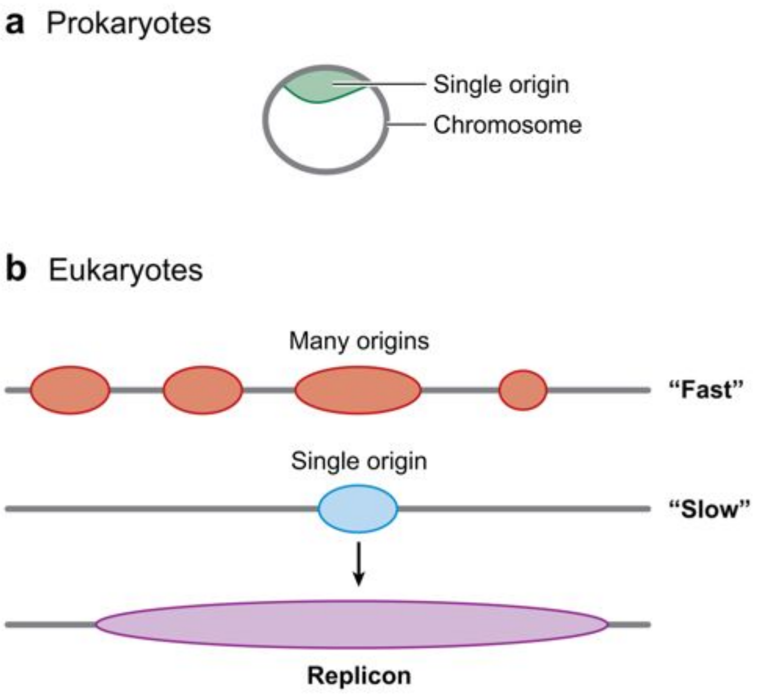


Figure 1.1 Regulation of DNA replication by origin usage. a) Prokaryotes have a single origin on a circular chromosome. b) In eukaryotes, multiple origins are found on a single, linear chromosome. This is useful to achieve a "fast" replication, whereas if only one origin were used in this region replication

would be “slow”. Replication proceeds bi-directionally from an origin to form a replicon (bottom). Taken from Sclafani *et al*. [6].

As a process, DNA replication can be divided into three steps: initiation, elongation and termination. During initiation [7], a DNA sequence is selected to be an origin (i.e. the start site) of DNA replication, usually in correspondence to AT-rich sequences, and initiator proteins assemble thereon. This results in the formation of a multi-protein complex which is responsible of the local melting of the duplex, that is necessary for the proteins to have access to the template strands; subsequently, the complex stabilizes the ssDNA that is formed, and two replication forks comprising DNA helicases and polymerases start to replicate the two parental DNA strands in opposite directions. The elongation [8, 9] is actually the continuation of the unwinding activity by the two fork complexes that ensures simultaneous replication of both parental DNA strands also outside of the origin sequence. When two replication forks converge, they merge and termination of replicon duplication can occur [10, 11].

Significant differences exist between DNA replication mechanisms in lower and higher eukaryotes. In the former ones, replication occurs at site-specifically defined replication origins (e.g. 250 - 400 in *S. cerevisiae*) that share conserved features in their sequence [3]. Conversely, higher eukaryotic organisms display a number of origins that is at least 100-fold higher and, at present, no sequence specific replicators have been found [12-15]. For example, it is estimated that in human cells replication initiates from 30,000 replication origins, a complexity that is difficult to harmonize with the simple replicon model. In spite of theses disparities between lower and higher eukaryotes, the proteins that regulate replication are highly conserved in function from yeast to *Drosophila*, from *Xenopus* to human, suggesting a common mechanism in the replication function that does not depend on the origin sequence itself [16-19]. In addition, structural and functional homologues of many proteins involved in DNA replication also exist in *Archea*. Even in these organisms, the starting mechanism of DNA replication is similar to eukaryotes and involves the binding of a multi-protein complex to the replication origin [20].

1.1.2 Multiple levels of DNA replication regulation within cell-cycle progression

The complexity of DNA replication in higher eukaryotes implies many levels of regulation including origin decision, timing of origin firing and inhibition of origin re-firing. Although DNA is replicated only during the S phase of the cell-cycle, these mechanisms take place at different times during cell-cycle progression, to ensure a stepwise regulation of the process and its coordination with the other events of the life cycle of the cell. The first regulatory step of DNA replication concerns the initiation, i.e. the activation of the replication origins. This occurs starting from the end of M phase and onset of G1 phase [21], when several proteins taking part to the pre-replicative complex (pre-RC) select the DNA sequences to be replication origins and, by binding to these regions, commit the recruitment of other proteins involved in origin firing. By the beginning of S phase, after origin firing, the pre-RC gets reorganized at the origin due to the modification or degradation of several of its members, as regulatory mechanisms to avoid re-replication [22-24]. It is actually the temporal separation of pre-RC assembly from origin activation that ensures that new pre-RC cannot assemble on origins which have already fired [22, 25]. These mechanisms rely on the activity of cell-cycle regulated kinases, among which CDK (cyclin-dependent kinase) that act on several target proteins [7], and will be listed in detail in the next **paragraph 1.1.3**. Because this kinase activity remains high from S phase onset to the end of the following mitosis, re-licensing cannot occur until the beginning of the next cell cycle [22, 26, 27].

Meanwhile in G1 phase, for all the sequences selected as origins, a timing of replication initiation is assigned and only a subset of these will fire immediately after entry into S phase (early origins). The rest (middle and late origins) are programmed to fire in ordered manner later on after early origins. This results in an organized spatio-temporal activation of replication clusters of different subchromosomal domains at different times during S phase [28, 29]. CDKs have also been implicated in controlling the time of replication initiation at specific origins [29, 30]. However, the mechanism that regulates the timing of replication is not completely understood and currently under intense investigation. Originally, because transcriptionally active euchromatic regions replicate early and inactive heterochromatic regions late, it was thought that early replication is a prelude to transcription. Under this hypothesis, transcription requires an opening of chromatin which allows also easy access to replication factors. Up to date, the relationship between transcription and replication in regulating the temporal program is unclear in that it has never been established whether transcription of these regions causes replication or vice versa [6, 31]. Furthermore, the timing of origin activation has been reported to correlate with a developmental program rather than with transcription *per se* [32, 33]. Recent advances in DNA microarray technology have enabled eukaryotic replication to be studied at whole-chromosome and genome-wide levels. These studies, in both *S. cerevisiae* and higher eukaryotes, have provided new insights into the temporally coordinated activation of replication initiation [34]. In detail, they have revealed clear connections between chromosome organization and replication timing. For example, in yeast, the centromeric proximal sequences are consistently early replicating and telomeric regions are consistently late replicating.

Studies in metazoa have so far confirmed the recurring correspondence between replication and transcriptionally active regions [14, 15, 34].

During the course of S phase, several other proteins ensure the fork progression during DNA replication elongation. Given the complexity and the importance of this phase for the maintenance of genome integrity, many different checkpoint pathways are active within this time window, as demonstrated by studies in the yeast model [35]. These checkpoints encompass the whole phase of DNA synthesis, as well as the switch to G2 phase, and comprise a variety of mechanisms to prevent replication defects, repair damaged replication forks, and enable fork reactivation. For their role in the overall control of the cell-cycle progression, as well as the control of genomic stability, they are often referred to as cell-cycle checkpoints [36, 37]. Very interestingly, often the induction of a cell-cycle checkpoint results in the retroactive regulation of the recruitment of key members of the pre-RC to the origin site. For example, in budding yeast, hydroxyurea treatment not only blocks fork progression from early-origins but also prevents the firing of late-origins, and this mechanism was shown to depend on Rad53 and Mec1 [38]. The same conclusion was obtained also following induction of double-strand breaks, and the protein involved in this regulation was shown to be yeast ORC2 [39]. Moreover, a post-S phase role was recently proposed for the control of re-replication also for the initiator protein Cdc6 [40, 41].

Altogether, these data point to the same factors involved in DNA replication initiation as important actors also in the regulation of the replication process at different stages during the cell-cycle, being the targets of many checkpoint controls. It is therefore worth to go into more depth with regard to pre-RC proteins and the switch from initiation to elongation, as will be discussed in the next paragraph.

1.1.3 From initiation to elongation in eukaryotic DNA replication

DNA replication starts from the stepwise recruitment of the replication machinery to the various origins on the chromosome. The recruitment process is called initiation, to distinguish it from the subsequent replication of the DNA by the replisome that is called elongation. As explained in the previous paragraph, initiation is a major point at which DNA replication is regulated: the ordered recruitment of these proteins onto the origin is indeed responsible for controlling the process of initiation of DNA replication in terms of both space and time; furthermore, their subsequent inactivation or removal is believed to be necessary to prevent re-replication during S phase. For these reasons, initiator proteins are crucial in regulating origin activity.

The basic mechanism of initiation occurs in several steps that finally lead to bidirectional replication from the origin (Figure 1.2), namely:

1. **Recognition**: labeling of the origin by ORC (Origin Recognition Complex), Cdc6 and Cdt1;
2. **Initiative assembly** (or Licensing): loading of the MCM DNA-helicase to form the pre-RC;
3. **Unwinding**: activation of the MCM DNA helicase;
4. **Elongative assembly**: loading of the complete replisome, including DNA polymerase enyzmes and SSB (single-stranded DNA binding protein).

The ordered sequence of these four steps allows the switch from initiation of elongation; each of them will be briefly summarized below.

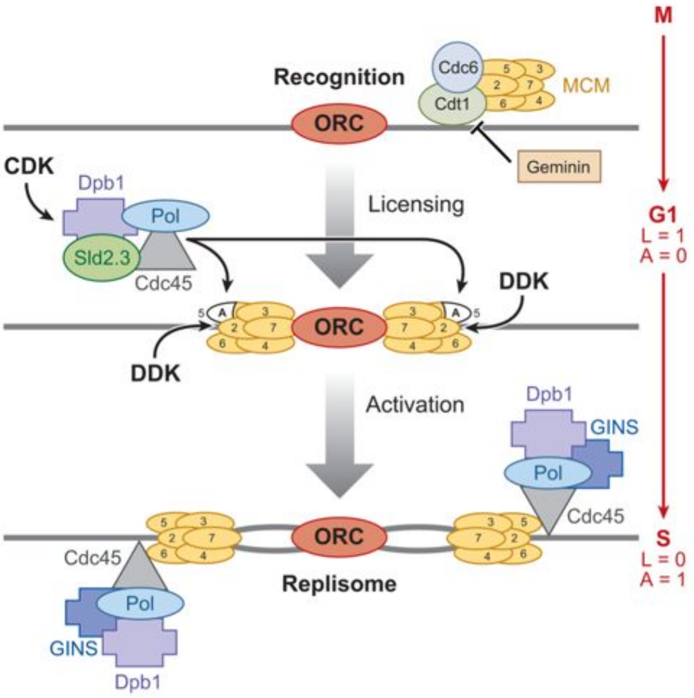

M
Recognition
Cdc6
Cdt1
MCM
ORC
Geminin
CDK
Dpb1
Pol
Sld2.3
Cdc45
Licensing
G1
L = 1
A = 0
DDK
ORC
DDK
Activation
Dpb1
GINS
Pol
Cdc45
S
L = 0
A = 1
ORC
Replisome
Cdc45
Pol
GINS
Dpb1

Figure 1.2 Model of the regulation of DNA replication. In eukaryotic DNA replication, a replication origin is recognized by ORC, and then Cdc6 and Cdt1 proteins load the MCM helicase to form the "licensed" (L) pre-RC in G1 phase (L = 1, A = 0). Geminin inhibits Cdt1 and pre-RC formation. CDK and DDK become active in late G1, activate (A) the MCM helicase and load on the replisome that contains the DNA polymerases. In addition, CDK inhibits any further licensing (L = 0, A = 1). To this end, CDK phosphorylates Sld2 and Sld3 proteins and DDK phosphorylates MCM proteins, which "pushes out" the "A" domain of Mcm5. Adapted from Sclafani *et al.* [6].

1. Recognition: in this step, ORC recognizes and marks the origins, which is proposed to occur between late M and G1 phase [21], and provides a “landing-pad” for other two proteins coming during the course of G1 phase, namely Cdc6 and Cdt1. The ORC is a six-protein complex containing ORC1–6 proteins (Figure 1.3) in equal stoichiometry and was first isolated in yeast cells for the specific recognition of origin sequences [42]. Although the ORC1–6 proteins are conserved in evolution (Figure 1.3B), the recognition of specific sequences is a property lost in ORC of most higher eukaryotes, and only a preference for AT-rich sequences can be observed. This is expected because origins have no consensus DNA sequence in metazoa (see **paragraphs 1.1.1** and **1.1.5**). The most striking example of this phenomenon is when recombinant ORC1–6 proteins from human replaced the frog ORC1–6 proteins *in vitro* to initiate DNA replication in a sequence-independent manner [43].

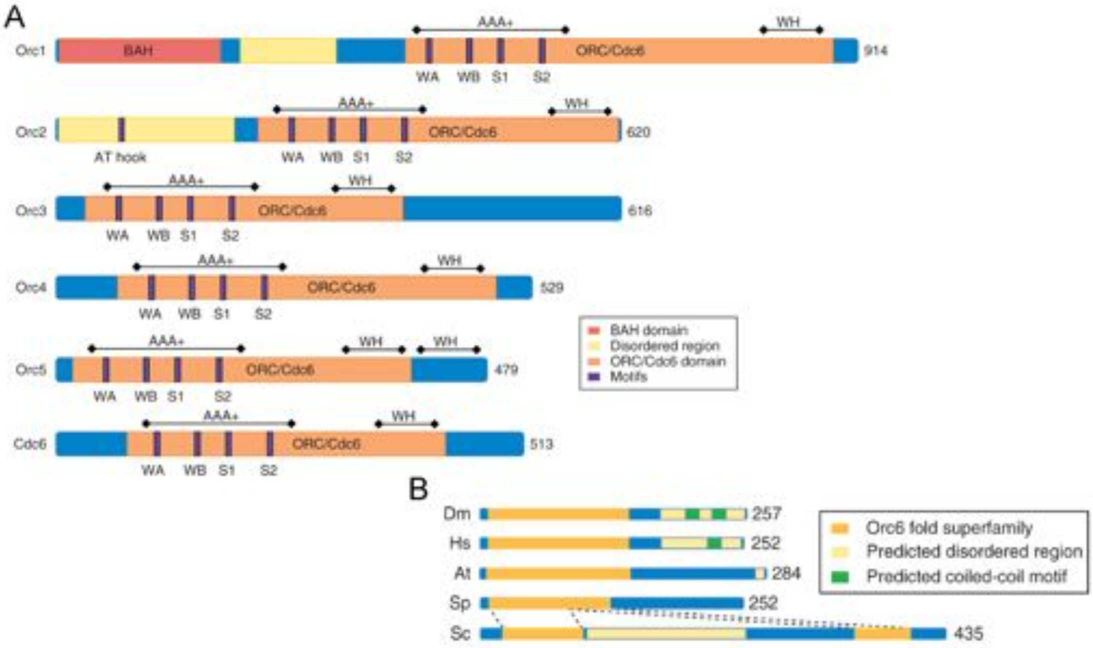

A
Orc1
BAH
AAA+
WH
ORC/Cdc6
914
WA WB S1 S2
Orc2
AAA+
WH
ORC/Cdc6
620
AT hook
WA WB S1 S2
Orc3
AAA+
WH
ORC/Cdc6
616
WA WB S1 S2
Orc4
AAA+
WH
ORC/Cdc6
529
WA WB S1 S2
Orc5
AAA+
WH
WH
ORC/Cdc6
479
WA WB S1 S2
Cdc6
AAA+
WH
ORC/Cdc6
513
WA WB S1 S2
BAH domain
Disordered region
ORC/Cdc6 domain
Motifs
B
Dm 257
Hs 252
At 284
Sp 252
Sc 435
Orc6 fold superfamily
Predicted disordered region
Predicted coiled-coil motif

Figure 1.3 Schematic pictures of ORC1-6 and Cdc6 proteins from *S. cerevisiae*. A) ORC1-5 and Cdc6 each contain an AAA+ domain as part of a larger ORC/Cdc6 homology domain (orange). Motifs within the AAA+ domain include Walker A (WA), Walker B (WB), Sensor-1 (S1) and Sensor-2 (S2). The winged-helix domain (WH) is involved in DNA binding. ORC1 contains an additional BAH (bromo-adjacent homology) domain (pink). ORC1 and ORC2 have disordered regions (yellow); a DNA-binding AT- hook motif (here PRKRGRPRK) is identified in *S. cerevisiae* ORC2, and several of these have also been identified in disordered regions of *S. pombe* ORC4. B) Homology between ORC6 in *D. melanogaster* (Dm), *H. sapiens* (Hs), *A. thaliana* (At), *S. pombe* (Sp), and *S. cerevisiae* (Sc). The conserved domain is depicted in orange. ORC6 has no recognizable homology to ORC1-5 or AAA+ domains. The C-terminal region of ORC6 in *D. melanogaster* has been proposed to be a coiled-coil motif. The number of amino acids for each protein is indicated at the right side. Adapted from: Duncker *et al.* [44].

To date, it is not clear which DNA or chromatin structure the ORC from most organisms recognizes. ORC may recognize a unique chromatin structure dictated by epigenetic determinants and not primary DNA sequence. This recent theory is supported by several observations. Histones at the origins in *Drosophila* follicle cells are hyperacetylated and changes in the acetylation level affect ORC binding [45]. In frog, acetylated histones are preferentially found at active origins [46]. Recently, ORC was found to map at a precise nucleosome positioning [34]. Overall, these observations fall within the more general riddle of how replication origins are actually determined in metazoa, which will be addressed in detail in **paragraph 1.1.5**.

Most ORC subunits belong to the superfamily of AAA+ ATPases (ATPases Associated with various cellular Activities) and share conserved motifs, except for ORC6 [44] (Figure 1.3). The ATP-binding activity is required in the process of origin DNA recognition. Indeed, in *S cerevisiae* the ORC1 ATPase activity is inhibited until Cdc6 protein, which is also an AAA+ ATPase (Figure 1.3A), is recruited and activates ORC1 ATPase, thus resulting in the specific recognition of the origin (see also Figure 1.7A and next paragraph) [47]. The role of the ORC6 protein is controversial. ORC6 is required for viability in yeast but is not required for DNA binding *in vitro*. In metazoan cells, complexes with lower amounts of ORC6 than the other ORC1–5 proteins still are active and in *Drosophila* all six subunits are needed [44].

In yeast, the ORC is bound to origins throughout the cell cycle and re-replication is avoided by phosphorylation of ORC2 and ORC6 by CDK1. However, in other eukaryotes, ORC binding is regulated, based on a mechanism known as the "ORC cycle" [48]. This is used to avoid re-replication and consists in the dissociation of ORC1 from the chromatin-bound ORC2–5 complex and its subsequent degradation in cells at the end of G1 phase and beginning of S phase [23, 49]. The process is regulated by CDK1-cyclin A phosphorylation [50].

Recent research avenues have identified roles for ORC proteins other than direct controlling DNA replication initiation. ORC1 has been reported to participate in gene silencing via its BAH domain (Figure 1.3) providing a direct interaction with heterochromatin protein Sir1 in *S. cerevisiae* [51], as well as with heterochromatin protein 1 (HP1) in *Xenopus* and *Drosophila* [52]. In both cases ORC1 helps Sir1 and HP1 to propagate silenced chromatin. Also ORC2 has been reported to be important for heterochromatin maintenance [53]. In metazoan cells, ORC localization clearly extends beyond origin sequences (reviewed by Chesnokov *et al.* [54]). Studies in human cells have revealed that ORC6 also localizes to the cleavage furrow in dividing cells, and a role for this protein in cytokinesis has been confirmed by RNA interference depletion of the protein [55]. In addition, human ORC6 was shown to localize to kinetochores and reticular-like structures around the cell periphery during mitosis, and to be required for the proper progression of this cell-cycle stage [56]. Human ORC2 also localizes to the centrosome throughout the cell-cycle and its depletion results in mitotic defects and multiple centrosomes [53]. Recently, a similar role in controlling centrosome copy number was reported for human ORC1 [57].

2. Initiative assembly: The next step is to load the DNA helicase onto the origin (Figure 1.2). This is accomplished by at least two proteins, Cdc6 and Cdt1, that recruit the MCM helicase to finally achieve the pre-RC assembly onto the origin. "Replication licensing" is a useful term that is used to describe the process in which origins are "licensed" when the MCM helicase is loaded onto them in G1 of the cell-cycle [58]. Thus, pre-RC formation equates with "licensing."

Cdc6 is also an AAA+ ATPase (see Figure 1.3A), which is required to load on the MCM helicase in G1 phase, as shown in experiments performed in budding yeast which also revealed the importance of its ATPase activity to exert this function [59]. In detail, the Cdc6 and ORC ATPases act sequentially, with Cdc6 required initially. In a recently proposed model, Cdc6 and origin chromatin set off a molecular switch in ORC for pre-RC assembly [47]. Indeed, in *S. cerevisiae* the ORC1 ATPase activity is inhibited until Cdc6 protein is recruited and activates ORC1 ATPase. This then produces a conformational change in the ORC-Cdc6-DNA complex to achieve a ring-like structure (see below, Figure 1.7A) with increased specificity for the origin sequence. Origin DNA inhibits ATP hydrolysis by Cdc6 and stabilizes the complex, whereas mutations in the origin sequence can increase Cdc6 ATPase activity, resulting in a less stable Cdc6-DNA complex. This means that ORC binding to the origin is not specific unless Cdc6 is also bound, i.e. Cdc6 rather than ORC is responsible of origin selection [47]. The proposed structure of Cdc6, which was deduced by comparison with the ORC structure, is similar to the atomic structure of the archaeal homologue, ORC1/Cdc6. ORC1 and Cdc6 proteins are homologues (Figure 1.3), and indeed archaeal species have one protein Orc1/Cdc6 that does both functions, i.e., origin recognition and the loading of the MCM helicase. Some archaeal species, such as *Sulfolobus solfataricus,* have multiple Orc1/Cdc6 proteins that bind to two different origins [60].

Cdt1 protein, like Cdc6 protein, is also required to load the MCM helicase during G1 of the cell cycle of eukaryotes [7]. This protein, which was initially found in fission yeast, is clearly conserved in eukaryotic evolution. As Cdc6 ATPase is needed for Cdt1 binding on the origin *in vitro*, it has been proposed that a Cdt1-MCM complex is loaded onto the ORC-Cdc6-origin complex during initiation [59, 61]. Cdt1 and Cdc6 then dissociate and finally ATP hydrolysis by ORC completes the MCM helicase loading reaction [47, 59].

As stated in **paragraph 1.1.2**, licensing is blocked during S, G2, and M phases of the cell-cycle to prevent re-replication. Re-replication is actually avoided by the concurrence of several, redundant mechanisms that block MCM loading during S, G2 and M phases. In other words, pre-RCs can be assembled only in G1 phase, but are activated for origin firing only during S phase. There never exists a cell-cycle phase in which pre-RC formation (licensing) and activation can occur. In binary terms, if off = 0 and on = 1, then there is no phase in which replication licensing (L) and activation (A) are both 1. In G1 phase, L = 1 and A = 0, while L = 0 and A = 1 in S phase [6] (Figure 1.2). A major level of regulation is catalyzed by CDK, which acts at many redundant levels to block licensing in most eukaryotes [62]. These levels include the degradation and localization of several pre-RC components. Besides the already mentioned ORC1 protein in higher eukaryotes, another modified protein is Cdc6: in yeast it is degraded after CDK phosphorylation [63], while in mammals Cdc6 is exported from the nucleus after CDK phosphorylation [64]. Another level of regulation to block re-replication occurs through a protein known as Geminin (Figure 1.2), which was discovered in frog egg extracts [65] and is only found in metazoans. Geminin binds to and inhibits Cdt1 and thus prevents replication licensing by blocking the loading of the MCM helicase [65]. The same role for Geminin is reported to occur also in human [24]. The finding that Geminin specifically recognizes and binds HOX proteins [66] has also shed light on a possible concomitant role for this protein in development, as will be reviewed in **paragraph 1.2.5**.

The redundancy of these mechanisms avoiding re-replication has been proposed to provide a key driving force in the evolution of licensing control [22, 67] (Figure 1.4). Because no single mechanism for inhibiting pre-RC components can be completely effective, multiple mechanisms are required for achievement of an efficient block to re-replication.

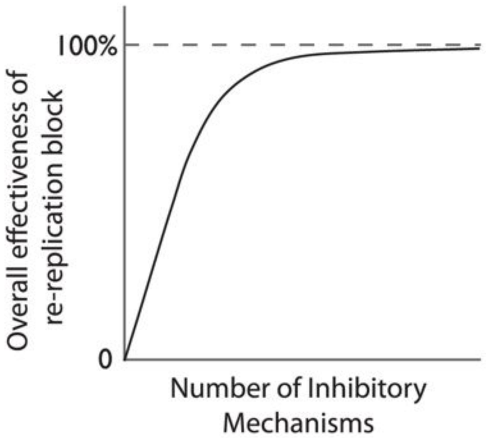


Figure 1.4 Proposed model for the evolution of multiple pre-RC regulatory pathways. Details are in the main text. Taken from: Drury *et al.* [63].

However, as the number of inhibitory mechanisms increases, the relative importance of any single mechanism decreases. During evolution, these regulatory mechanisms may be gained or lost, and an organism may ‘‘sit’’ at different positions along the curve in Figure 1.4. This could explain the different regulatory mechanism of Cdc6 in yeast and human as well as the appearance of Geminin in metazoa to provide an additional mechanism for preventing re-replication, which may have been important in supporting an increase in genome size with respect to lower eukaryotes.

3. Unwinding and 4. Elongative assembly: these two steps refer to the activation of the helicase activity at the origin and to the replisome assembly, respectively; they are closely interconnected, and will therefore be reviewed together. The MCM complex is believed to be the engine of the replicative helicase. This complex is a hexamer comprising six related polypeptides (Mcm2–7) coded by a family of six paralogous genes, which are conserved from yeast to human. All six members of the gene family are essential genes in both budding and fission yeast. All are AAA+ ATPases with similarity to DNA helicases. A pioneering work in fission yeast first identified a complex that contained all six subunits in 1:1:1:1:1:1 stoichiometry and had a ring-like structure [68]. Thereafter, a large body of work has been done in order to elucidate the structure of the MCM ensuring the helicase activity to the replisome, as will be discussed in detail in the next paragraph. Here it is just worth to recall that most of these studies have been performed with MCM from yeast and *Archea*.

In G1 phase, pre-RCs with the Mcm2–7 helicase bound are present on almost all origins. Indeed, nearly 90% of all origins that are bound by ORC also have MCM complex bound [69]. Nevertheless, MCM is loaded inactive in the pre-RC, when CDK activity is low. The next step is to activate the MCM helicase. This is achieved by the binding of several other proteins to the origin, up to the loading of the replisome (Figures 1.2 and 1.5). The multi-protein complex assembled on the origin at this stage is referred to as pre-initiation complex (pre-IC), and is required for the activation of the Mcm2–7 helicase. In this manner, the cooperation between helicase activation and replisome loading assures coordinated replication.

Both helicase activation and replisome loading require phosphorylation by CDK (cyclin-dependent kinase) and DDK (Dbf4-dependent kinase, comprising a heterodimer of Dbf4 and Cdc7) enzymes. In budding yeast, there is only Cdk1 or Cdc28 enzyme, but there are six B-type cyclins (Clb1–6) needed for S and M phases [70]. Most likely, Cdk1-Clb5 complexes are the most active in regulating DNA replication and Cdk1-Clb2 for regulating mitosis. In mammals, while there is only one DDK, there are many different CDKs and cyclins with at least four CDKs (Cdk1–4) and four classes of cyclins (A, B, D, and E) required for cell-cycle progression [70]. The Cdk2 homologue is probably used in DNA replication. By analogy, Cdk2-cyclin E and Cdk2-cyclin A act as yeast Cdk1-Clb5 for DNA replication, whereas Cdk1-cyclin B act as yeast Cdk1-Clb2 for mitosis. Thus, it is substrate specificity by different Cdk-cyclin complexes that drives the cell-cycle. CDK and DDK enzymes are regulated independently of each other, but by similar mechanisms [70]. Both kinase subunits are inactive in monomeric form and are activated by the binding of an unstable activating subunit, Cyclin and Dbf4/Drf1 protein for CDK and DDK, respectively. Thus, CDK is Cyclin-dependent kinase and DDK is Dbf4-dependent kinase (in vertebrates, Drf1 is the Dbf4 paralogue). Cell-cycle regulation of the unstable subunit assures cell-cycle regulation of the kinase activity. With CDKs, other levels of regulation occur including protein inhibitor binding, phosphorylation by other kinases, and cyclin subcellular localization [70]. With DDK, it is simpler in that Dbf4 protein is absent in G1 phase because it is targeted for proteosomal degradation by the APC (anaphase promotion complex). As cells enter S phase, the APC is inactivated by CDK phosphorylation and Dbf4 is stabilized. Thus, CDK might have evolved later to coordinate the cell-cycle with DNA replication, while DDK is simpler as it has only a very specific role in DNA replication.

How do the two protein kinases activate the MCM helicase and load on the replisome (Figure 1.2)? A large body of evidence indicates that the Mcm2–7 complex is a target of phosphorylation by DDK, and this occurs in several eukaryotes [7, 71]. Studies performed in yeast have identified phosphorylation sites in the N-terminus of Mcm4, Mcm2 or Mcm6 to be important for formation of the pre-IC and for DNA replication [72]. The hypothesis is that there is considerable redundancy in the system with any of 3 different subunits of the MCM complex acting as targets for DDK. Phosphorylation of the MCM complex by DDK leads to the loading of Cdc45 protein onto origin chromatin with a mechanism conserved from yeast [73] to human [74]. Cdc45 protein is needed for loading of the replisome, including DNA polymerases and RPA, the eukaryotic SSB (Figure 1.5), and moves with the replication fork [75]. There is a conserved genetic interaction between Cdcc45 and the Mcms; specific mutations in these genes can either suppress or show synthetic lethality with the other [30]. How would phosphorylation of Mcm2, Mcm4, or Mcm6 by DDK activate the helicase and load the replisome? One hypothesis is that DDK phosphoryation results in a conformational change in the Mcm5 protein that activates the helicase and is a signal for the binding of Cdc45 protein (Figures 1.2 and 1.5) [6]. In some cases, Cdc45 has been reported to bind the origin earlier in G1 phase, before the MCM activation by DDK. In this case, it is possible that Cdc45 protein may be weakly bound to origin chromatin in G1 phase, and that it is later stabilized by CDK rather than DDK regulation.

The role of CDK in promoting origin activation has also been thoroughly investigated. In yeast, CDK phosphorylates Sld2 and Sld3 [76], causing them to bind Dpb11 (DNA Polymerase B possible subunit, a subunit of DNA polymerase ε holoenzyme, also called Pol2 or PolB), which in turn serves as an anchor for DNA polymerase, RPA and the GINS complex to reach the replisome. The GINS complex, which name is based on the numbers 5, 1, 2, and 3 in Japanese (Go, Ichi, Nii, San), is composed of the Sld5, Psf1, Psf2, and Psf3 proteins, and is required for replication by functioning interdependently with Cdc45 protein in the loading of the replisome. Most of these proteins are conserved in eukaryotic organisms [74]: only yeast Sld3 does not have any homolog in metazoa, while Sld2 and Dbp11 are related to mammalian RecQ4L and TopBP1, respectively; the GINS complex is highly conserved in yeast [73], *Xenopus* [77] and human [78]. Recently, another *mcm* member was identified to take part to the pre-IC assembly, namely Mcm10, which is not a Mcm2–7 homologue [79]. Mcm10 is needed for the loading of the Cdc45 protein after pre-RC formation and for stabilizing the replisome, a mechanism conserved up to human [80]. Mcm10 protein may act by stimulating DDK and DNA polymerase activity at the fork.

To summarize, a large number of proteins is needed to load the replisome onto the origin (all of them appear in Figure 1.5). These proteins help to activate the MCM helicase. Indeed, the association of Mcm2–7, Cdc45, and GINS constitutes a complex named CMG complex which, when purified from *Drosophila* embryos, has helicase activity *in vitro* [81]. Moreover, these proteins bring the DNA polymerases onto the origin, thereby coupling helicase activation and replisome loading. It is also evident that CDK and DDK regulate similar events independently. This explains why both kinases are needed for replication in all eukaryotes examined [6]. In this proposed model, DDK phosphorylates any one of Mcm2/4/6 proteins, resulting in the Mcm5 protein structural change, which together with CDK phosphorylation of Sld2 and Sld3, recruit Cdc45, Dpb11, Mcm10 and GINS proteins; this results in the activation of the Mcm2–7 helicase. The DNA is unwound by the helicase and the DNA is replicated by the replisome. It is not clear what the exact role of Mcm10 is in this model but it is known to be required for Cdc45 loading and replisome stability. In order to avoid confusion in the nomenclature, in all subsequent paragraphs and chapters the proteins listed so far to be part either of the pre-RC or of the pre-IC, will be generally referred to as replicative-complex proteins (RC proteins).

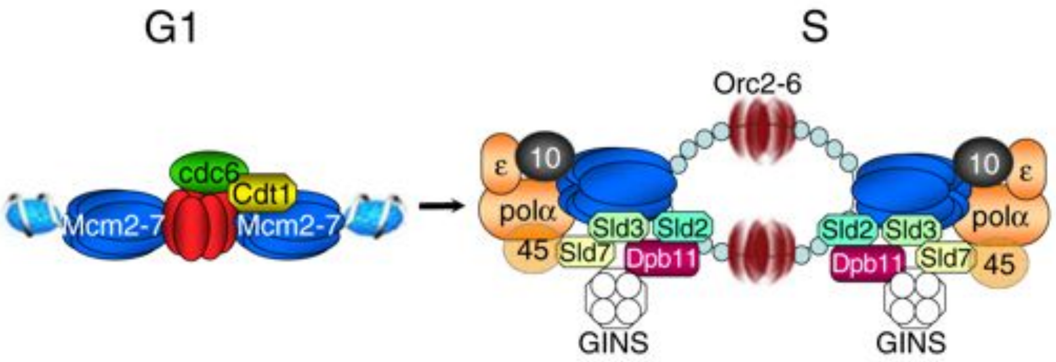


Figure 1.5 Helicase activation and replisome loading at the onset of S phase. Pre-RCs are assembled in a window of opportunity during the G1 phase of the cell-cycle when the CDK and DDK activities are low. In a DDK- and CDK-dependent process, additional complexes bind to origins to form the pre-initiation complex (pre-IC). This includes the Cdc45, Mcm10, the GINS complex (Sld5, Psf1, Psf2, Psf3), Sld2, Sld3, Sld7 and Dpb11, which finally lead to the loading of the polymerase α–primase complex and polymerase ε. RPA (light circles) stabilizes locally unwound ssDNA. Adapted from: Schepers *et al*. [31].

1.1.4 Recurrence of the ring-shape geometry in RC protein structures

In the last years, a large body of work has been done to elucidate the exact mechanisms by which MCM is recruited to the origin, as well as it subsequently exerts DNA unwinding during chromosomal replication. In the latter case, several mechanisms have been proposed to explain how the helicase activity works, namely: i) the *SV40 T-antigen model*, in which MCM resembles the SV40 T-antigen in that two hexamers are put together forming a loop. In this model, the DNA is pumped into the channel of the double hexamer and then extruded out the holes in the outside C-terminal domains of MCM; ii) the *pump-in-ring model*, in which each single hexamer translocates on a different strand of DNA; iii) the *ploughshare model*, in which a double MCM hexamer is loaded onto the origin, then single hexamers translocate in opposite directions along the dsDNA with a ploughshare protein acting as a wedge and helping to keep the ssDNA unwound as it emerges from behind the helicase; iv) the *rotary pump model*, in which different single hexamers twist the DNA at a distance resulting in topological strain and unwinding in the center. These models are reviewed in detail by Sclafani *et al*. [6] and are here summarized in Figure 1.6.

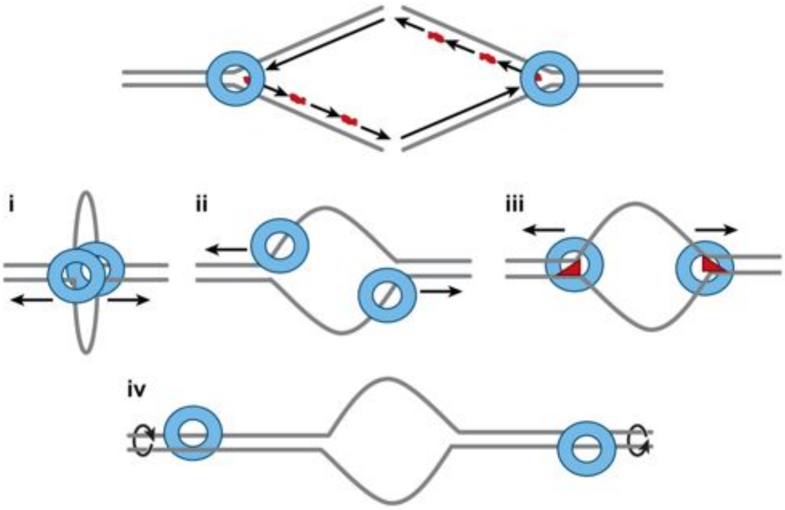


Figure 1.6 Models for MCM helicase activity. Top: single hexameric MCM is depicted as a blue ring at the ends of a conventionally drawn replication fork. Lagging strand Okazaki fragments are shown with red RNA primers at their 5’ ends. i-iv) four models proposed for MCM unwinding activity, as discussed in the main text. Adapted from: Sclafani *et al*. [6].

Clearly, what mechanism is exactly used by MCM for DNA unwinding depends on the real geometry displayed by the MCM complex when loaded on the origin. For example, both models reported in panel i) and iii) of Figure 1.6 (the SV40 T-antigen and the ploughshare models, respectively) assume that MCM is loaded as a double hexamer onto the origin. Therefore, a lot of effort has been put in trying to understand the geometry adopted by MCM, as well as by the other RC proteins recruiting MCM, when bound to replication origins. In this perspective, transmission electron microscopy applied to image purified RC components (usually from yeast, but often from *Archea* and sometimes also from human) turned out to be an extremely useful method. It was first used in a pioneering work in fission yeast where the six MCM subunits were found to display a ring-like structure [68]. Strikingly, this approach has revealed that many RC proteins share a ring-shape structural geometry, so that this can be considered as the leitmotif of the RC assembly.

The first to be recruited onto the origin is the Cdc6-ORC1 ATPase complex (Figure 1.7A). The overall shape and dimensions of the ORC–CDC6 complex match the structure of a ring, with dimension similar to a single Mcm2-7 hexamer [47]. This has suggested that CDC6, when bound to ORC (indeed, ORC1) acts as a “clamp loader” for the recruitment of MCM, i.e. it uses the energy derived from ATP hydrolysis to engage ring-shaped molecules with similar dimensions. The loading of MCM by CDC6 and ORC could be achieved by a series of conformational changes, coupled to ATP binding and hydrolysis, which open and close the MCM ring around the DNA. However, additional structures of ORC, CDC6 and MCM, ideally in complex with each other, will be required to construct a structural model for the DNA helicase loading reaction.

The electron microscopy approach has been instead recently decisive in elucidating that MCM is loaded onto origin chromatin as a head-to-head double hexamer (Figure 1.7B and [61]), capable of passive sliding along dsDNA. This actually leads to only two possible mechanisms for its helicase activity once the replisome is assembled: the SV40 T-antigen or the ploughshare model (Figure 1.6 i and iii, respectively). However, the observation that also the GINS complex has a ring-shape structure conserved up to human (Figure 1.7C and [78]) makes the latter more likely. The proposed model in this case is depicted in Figure 1.7D: MCM double hexamer exploits its passive sliding along dsDNA and acts as a dsDNA translocase, and GINS and/or Cdc45 play a direct, structural role in strand separation acting as a plough.

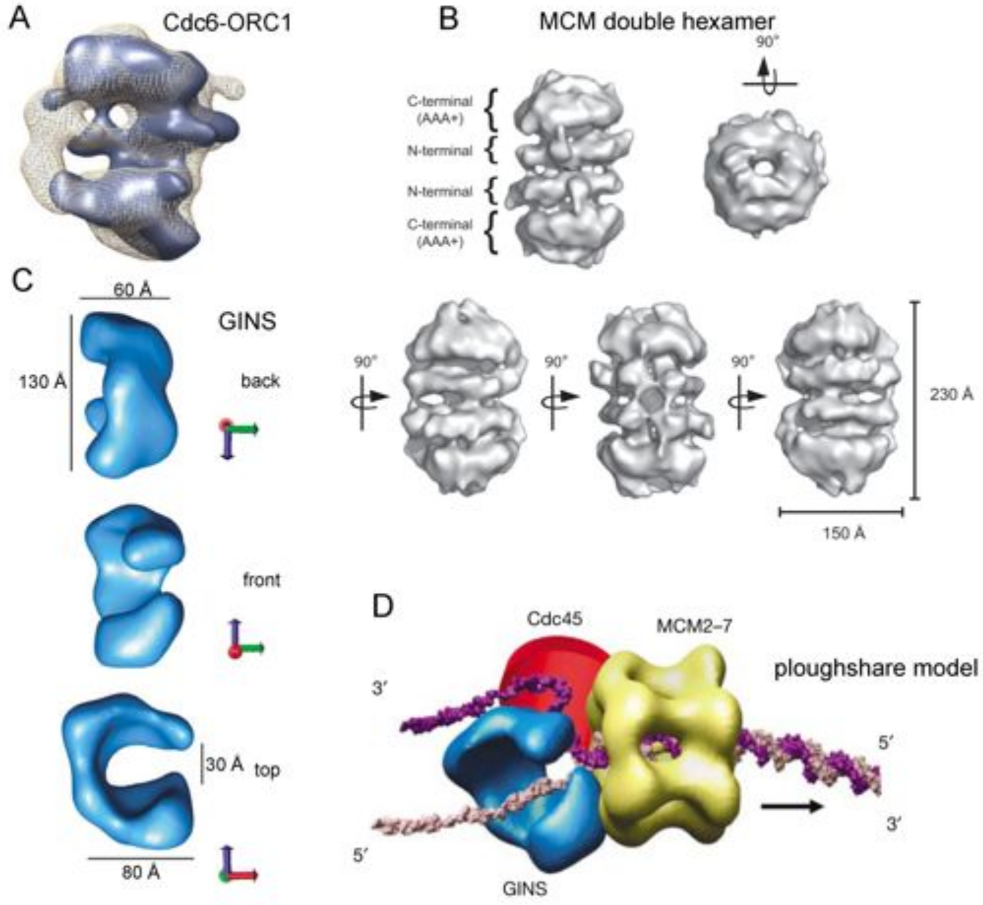

A
Cdc6-ORC1
B
MCM double hexamer
90°
C-terminal (AAA+)
N-terminal
N-terminal
C-terminal (AAA+)
C
60 Å
GINS
130 Å
back
front
30 Å
top
80 Å
90°
90°
90°
230 Å
150 Å
D
Cdc45
MCM2-7
ploughshare model
3′
5′
5′
3′
GINS

Figure 1.7 3D reconstructions of the RC proteins structure imaged by electron microscopy. A) Cdc6-ORC1 complex in yeast, taken from Speck *et al*. [47]. B) Yeast Mcm2-7 complex is loaded onto the origin as a double hexamer, taken from Remus *et al*. [61]. C) Human GINS complex, taken from

Boskovic *et al*. [78]. D) Picture of the proposed ploughshare model (same as Figure 1.6, iii), with the GINS acting as the wedge. Adapted from: Boskovic *et al*. [78].

1.1.5 The riddle of origin specification in metazoa

As already mentioned in **paragraph 1.1.1**, the selection of defined and adequately distributed replication origins seems to represent the safest way to achieve complete genome duplication in eukaryotes. Specific sites (named ARS: Autonomously Replicating Sequence), determined by precise sequence motifs were found in *S. cerevisiae* [82]. However, this revealed not to be the case for other eukaryotes, in which the sequences directing replication initiation appear to be far less defined. Extreme situations have been reported in *Drosophila* and *Xenopus* early embryos, in which replication initiation occurs at random sites along the chromosomes. Strikingly, during embryonic development, in correspondence to remodeling of nuclear structure and chromatin organization, initiation events become restricted to preferred regions [83]. The factors determining the confinement of initiation events to specific regions during embryogenesis have been in the last years intensively studied, but are still unclear [13, 17]. In agreement with the fact that preferred sites of initiation are selected during development, DNA synthesis does not start at random locations in somatic mammalian cells. Also in this context, the mechanism that governs the selection of replication origins in metazoan genomes taking place in the G1 phase of the cell-cycle, is still not clear. What makes the understanding of origin specification difficult is mainly the high degree of degeneracy of metazoan origin sequences. In human cells, only few origins are well characterized and they share no evident sequence similarity [84-86]. More recently, genome-wide approaches have led to the identification of several origins, but still no consensus has been clearly identified, besides a relative frequency of CpG islands [14] and asymmetric A/T stretches [15] in correspondence to highly active origin sequences (Figure 1.8).

On the other hand, also ORC, the protein structure that marks all replication origins and which is needed for the sequential assembly of the full RC, exhibits little sequence specificity in higher eukaryotes spanning from *Drosophila* to mammals [13, 17, 43, 87, 88], as already mentioned in **paragraph 1.1.3**.

All these observations suggest that the binding of ORC at origins might be determined not only by the recognition of specified DNA elements (*e.g*. AT-rich sequences and CpG-islands reported in Figure 1.8, as well as promoter regions, dinucleotide repeats, matrix attachment regions), but also by a series of other factors which could facilitate the ability of a DNA region to function as an origin, namely: i) the overall chromatin structure, comprising also the local DNA topology and possible epigenetic marks; ii) ORC-chaperon proteins that function as auxiliary ORC-targeting factors [31].

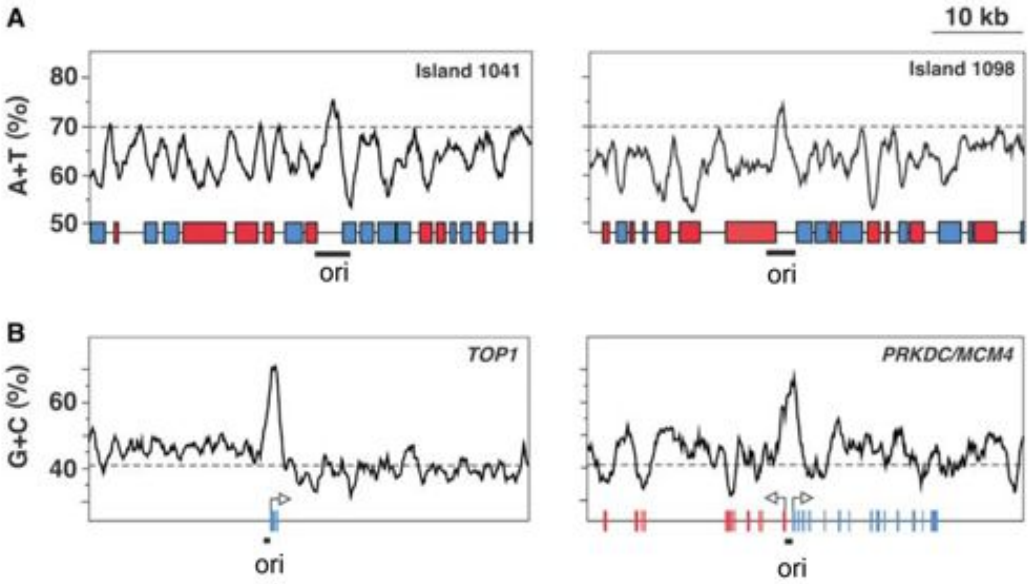


Figure 1.8 AT-rich and CpG-rich islands at DNA replication origins. A) AT content (%) across the 50 kb regions including the 1041 and 1098 AT-rich islands in *S. pombe*. Red and blue rectangles represent genes transcribed towards the left and the right, respectively. Broken lines indicate the average intergenic AT content (70%). B) CpG content (%) across the 50 kb regions spanning the first two exons of the human TOP1 gene and the bidirectionally transcribed PRKDC and MCM4 genes. Arrows indicate the direction of transcription. Red and blue bars represent exons. Broken lines indicate the human average genomic CpG content (41%). In both panels, the black bars underline the presence of an active origin in correspondence to peaks of AT% or CpG% over the average values. Scale bar: 10 kb. Adapted from: Antequera *et al.* [89].

One of the most probable candidates, which could likely contribute to origin specification, is a local chromatin environment ideally suited for the pre-RC assembly. This could result from the combination of several elements, such as an open, transcriptionally active chromatin structure, bent DNA structures, gene promoters close-by, binding sites for sequence specific proteins or asymmetric AT-rich stretches. According to this scenario, origins could take advantage of—or parasitize—regions that are maintained in an accessible conformation for structural reasons or to facilitate transcription, as suggested by the preference of origins to map near promoters in many cases [89]. This "opportunistic" origin specification would remove the selective pressure to maintain each single origin sequence in the genome for its individual contribution to replication. This model is supported by at least two considerations: i) eukaryotic origins are present in excess through the genome; not all of them fire in every S phase and many remain silent and are inactivated by replication forks passing by during S phase [90]; ii) open chromatin appears to be the underlying feature that is deterministic for ORC binding as revealed by genome-wide approaches [34]. Thus, origin specification could be achieved by a progressive restriction of the potential to initiate replication from too many or undesired sites; the driving force of this selection could be the chromatin re-organization associated with development and/or with the cell-cycle progression, particularly with the G1/S transition. These assumptions characterize a model [13, 17, 91] in which two stages need to be passed in G1 to achieve origin specification: the former is the timing decision point (TDP), where early and late replication domains are established, and the latter is the origin decision point (ODP), which selects only a fraction of the sites previously licensed to be used in the next S phase (Figure 1.9).

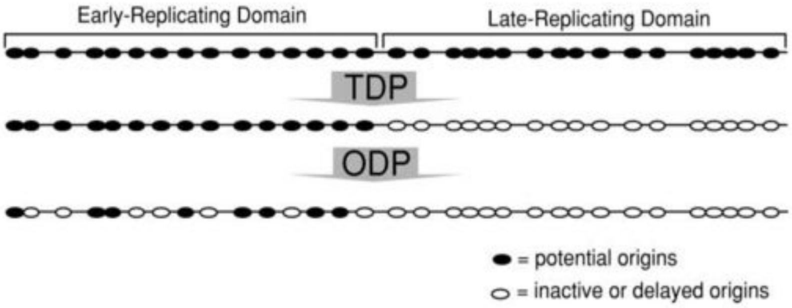


Figure 1.9 Model describing the "opportunistic" origin specification during G1 phase. In early G1 phase, many sites distributed throughout the genome have an equal potential to be used as early replication origins. At the TDP, late replicating chromosomal domains become excluded from the pool of potential early replicating origins. At this time, origins within these early replicating domains still have an equal potential for initiation regardless of their position within the domain. At the ODP, a subset of these potential origins is chosen for initiation in the upcoming S phase. Taken from: Li *et al.* [91].

It must be underlined that this view hints at a role also for DNA topology in establishing an origin of DNA replication. A particular chromatin structure is indeed maintained by a peculiar DNA conformation and topology. If this model holds true, topoisomerases, which are enzymes able to alter the topology of a DNA region, should play a determining role in origin function; in detail, topology-modifying events may be required for the formation of the pre-RC. And indeed, both topoisomerases I and II were found to interact with the lamin B2 origin, and to be essential for origin firing in close interaction with ORC [88, 92].

Chromatin accessibility, however, is unlikely to be the only requirement for origin specification, as several specific sequences ranging in size from 1 to 6 Kb have been described that are capable of maintaining their activity at ectopic positions in the genome [93]. Thus, likely origin specification relies on other factors; these could include proteins, which display a preference for certain origin features (like sequence or other structural properties) and target the RC proteins onto the origin by direct or indirect protein-protein interaction. So far, several proteins have been identified, that can either specify sites of ORC binding, or in any case have a role in DNA replication initiation. Among these it is worth to mention AIF-C [94], Trf2 [95], Ku80 [96, 97], EBNA1 [98], and HMG1a protein [99]. In many of these cases, the proposed proteins were shown to function as “ORC-chaperons” in targeting ORC to chromatin regions thus contributing to origin formation and to a more specified binding of ORC. Interestingly, also an important transcription factor (and proto-oncogene) like Myc was recently found to bind the well-characterized human replication origin sequences and participate in origin activation [100], again hinting at transcription events as actors in origin specification.

In this context, it was very interesting to find out that another family of transcription factors, namely the homeotic proteins, displays an affinity for the origin sequences [101], which seems to correlate with cell-cycle as well as DNA replication regulation [102, 103]. A possible role for homeotic proteins in origin decision would be particularly intriguing, because it could represent a basis of the interplay between DNA replication and development, as well as explain the proto-oncogenic properties often displayed by these proteins [104]. In order to better understand to what extent could homeotic proteins be involved in origin function, some knowledge about their structure and function is required; accordingly, the next paragraph will be focused on this family of transcription factors involved in the regulation of development and often in cancer.

1.2 The family of HOX proteins

Homeotic proteins are the product of the regulatory family of homeobox genes. First identified in the fruit fly *Drosophila melanogaster* [105], they are highly conserved across different species, and share a functional activity as transcription factors during normal development [106-108]. They are characterized by the presence of a highly conserved motif about 61 aminoacids long, called the homeodomain. This enables homeodomain proteins to bind DNA at specifically recognized binding sites and transcriptionally activate (or repress) their target genes, as extensively reviewed [109-111]. Some variations besides the common homeodomain sequence allow grouping them into different families. Most strikingly, however, is their seemingly universal involvement in cellular specification and identity. They all seem to be key regulators of cellular identity, from the specification of regional identity of several skeletal and neurological areas to the cooperative role in the regulation of epithelial-mesenchymal cell interactions in extra-embryonic tissues.

At present, a large number of homeobox genes have been characterized and novel isolates are still being identified [107]; up to date, the list comprises more than 350 known homeodomain sequences, making up an extensive super-family of regulatory genes, that can be grouped in at least 10 different classes, with members found in all animal species, and even in many plants. In other words, there is a large catalogue of divergent homeobox sub-families (estimated to be around 0.1-0.2% of the whole vertebrate genome [112]), of which the best characterized and most extensively studied is the *Hox* family.

As the homeotic proteins were identified quite early as regulatory transcription factors with DNA-specific recognition via the conserved homeodomain, much effort has gone into trying to elucidate the exact mechanisms by which the HOX proteins gain their functional specificity and where in the regulatory hierarchy they exert their effect. In parallel, obtained results have also strongly suggested that these proteins are key regulators of many biological processes in adult eukaryotic cells. Among these processes are cell-cell and cell-extracellular matrix interactions [113], capillary morphogenesis and angiogenesis [114], cell cycle control [115], DNA replication and cell-cycle regulation [66, 101] and cell growth and differentiation [116].

For the many attractive, interdisciplinary implications raised by the study of HOX proteins, these will be actually the focus of this section (although the most general features can be extended also to other homeobox sub-families). First, we shall present the classification of HOX proteins in vertebrates and, indeed, in mammalian species; furthermore, we will present the structural basis of DNA-HOX recognition and then switch to the structure-function related aspects of HOX activity; finally, we will present the most recent evidence that HOX proteins are involved as proto-oncogenes in cancer and that this could depend on an involvement of these proteins in the regulation of DNA replication and cell-cycle progression.

1.2.1 *Hox* genes encode HOX proteins regulating development

Due to the high number present in the vertebrate genome, homeobox genes have been organized into classes characterized by their conserved aminoacid variants in the homeodomain and other parts of the protein. Those homeobox genes which are structurally and functionally homologous to the *Antennapedia* complex (*Ant*-C) and the *Bithorax* complex (*Bx*-C) of *Drosophila* are referred to as class 1 of homeobox genes, or

Hox genes [117]. The numerous other genes encoding homeoproteins, which are found outside the *Hox* gene cluster, are referred to as divergent homeobox genes.

All vertebrate species display several *Hox* members; although the number of *Hox* genes is variable, they are always localized together in four clusters on different chromosomes in an evolutionary conserved manner, where their positioning on the chromosomal coordinate can be directly linked to their spatio-temporal expression pattern. *Hox* genes encode transcription factors that operate as transcriptional *trans*-regulators, exerting their regulatory potential from high up in the genetic hierarchy. Their target can be genes either directly involved in development ("realizators"), or rather transcription factors [118] or other signal molecules [119], which in turn may control the actual "realizators". The clustered arrangement of *Hox* genes hints towards a complex genetic unit that exerts its function as a single entity. Indeed, no one HOX protein can alone control the fate of a given downstream target gene and subsequently a particular morphogenetic pathway. HOX members rather seem to share functional complementation with other members of the *Hox* cluster [120].

In mammalian species, there are 39 *Hox* genes organized in four clusters located on four different chromosomes in the genome [121]. They are thought to have arisen through the combination of *cis*-amplification and *trans*-duplication of the *Drosophila antennapedia* and *bithorax* complexes during separate evolutionary events. In detail, it is believed that the evolutionary amplification of the *Hox* genes started with a *cis*-amplification of a primordial *Hox* gene, producing 13 members, which was followed by a *trans*-duplication of most of the *Hox* complex. The *trans*-duplication is further believed to have occurred twice, leading to the four *Hox* clusters evident in the mammalian genome [122-124].

As depicted in detail in Figure 1.10, the four different chromosomal groups are each identified by a single letter (A, B, C or D), which is followed by a number from 1 to 13, depending on their position in the 5' to 3' chromosomal coordinate. The numbering correlates with a differential morphological expression of these genes during development: those with the lowest numbers are expressed most anteriorly, with a more posterior expression corresponding to increasing numerical value. Thus, the *Hox* genes are expressed in a spatio-temporal fashion during development, where they confer segmental identity along the primary body axis, from posterior to anterior.

The figure also shows that duplication events from the homeotic complex (*Hom*-C) of *Drosophila* have had a direct consequence on the striking homology seen between the 39 *Hox* genes. The *trans*-paralogous genes (genes occupying the same relative position along the 5' to 3' chromosomal coordinate) share a higher degree of sequence similarities than the *cis*-paralogous genes (genes occupying adjacent positions on the same chromosome). It should be noted that columnar correspondence (i.e. *trans*-paralogous relationship) of genes in the four clusters are based solely on the homeodomain sequences. For example, by comparing the homeodomain peptide-sequences of HOXA1 with the *trans*-paralog HOXB1 and both its adjacent and distant *cis*-paralogs HOXA2 and HOXA13, it becomes evident that HOXA1 shares as much as 53 of the 60 amino acid residues with HOXB1 (~88% homology), but only 40 residues with HOXA2 (~67% homology), and 29 residues with HOXA13 (~48% homology). This strongly indicates a more similar expression pattern and functional complementation between the *trans*-paralogs than between the *cis*-paralogs, as it has been demonstrated for mouse *HoxA3*, *HoxB3* and *HoxD3* by Greer and colleagues [124]. They showed that the three *trans*-paralogs are capable of carrying out identical biological functions even though they do not show functional equivalence in their normal, wild type chromosomal environment. They found that in knockout mice lacking either one of the three *Hox* genes there was no obvious overlap in phenotype, but rather a unique mutated phenotype. However, by replacing any of the deleted *Hox* genes with one of its *trans*-paralogs, the normal phenotype was restored. Based on their findings, they postulated that the observed non-equivalence of the *trans*-paralogs in wild-type tissues might be attributed to quantitative differences in expression levels rather than qualitative differences.

Thus, not only do the HOX proteins show specific spatio-temporal expression patterns in the developing organisms, but also individual *Hox* genes, at least in higher animals, show a degree of functional redundancy, potentially caused by the duplication events of the *Hox* cluster. It is therefore possible that morphologic specification by HOX proteins is partially determined by the concentration of individual HOX members in a dynamic interactive play with other HOX proteins, collectively working together as a specific “blueprint”.

Hox genes are transcribed long after the cells have differentiated and gained their morphological and phenotypical traits; the expression pattern of the *Hox* cluster differs greatly between organs and tissues [125]. The actual function exerted by HOX transcription factors in the fully differentiated cells is yet largely unknown, but indicates a role for these proteins in the maintenance of cell identity. Studies on malignant transformation in mammalian systems however, elicit intriguing hypotheses with respect to HOX function and expression patterns and will be reviewed in the following **paragraphs 1.2.4** and **1.2.5**.

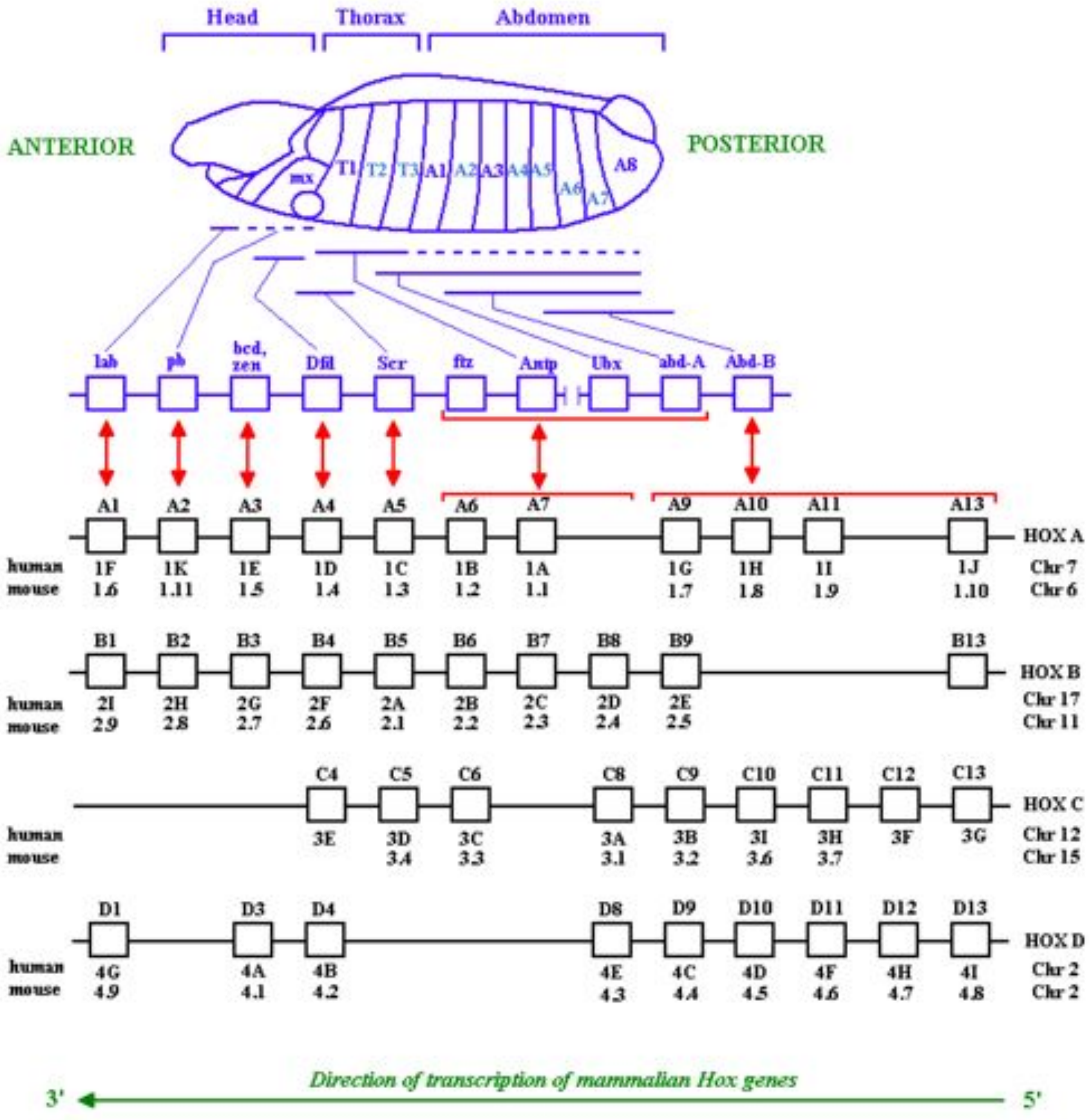

Head
Thorax
Abdomen
ANTERIOR
POSTERIOR
mx
T1 T2 T3 A1 A2 A3 A4 A5 A6 A7 A8
lab
pb
bcd, zen
Dfd
Scr
ftz
Antp
Ubx
abd-A
Abd-B
A1 A2 A3 A4 A5 A6 A7 A9 A10 A11 A13
HOX A
human 1F 1K 1E 1D 1C 1B 1A 1G 1H 1I 1J Chr 7
mouse 1.6 1.11 1.5 1.4 1.3 1.2 1.1 1.7 1.8 1.9 1.10 Chr 6
B1 B2 B3 B4 B5 B6 B7 B8 B9 B13
HOX B
human 2I 2H 2G 2F 2A 2B 2C 2D 2E Chr 17
mouse 2.9 2.8 2.7 2.6 2.1 2.2 2.3 2.4 2.5 Chr 11
C4 C5 C6 C8 C9 C10 C11 C12 C13
HOX C
human 3E 3D 3C 3A 3B 3I 3H 3F 3G Chr 12
mouse 3.4 3.3 3.1 3.2 3.6 3.7 Chr 15
D1 D3 D4 D8 D9 D10 D11 D12 D13
HOX D
human 4G 4A 4B 4E 4C 4D 4F 4H 4I Chr 2
mouse 4.9 4.1 4.2 4.3 4.4 4.5 4.6 4.7 4.8 Chr 2
Direction of transcription of mammalian Hox genes
3'
5'

Figure 1.10 *Hox* Genes: evolutionary conservation of genomic organization and expression patterns from *Drosophila* to mammals. The top of the figure (blue) depicts a 10hr *Drosophila* embryo: the horizontal bars indicate the approximate extents of the epidermal expression domains of the eight homeotic genes, whereas a dotted extension indicates posteriorly overlapping expression domains. T1, T2 and T3 depict the thoracic segments and A1-A8 the abdominal segments. The *HOM*-C cluster is divided in the middle, with two clusters (*Ant*-C and Bx-C) organized on the same chromosome. The lower part (black) depicts the mammalian *Hox* gene complex and its genomic organization (chromosomes are indicated on the right side). The official names for the *Hox* genes are those present on top of each square, whereas the most commonly used old names of the human and mice *Hox* genes are present below each corresponding square. The red arrows align the *Drosophila* and mammalian gene orthologs, also indicating corresponding expression patterns. Note also the conserved direction of transcription of the *Hox* genes, which is conserved from *Drosophila*, with the exception of *Dfd*, of which transcription occurs in opposite direction. Modified from: [110, 120, 121].

1.2.2 The homeodomain and other homeobox signatures

Although the early work of Lewis *et al.* [105] allowed for the characterization of key genetic loci controlling *Drosophila* development, including homeotic genes, the actual homeobox was discovered some years later as a highly conserved 183-bp DNA sequence, contained in all the above-mentioned loci, and encoding for a 61-aminoacids domain, termed the homeodomain [111, 126].

After the first pioneering structural studies by NMR spectroscopy [127, 128], several crystal structure studies of both free and DNA-complexed homeodomains have followed [129-135], so that by now the tertiary structure of the homeodomain is exhaustively described. It consists of a flexible N-terminal arm preceding three helical regions, where helices 2 and 3 form a helix-turn-helix motif (Figure 1.11). The relative arrangement of helices displayed by the helix-turn-helix motif seems not to be a direct requirement for DNA-binding, but actually helps to stabilize a global fold with helix 3 exposed so that it can penetrate the DNA major groove (Figure 1.12A). This is indeed known as “recognition helix”: it spans over the entire major groove, so that residues from the helical turns can reach the DNA bases; in the conventional homeodomain numbering (see *e.g.* Billeter *et al.* [136]), these are residues 47, 50, 51, and 54, reported also in panel B of the same figure. On the other side, conserved arginine residues 3 and 5 of the N-terminal flexible arm preceding the three helices of the homeodomain form additional contacts with the DNA bases, in this case by inserting into the minor groove. Several studies elucidating DNA motifs recognized by the HOX proteins have by now identified a canonical ($T_1A_2A_3T_4X_5X_6$) consensus sequence as the core of HOX-DNA binding. In detail, residues 3 and 5 of the homeodomain N-terminal arm contact the bases in positions 1 and 2 at the minor groove, and amino acid residues 47 and 51 of the homeodomain helix 3 contact the bases in positions 3 and 4 at the major groove (Figure 1.12B). Interestingly, 47 and 51 are conserved isoleucine (or other hydrophobic residue) and asparagine, respectively; similarly, residues 3 and 5 are conserved arginine residues; conversely, residues 50 and 54 are less conserved and this is supposed to serve to differentiate the binding properties of different homeodomains, by recognition of the less conserved DNA bases at position 5 and 6 [137].

Overall, these protein-DNA contacts are referred to “specific” contacts, i.e. interactions of the protein with DNA bases, which are responsible for the selection of a certain DNA sequence to be bound. Other “unspecific” contacts are displayed by the homeodomain with DNA backbone, and mostly serve to increase the stability of the complex. Many of these interactions involve basic residues that contact DNA phosphate groups, like conserved arginines at positions 31, 52, 53, or lysine 55.

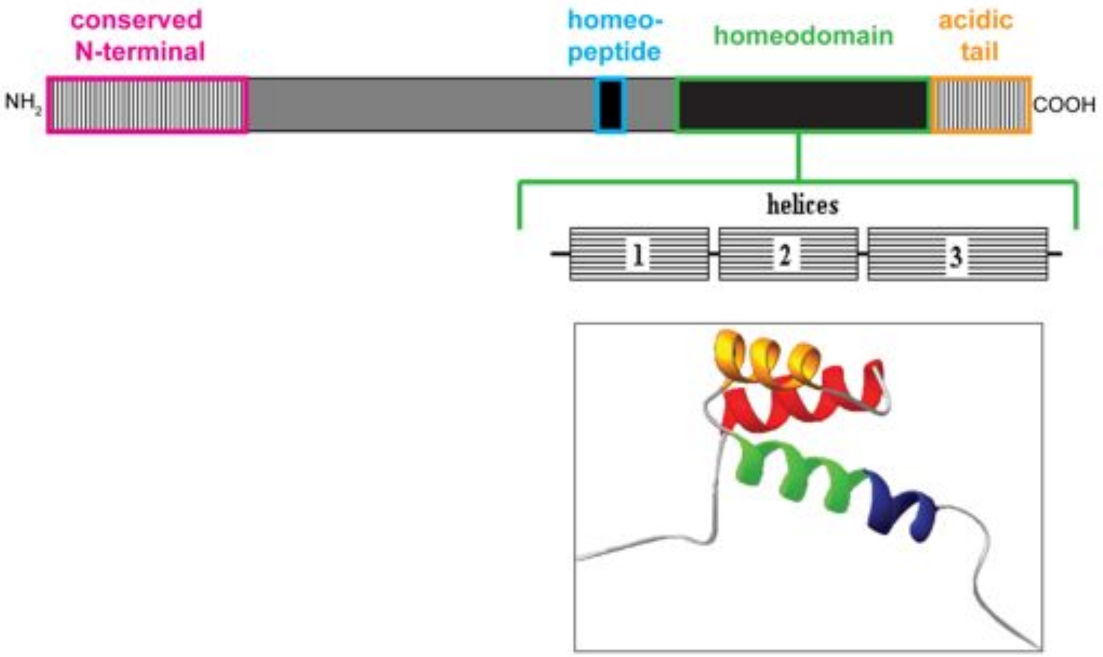


Figure 1.11 The homeodomain tertiary structure. Schematic representation of a prototypic homeodomain protein. The homeodomain is further highlighted with the three alpha helices. Taken from: Castronovo *et al*. [138]. In the inset, the structure of the homeodomain from the *Drosophila* Antennapedia protein is taken as a reference [127, 128] for the visualization of the helix-turn-helix structural motif. The Antennapedia homeodomain is 68 amino acid residues long. It contains three alpha helices: Helix 1 (res.10-21) = Red, helix 2 (res.28-38) = Yellow, helix 3 (res.42-52) = Green. The C-terminal part of helix 3 (res.53-59) is depicted in blue, because it exhibits lower stability in free homeodomains and can sometimes be reported as an actual fourth helix. Helices 2 and 3 adopt a helix-turn-helix motif from residues 28 to 52. The three helices making up the *Drosophila* Antennapedia homeodomain are prototypic for all homeobox proteins, serving as the core structure even for divergent homeobox families.

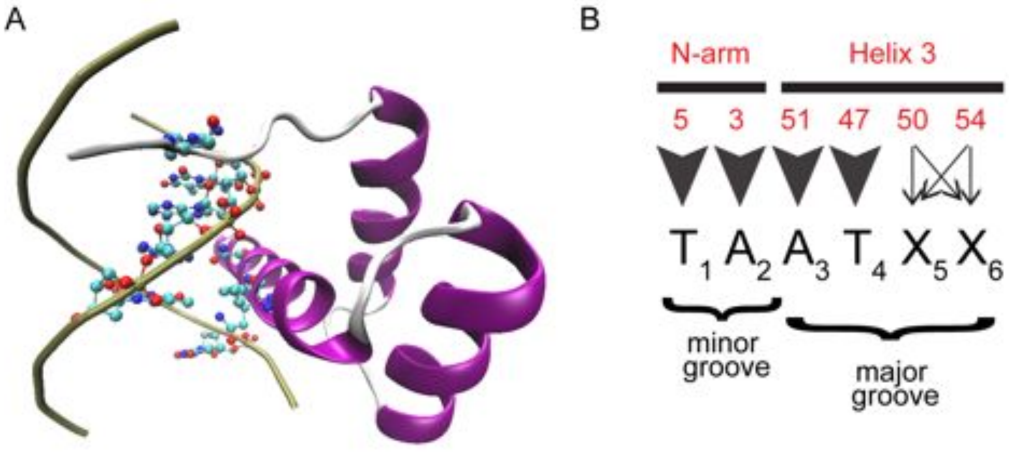


Figure 1.12 Homeodomain-DNA interactions between HOX proteins and the canonical TAAT core. A) The co-crystal structure of the *Drosophila* engrailed homeodomain with DNA [130] (PDB no.: 1HDD) was taken as a reference and visualized using VMD software (a courtesy of Fabio Trovato, Scuola Normale Superiore). The homeodomain contacts DNA both at the major groove by helix 3 and at

the minor groove by its N-terminal flexible arm. B) The base-specific contacts made between the homeodomain amino acid residues 3, 5, 47 and 51 and the TAAT binding motif. Residues 3 and 5 of the homeodomain make contacts in the minor groove, whereas amino acids 47 and 51 make contacts to the DNA in the major groove. Residues 50 mainly, but also 54 and possibly others docking in the major groove, are further thought to make more stable contacts and thus ultimately impose a higher specificity. Nucleotides 5 and 6 are also thought to make specific contact with the homeodomain, as discussed in the main text. Modified from: [139].

In addition to the homeodomain, the homeobox genes also encode other highly conserved structures (reported, as well, in Figure 1.11). The homeodomain itself ends with an acidic tail, with a yet undetermined structure. Furthermore, many homeodomain proteins, albeit not all, contain a highly conserved hexapeptide (consensus: hydrophobic-Y/F-P-W-M-K/R), named homeopeptide, just upstream of the homeodomain, to which it is linked by a linker that varies in length and sequence among different HOX proteins and species. The homeopeptide has been reported to be essential for the HOX proteins to achieve cooperative binding with HOX cofactors [140] (see *paragraph 1.2.3.1*). Finally, homeobox proteins contain a conserved amino-terminal region followed by a variable region, usually rich in alanine, serine, glycine, proline or glutamine [138]. All proteins that contain a homeodomain and, possibly, other homeobox signatures, usually play a role in development, and many of these are sequence-specific transcription factors.

1.2.3 The "HOX paradox": transcriptional and functional specificity

Hox genes encode transcription factors with a canonical homeodomain-structure responsible for DNA binding, with subsequent transcriptional activation or repression of the target gene. The recurrence of the homeodomain as a ubiquitous DNA binding domain found in all metazoans, coupled with the wide diversity in binding sites observed for these proteins, implies that this fold is highly versatile with respect to the possible target DNA sequences.

In general, when thinking about what a HOX binding site may look like, it is useful to distinguish between three types of HOX target genes: 1) those that must be highly specific for one HOX paralog ("paralog-specific"), 2) those that are shared by a subset of HOX proteins ("semi-paralog-specific"), and 3) those that are regulated by most or all *Hox* genes ("general"). In any case, the conserved TAAT motif recognized by the homeodomain *in vitro* (see Figure 1.12B) is by itself too short to account for the wide range of transcriptional specificity of the HOX transcription factors. Indeed, this four base-pair motif is potentially present in a very large number of promoter elements throughout the genome. For example, there are 87,307 copies of the sequence TAATTA and 86,201 copies of the sequence TAATGA in the *Drosophila* genome. Moreover, as elegantly illustrated by the homeodomain-binding site survey studies (reviewed by Mann et al. [141]), TAAT[t/ g][a/g] is readily bound by most HOX homeodomains, as well as many non-HOX homeodomains. It is therefore clear that the large number of known HOX proteins definitely needs further regulatory principles if the functional specificity of individual HOX proteins is to be accounted for. These considerations, that DNA binding by an isolated HOX homeodomain does not provide the specificity necessary for context-specific gene regulation *in vivo*, constitute a conundrum termed the “HOX paradox”.

Several mechanisms have been reported, that could explain how selective DNA binding is achieved *in vivo* by HOX proteins, including: i) an intricate network of water-mediated interactions [136]; ii) different flanking base pairs present both at the 5’ and 3’ of the canonical ATTA motif [142]; iii) HOX protein post-translational modifications, such as phosphorylation [143, 144]; iv) cooperativity with other trascription factors often bearing homeodomains, named HOX cofactors [140, 141]; v) sequence-dependent alterations of the DNA geometry and shape [145, 146]. The last two points of the list deserve a few more comments as they have recently shown to be the most crucial factors influencing HOX specificity.

1.2.3.1 HOX cofactors and collaborators

One critical finding explaining how HOX transcription factors could gain further specificity in their *in vivo* environment, was the identification of HOX cofactors. To date, the best-characterized cofactors are all TALE (Three Aminoacid Loop Extension) homeodomain proteins [147, 148]. In *Drosophila*, the known TALE HOX-cofactors are Extradenticle (Exd) and Homothorax (Hth). In the mouse, there are four Exd-related proteins (Pbx1, Pbx2, Pbx3, Pbx4) and five Hth-related proteins (Meis1, Meis2, Meis3, Prep1, and Prep2). These proteins all have the ability, at least on some DNA sequences, to bind with HOX proteins in a highly cooperative manner (reviewed by Mann *et al.* [141]). Variations to the TAAT motif have been characterized, with various HOX transcription factors binding to ATTA, TGAT, TTAT and TTAC sequences in conjunction with other cofactors [149, 150]. Several co-crystal structures of HOX proteins and cofactors bound to DNA have been reported [140, 141, 151]. These have revealed that the homeobox hexapeptide (Figure 1.11) is critical for the HOX-cofactor interaction, as one of its hydrophobic residues (usually a tryptophan) makes direct contacts with a hydrophobic pocket of the TALE motif in the cofactor homeodomains. For those HOX proteins that do not have an obvious homeopeptide motif, in particular the Abd-B paralogs, there is a conserved tryptophan residue that, at least for a subset of Abd-B paralogs, plays an important role in this protein-protein interaction [149, 150].

However, increasing evidence suggests the more general idea that cofactors may have modes of interaction with other HOX proteins that are in addition to the classical homeopeptide-TALE interaction, and this could potentially expand their ability to recruit additional transcription factors at the target gene. Indeed, signal effector transcription factors are being identified as HOX collaborators in several *cis*-regulatory elements, such as SMADs, effectors of the TGF-beta and Decapentaplegic (Dpp) pathways in both vertebrate and *Drosophila* (reviewed by Mann *et al*. [141]). The HOX-SMAD interactions may be critical for building an enhanceosome-like structure on HOX-targeted *cis*-regulatory elements. Although the number of examples shown to have direct inputs by signaling effectors is currently low, genetic analyses suggest that this phenomenon is likely to be extended to other signaling pathways, including Hedgehog (Hh), Wnts, and Notch [141]. Overall these considerations allow drawing a two-step mechanism by which HOX proteins can gain specific transcriptional activity (depicted in Figure 1.13). The first step is DNA binding by HOX proteins: cofactors selectively target specific HOX proteins, with a subsequent selective DNA binding due to the cooperative HOX-cofactor action. The second step involves the recruitment of additional factors, generally named HOX collaborators, to the *cis*-regulatory element. The recruitment of these factors may depend on contacts between them and the DNA and/or protein-protein contacts between them and the HOX-cofactor complex. It is the recruitment of these collaborators that ultimately determines the sign of the transcriptional regulation.

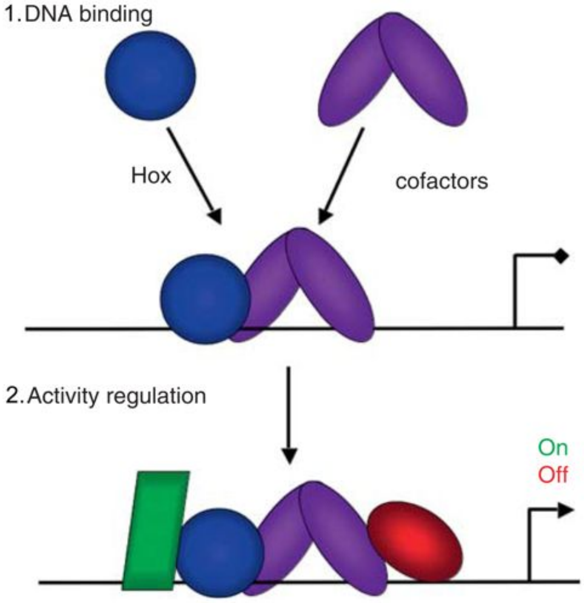


Figure 1.13 Model of the two-step mechanism of HOX specificity. Details of the model are reported in the main text. Adapted from: Mann *et al.* [141].

Yet, despite the flexibility of this model, it is critical to stress that the HOX protein plays the central role in the function (and/or the assembly) of these multiprotein complexes, because without them, these complexes cannot function. Therefore, these multiprotein complexes containing the HOX proteins, their cofactors, and their collaborators have been named “Hoxasome”.

1.2.3.2 Role of DNA shape in HOX specific DNA recognition

In many protein-DNA complexes, the DNA assumes conformations that deviate from the structure of an ideal B-form double helix, in some cases bending, and in other undergoing large conformational changes in order to optimize the protein–DNA interface. These recognition mechanisms, relying on the propensity of a given sequence to assume a conformation that facilitates its binding to a particular protein, are often referred to with the term “indirect readout” [152]. A new mechanism of indirect readout has been recently proposed to explain the DNA binding specificity of HOX proteins by Joshi *et al*. [146]. The authors observed that only residues of the homeodomain N-terminal arm and of the linker just upstream of it are critical for the specific *in vitro* and *in vivo* recognition properties of the *Drosophila* Src homeodomain variant. They further notice that these residues insert into an unusually narrow region of the minor groove, which in turn creates a local dip in electrostatic potential. Therefore they propose a two-step model to explain how HOX protein realize a specific DNA binding, as depicted in Figure 1.14: first, HOX proteins recognize generic AT-rich sequences through the conserved interactions occurring both at the major and minor grooves (see Figure 1.12); but it is only the N-terminal arm and the upstream linker residues, likely helped by the presence of a cofactor, which select among these sites by reading the structure and electrostatic potential of the minor groove. Interestingly, two years later the same group reported that the creation of specific binding sites for positively charged aminoacids, typically arginines, within narrow minor grooves, is a general mechanism widely used in protein-DNA recognition [152].

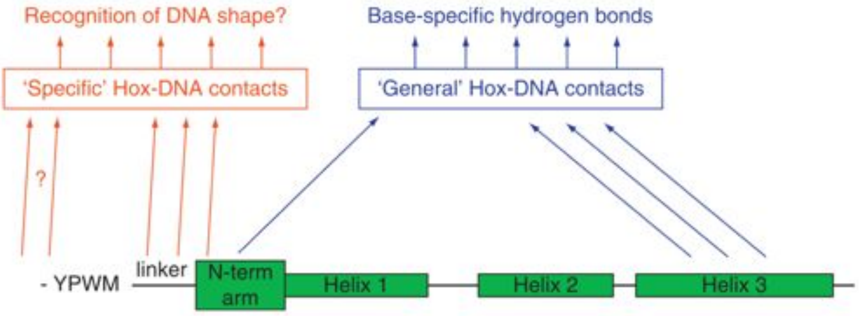

Recognition of DNA shape?
Base-specific hydrogen bonds
'Specific' Hox-DNA contacts
'General' Hox-DNA contacts
?
- YPWM
linker
N-term arm
Helix 1
Helix 2
Helix 3

Figure 1.14 Model of the two levels of HOX-DNA binding specificity. DNA contacts made by residues of the N-terminal arm and of the third helix (blue) are used by all HOX proteins to bind general AT-rich DNA sequences, but are not good at distinguishing between Hox-paralogs specificities. With the help of

cofactors, paralog-specific DNA contacts are mediated by the homeopeptide, the linker and N-terminal arm residues. These "Paralog-specific" DNA contacts may read a DNA structure, such as the narrowing of the minor groove, rather than forming specific hydrogen bonds. Taken from: Mann *et al.* [141].

1.2.4 HOX proteins and cancer

The proposed involvement of homeobox genes in cancer is fairly recent, when compared to the first studies of the homeotic proteins, and was first suggested after studies on mouse myeloid leukemia cell lines in early 1990s. This first hypothesis proposed that overexpression of a homeobox gene could lead to leukemia, as reviewed by Lawrence *et al.* [153]. Following these findings, the homeobox genes became a focal point for several research groups trying to elucidate the genetic control of carcinogenesis.

The molecular mechanisms involved in carcinogenesis deal with the acquisition, regulation and maintenance of cell identity; in this context, the homeodomain-containing transcription factors represent potential candidates as oncogenic regulators, given their functional role in embryogenesis and normal development. Indeed, since the homeobox proteins are involved in the physiological processes of cell regulation during early development, it may be postulated that a deregulation of these genes in mature cells may lead to altered cell identities and possibly trigger the onset of cancerogenesis [154, 155].

To date, several *Hox* genes have been isolated from selected malignant tumours, among them: HOXA9, HOXB7 and HOXD11 have all been found to be misexpressed in primary colon cancer [117]; a frequent overexpression of HOXA7, HOXB7 and HOXA10 was reported in human lung cancer [156]; HOXD9 showed an ectopic expression in cervical cancer [157]; and several HOX members were found misexpressed of in human anaplastic thyroid cancer, including the silencing of HOXD9 in all tested cancer cell lines compared to its expression in normal thyroid [125]. Notably, in T-cell malignancies as acute myeloid leukemia, the misexpression of the *Hox* genes often occurs by their rearrangement with the *NUP98* locus to form a chimeric *NUP98/Hox* gene. This is the case for at least seven *Hox* genes: HOXA9, HOXA11, HOXA13, HOXD11, HOXD13, HOXC11, and HOXC13 [158, 159]. The mechanism of translocation to nucleoporin NUP98 has been reported for other proto-oncogenes, such as topoisomerase I and topoisomerase II [160, 161].

Thus, a significant change in the expression pattern of the *Hox* cluster, intended either as overexpression or as lack of expression, is observed in malignant phenotypes. Whether or not this misexpression is a cause or an effect of neoplastic transformations remains to be elucidated, although some data are strongly suggestive that both might be the cases. In recent years, many studies have been carried out to elucidate what is the role of *Hox* genes in cancer development. This task has proven to be far more difficult than expected, due mainly to the fact that homeobox genes typically are expressed at low levels and that several different homeobox genes seem to cooperate in specific combinations in the different cellular pathways within the cell [138]. Some studies have produced promising results however, giving some hope for the characterization of the key regulators responsible for cancerogenesis. Among them these should be mentioned: i) HOXB7 directly *trans*-activates the basic fibroblast growth factor (*bFGF*) gene in melanomas, thus possibly deregulating essential growth factor pathways in the tumour cells [113]; ii) HOXD3 overexpression mediates the conversion of vascular endothelial cells from the resting to the angiogenic/invasive state [162]; iii) a lack of expression of HOXA5 is implicated as a cause for the loss of expression of p53 in human breast cancer, which could enable human breast tumour cells to evade p53 induced apoptosis [163, 164]; iv) human SIX1 (a divergent homeodomain protein) overexpression is associated with abrogation of the G2 cell-cycle checkpoint in breast cancer [115].

It should be noticed that two of the above mentioned tumorigenic mechanisms deal with a possible role of the homeotic proteins in the regulation of cell proliferation. Interestingly, some parallel studies conducted in the field of DNA replication revealed an affinity of several HOX proteins for human DNA replication origins [101]. In view of a possible connection between key elements of development and differentiation with the main events regulating cell proliferation, we will now review the state-of-the-art about the possible involvement of HOX proteins in cell-cycle regulation and DNA replication origin function.

1.2.5 HOX proteins vs. DNA replication and cell-cycle regulation

A large body of evidence points to HOX proteins as actors in the regulation of cell-cycle progression, and in particular of DNA replication occurring at the S phase of the cell-cycle. As these two events are tightly interconnected, this paragraph actually summarizes, with chronological order, the most relevant findings concerning the both of them.

The search for cell-cycle regulatory and regulated genes in human carcinoma cells has allowed identifying several homeotic proteins. This was indeed the first indication that homeobox genes play a role in tumorigenesis and/or tumor progression, possibly through a cell-cycle regulative function. One of the first to be identified was HSIX1, a member of divergent homeobox genes (see **paragraph 1.2.1**). HSIX1 expression was found to be absent at the onset of and increased toward the end of S phase. Overexpression of HSIX1 in MCF7 cells resulted in abrogation of the G2 cell-cycle checkpoint in response to X-ray irradiation. HSIX1 expression was absent or very low in normal mammary tissue, but was high in 44% of primary breast cancers and 90% of metastatic lesions. In addition, HSIX1 was expressed in a variety of cancer cell lines [115].

Parallel studies focused on the search for proteins participating in the regulation of human DNA replication, identified the HOX proteins to specifically bind the sequence of human replication origins. The first indication came from a yeast mono-hybrid screen for human proteins with affinity for the lamin B2 origin [101]. This study identified three HOX proteins, namely HOXA13, HOXC10 and HOXC13, for the specific recognition of a central portion of the origin. This 74bp sequence already showed to be bound by the replicative complex (RC) [84, 165]. The HOXC10 and HOXC13 proteins were shown to bind the same origin sequence both *in vivo* (CAT assay) and *in vitro*. Further studies revealed that the HOXC10 protein is present only in proliferating cells, where it oscillates in abundance during the cell-cycle until it is degraded early in mitosis by the ubiquitin pathway [102]. Mutations of two destruction boxes avoid its degradation and delay the metaphase to anaphase transition, indicating that the protein is indeed involved in mitotic progression.

Soon after, the cell-cycle regulator Geminin, an origin licensing inhibitor [24], was found to associate with members of the *Hox*-repressing polycomb complex, with the chromatin of *Hox* regulatory DNA elements and with a number of HOX proteins. The interaction between Geminin and HOX proteins prevents the latter from binding to DNA, inhibits HOX-dependent transcriptional activation and displaces Cdt1 from its complex with Geminin. Thus, by establishing competitive regulation, Geminin was proposed to function as a coordinator of developmental and proliferative control [66], as schematically depicted in Figure 1.15.

More recently, other *abdominal-B* orthologs, namely HOXD13, HOXA13 and HOXD11 were found to interact *in vivo* with several origins (lamin B2, *TOP1, MCM4, c-MYC* and *FMR1*). In particular, HOXD13 was found to stimulate RC assembly and DNA synthesis; Geminin, which interacts with HOXD13 as well, blocks HOXD13-mediated assembly of pre-RC proteins and inhibits HOXD13-induced DNA replication. Thus, the HOX-Geminin competitive regulation works also in the context of replication origin licensing [103].

Overall, these studies suggest that HOX poteins do exert a role in the control of cell proliferation. In particular, data point to a direct intervention of homeotic proteins in origin regulation, with no mediation by transcription, previously considered as the only way through which HOX proteins exert their function. Moreover, the affinity for origin sequences displayed by HOX proteins *per se* suggests that they could be directly involved in origin function, besides via the competitive regulation of Cdt1-Geminin binding. An involvement in the regulation of origin activation of these proteins could finally help to explain the proto-oncogenic role often displayed by these proteins (see previous paragraph and [104]). Still, such considerations would need a thorough understanding of the origin activation/deactivation cycle to be achieved.

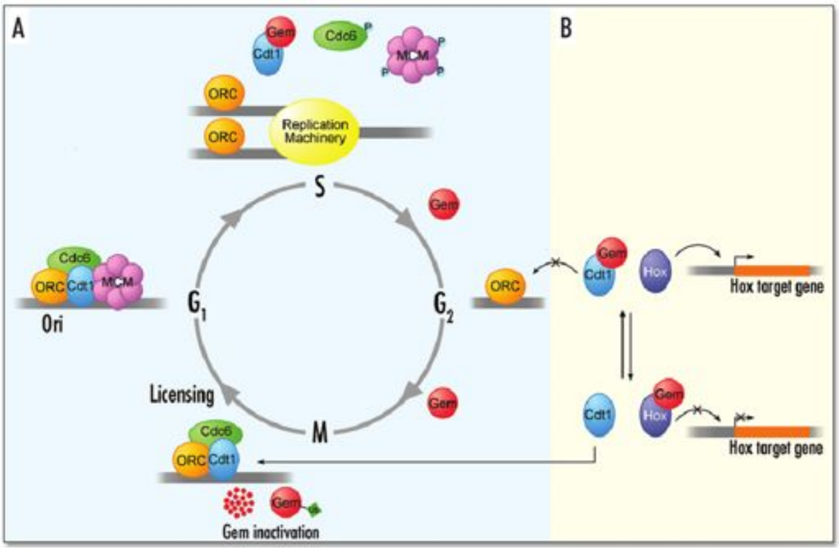


Figure 1.15 Model of Geminin interplay between proliferation and differentiation. A) Cdt1-dependent role of Geminin in cell-cycle regulation. B) HOX-Cdt1 competition for the binding of Geminin. Details are reported in the main text. Taken from: Luo *et al*. [166].

We chose to address the issue of HOX proteins involvement in origin function by exploring the spatio-temporal dynamics of a selected HOX protein in relation to DNA replication in living human cells. This approach took advantage of the use of several biophysical tools, which have so far proven helpful in the study of many biology fields, included DNA replication, as will be reviewed in the next paragraph.

1.3 Biophysical approaches for the *in vivo* study of nuclear structure and function

Recent advances in biophysical technologies combined to confocal or wide-field microscopy provide ideal strategies for approaching problems of various fields of biology. These advances were definitely favored by the discovery of the Green Fluorescent Protein (GFP), and by the subsequent development of numerous fluorescent protein (FP) derivatives spanning almost the whole UV-visible spectrum [167-172]. These actually constitute the unique category of genetically encoded fluorophores; since their introduction, the ability to study protein functions in an intact cellular environment has been revolutionized, in that fluorescent labeling is achieved non-invasively by the expression of the protein of interest fused to the sequence coding for the desired FP. It is therefore not surprising that the pioneers of this tool were awarded with the 2008 Nobel Prize in Chemistry.

The development of FPs has led to the rapid evolution of live-cell imaging methods. The full spectrum of FP color variants is being exploited in multicolor fluorescence microscopy experiments to track the distribution of different proteins in the same living cells, allowing for the direct visualization of protein recruitment and/or interactions, co-localization, and dynamics within subcellular compartments [173-175]. Through the combination of FPs and advanced digital imaging technologies, it is now possible to visualize diverse biological processes inside the living cell, providing an important complement to the biochemical methods, traditionally used as an exclusive tool in the field. Importantly, the availability of quantitative imaging approaches allows to extract the correct information from large digital imaging data sets; this is often the recommended way to gain reliable results in the typical context of cell-to-cell heterogeneity, which would make any qualitative analysis based on selected images difficult, if not impossible [176].

The mammalian cell nucleus represents a tempting subcellular environment where to test the potential advantages brought by biophysical approaches. It contains a variety of subnuclear domains where proteins with specialized functions are localized. These domains include spherical bodies (such as nucleoli, Cajal bodies, and the promyelocytic-leukemia-protein nuclear bodies or PML), irregular speckles (such as those of the subunits of the mRNA splicing machinery), as well as the diffuse structure of interphasic chromatin. The partitioning of the nucleus without intervening membranes indicates that subnuclear compartments likely formed by a process of self-assembly. The mechanisms that regulate and maintain these highly ordered assembly of proteins in intact nuclei must be defined to understand their function. Therefore, the structure and function of nuclear architecture has been in the last years extensively studied (for representative reviews see: [177-181]). Such studies have confirmed that biochemical approaches are invaluable in the characterization of the multi-protein complexes involved in the maintenance of nuclear structures. However, what is missed by the *in vitro* analysis is the role that the organized microenvironment within the intact nucleus plays in the regulation of these complexes. In this context, the visualization of nuclear compartments by both indirect immunofluorescence microscopy and by labeling of the constituent proteins with FPs has been particularly advantageous (reviewed by Voss *et al.* [176]). The latter approach in particular has allowed applying quantitative biophysical tools to explore nuclear dynamics with high spatial and temporal resolution. Among the most widely used biophysical techniques are FRAP and FRET microscopy, whose application in the nuclear context will be reviewed in the two next paragraphs.

1.3.1 FRAP and FRET applied to the study of the nucleus

Fluorescence Recovery After Photobleaching (FRAP) is a technique whereby the protein of interest is labeled by a fluorophore, *e.g.* it is fused to a FP and expressed in living cells; thereafter, it is photobleached in a region of interest (ROI) and the redistribution of the non-bleached protein population is monitored. From the plot of the recovery of fluorescence in the ROI vs. time important information about the mobility and diffusive properties of the protein can be derived. This procedure has been extensively used to analyze the dynamics of numerous nuclear proteins with particular regard to the kinetics of chromatin binding [182-188] (Figure 1.16). The quantitative analysis of nuclear FRAP experiments (reviewed in Appendix A of this thesis) has allowed to draw a dynamic vision of nuclear compartmentalization: nuclear proteins belonging to a particular subnuclear domain display a high tendency to actively exchange with other subnuclear domains. For transcription factors, this results in a partition of their distribution between nucleoplasm and chromatin, and in mean residence times onto chromatin of the order of seconds [182, 187-189]. This fast nuclear dynamics is supposed to be a means by which nuclear functions like gene expression and chromatin modification are regulated. Conversely, few nuclear proteins deviate from this rule and display stable chromatin binding and slow intranuclear exchange, namely histones [190], cohesin [191] and centromeric proteins [192]. Notably, these proteins share a structural role in the maintenance of nuclear architecture, which probably does not need the same plasticity as other nuclear functions for its regulation.

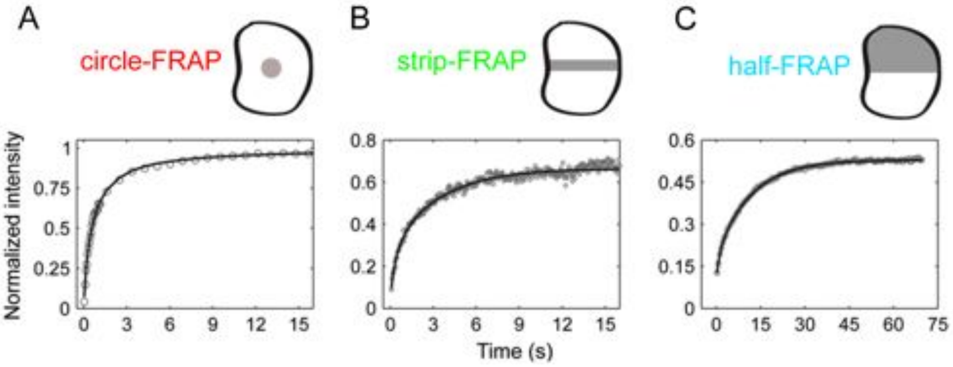


Figure 1.16 Nuclear FRAP procedures. Nuclear FRAP procedures differ mainly in the geometry of the ROI which is bleached (grey area within the schematically depicted nucleus). A) Bleaching of a small

circle inside the nucleus; B) Bleaching of a strip across the nucleus; C) Bleaching of half nucleus. Representative plots of fluorescence recovery in the ROI vs. time are reported below each bleached nucleus. Adapted from: Mueller *et al*. [182].

The formation of domains with separate functions in the nucleus requires the assembly of proteins which interact with each other. Thus, subnuclear organization likely reflects direct protein-protein interactions, but the optical resolution of the light microscope is not sufficient to detect this (Figure 1.17A). Fortunately, there are imaging techniques available that allow us to further define the spatial relationships between specific nuclear protein partners in living cells (reviewed by Voss *et al*. [176]). Fluorescence Resonance Energy Transfer (FRET) is used for the analysis *in vivo* of protein-protein interactions. A close proximity between two proteins of interest, i.e. a direct interaction between them, can be detected by their fusion to an appropriate couple of FP mutants with peculiar spectral properties (reviewed in Appendix B of this thesis), named donor and acceptor (Figure 1.17B). Because the distance within which energy transfer can occur between FPs is limited to less than about 8-10 nm, the detection of FRET provides measurements of the spatial relationship of the fluorophores on the scale of angstroms. There are several methods to detect FRET between two fluorolabelled interacting species. A very advantageous one is by detecting the reduction of donor lifetime in the presence of acceptor (graph of Figure 1.17B), as can be achieved by Fluorescence Lifetime Imaging Microscopy (FLIM) (see Appendix B of this thesis and [193]).

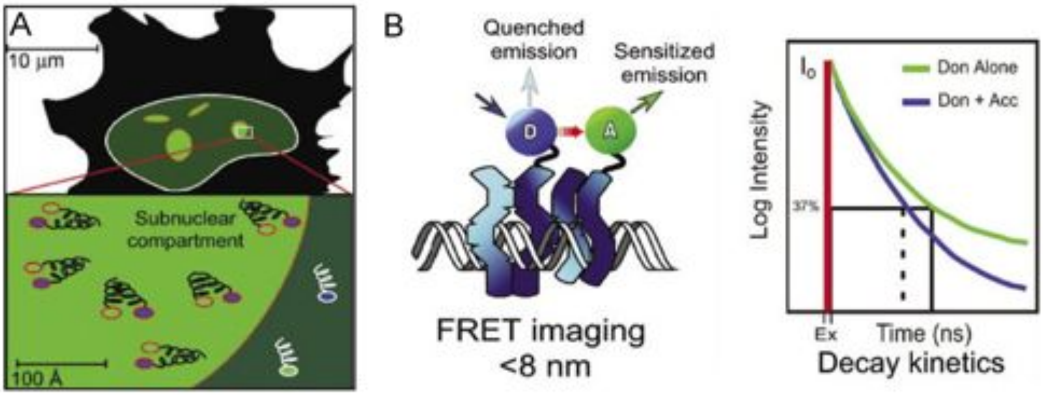


Figure 1.17 FRET to study mammalian cell nuclei. A) Fluorescence microscopy is limited by the diffraction of light to a resolution of approximately 200 nm, and objects that are closer will appear as a single object, so that considerable distances may actually separate proteins that appear co-localized by fluorescence microscopy. B) FRET microscopy detects the direct transfer of excitation energy (red arrow) from a donor *D* fluorophore to an acceptor *A*, when they are closer than 8-10 nm. When energy transfer occurs, the donor fluorescence signal is quenched. Right graph: FRET results in a shortening of the donor lifetime, i.e. the mean time that the fluorophore spends in the excited state prior to returning to the ground

state. FLIM (Fluorescence Lifetime Imaging Microscopy) allows the spatially-resolved quantification of FRET by donor lifetime measurements. Adapted from: Voss *et al.* [176].

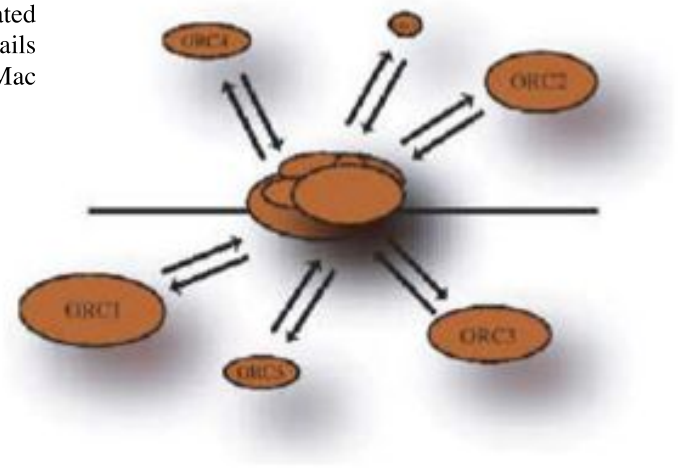
1.3.2 Study of DNA replication regulation in living cells

The mechanism of DNA replication, as summarized in **paragraph 1.1**, has also been studied in living cells by biophysical approaches, in order to elucidate both the dynamics of proteins involved in the DNA replication process, and the protein networks relevant for its regulation.

The *in vivo* study of DNA synthesis has unveiled interesting insights about how DNA replication is spatially organized in intact nuclei; furthermore, it has helped the understanding of how time-programmed origin activation (**paragraph 1.1.2**) is achieved throughout S phase progression. Replication sites were first detected by immunostaining of newly replicated DNA or replication enzymes [194] and later by monitoring GFP-tagged replication factors, like PCNA [195], Dnmt1 [196], RPA [28], DNA Ligase I [197]. To date, such fluorolabelled proteins constitute a very useful class of *in vivo* markers of S phase progression [197]. These approaches revealed that replication foci appear under the microscope as nuclear spots of varying size and position, yet their number per nucleus is far smaller than the expected number of genomic replication origins. The pattern of tagged replication loci changes significantly during S phase, consistently with the idea that replication origins self-assemble in replication foci [198]. Yet, the self-assembly relies on strong protein-protein interactions in the replication machinery, as FRAP experiments performed using FP fusions of PCNA have shown little if any turnover of the protein at replication sites [28, 199]. Collectively, these data have led to a model in which replication origins are clustered together at the start of replication to form several replication factories dispersed within the nuclear architecture. Unreplicated DNA is pulled into the factory, whereas replicated sister strands might be extruded in the course of replication. Interestingly, a spatio-temporal order is maintained in the activation of the replication factories in the course of S phase. This represents the way by which DNA replication timing regulation is exerted [200]. Indeed, replicon clusters that fire at different times during S phase occupy different subnuclear compartments, with early replicating foci showing enrichment in the nuclear interior, whereas foci replicating later during S phase are enriched in perinucleolar regions and the nuclear periphery [194, 201, 202] (see also Figure 2.7 at next Chapter, **paragraph 2.3**). The order in which these segments replicate is established during early G1 phase [200], and seems to be an epigenetic mark depending on the differentiation state of a cell [33].

The assembly of pre-RC occurring in G1 phase has also been investigated in living cells. In an interesting study by Mc Nairn and colleagues [203], GFP fusions of ORC1 and ORC4 subunits have been expressed in mammalian cells and their nuclear dynamics investigated by FLIP (Fluorescence Loss in Photobleaching). In FLIP, a single spot in the nucleus is repeatedly bleached and the loss of fluorescence is quantified at locations away from the bleach spot, so this represents a complementary technique with respect to FRAP, but it actually provides similar information [204]. The authors observe very fast dynamics for both proteins, thus providing evidence that ORC can sample many sites throughout the nucleus within a very short period of time (order of seconds). They comment that this dynamic behavior could account for the many origin sequences that mammalian ORC is supposed to bind as well as for the many biological functions in which ORC has found to be involved, besides pre-RC assembly (**paragraph 1.1.3**). Interestingly, dynamic exchanges characterize ORC1 and ORC4 binding to chromatin, while ORC complexes instead remain stably bound to chromatin during *in vitro* biochemical isolation procedures [205]. Therefore, they propose a model in which ORC forms stable chromatin-associated complexes consisting of rapidly exchanging components (as depicted in Figure 1.18).

Figure 1.18 Model of ORC dynamics. ORC may exist as a more stable chromatin-associated complex, consisting of dynamic subunits. Details are contained within the text. Taken from: Mac Nairn *et al*. [203]



Taken together, these data indicate that biophysical approaches can be used in the study of DNA replication in living cells; indeed, the combination of these techniques and traditional biochemical approaches provide complementary information, that are useful for a better understanding of the dynamics of this process.

1.4 Aim of the work

The mechanisms by which DNA sequences are selected to be replication origins in metazoan (indeed, human) genomes are still not completely understood (see **paragraph 1.1.5**). The finding that HOX proteins have an affinity for human replication origins [101], and take part to the regulation of cell-cycle as well as DNA replication (**paragraph 1.2.5**), makes them interesting candidates for a role in origin specification. An involvement in the regulation of origin activation of these proteins would not be surprising, in light of their morphogenetic and often proto-oncogenic role (see **paragraph 1.2.4** and [104]). Nevertheless, this consideration raises a number of questions concerning the possible general role played by these proteins in the context of DNA replication regulation: does the affinity for origin sequences of HOX proteins correspond to an actual origin binding *in vivo*? What is the dynamics displayed by the HOX proteins at the origins? Does the origin binding by HOX proteins correlate temporally and spatially with the sites of DNA synthesis, in the context of cell-cycle progression? Can HOX be considered as RC members at the origin?

In order to answer to these questions, we have selected the human HOXC13 protein to be studied. This protein is and Abd-B paralog of the HOX family and was indeed shown to specifically recognize the Lamin B2 origin sequence [101]; furthermore, it was identified to be an oncoprotein [158].

The combined advances in FP biology and imaging methods (as reported in **paragraph 1.3**) are here applied to the study in living human cells of HOXC13 protein localization, dynamics, and interactions in the context of DNA replication regulation.

2. RESULTS

2.1 Expression and localization of endogenous HOXC13 in human cells

The investigation of HOXC13 function *in vivo* required a preliminary understanding of the cellular expression, as well as of the subcellular distribution, of the protein in human cells. To this purpose, we first explored the localization of endogenous HOXC13 in two human cell lines, namely U2OS and T98G cells, which showed to express the protein in significant detectable amounts (Figure 2.1A).

Immunofluorescence of endogenous HOXC13 was performed using a specific anti-HOXC13 antibody (an antibody that does not recognize the homeodomain was chosen in order to increase the recognition specificity and to avoid epitope hindrance by the DNA in the DNA-binding moiety of the protein). Endogenous HOXC13 showed a mainly nuclear localization in asynchronously growing U2OS and T98G cells (Figure 2.1B). The latter displayed also a cytoplasmic staining of the protein in a substantial part (≈50%) of the investigated cell population.

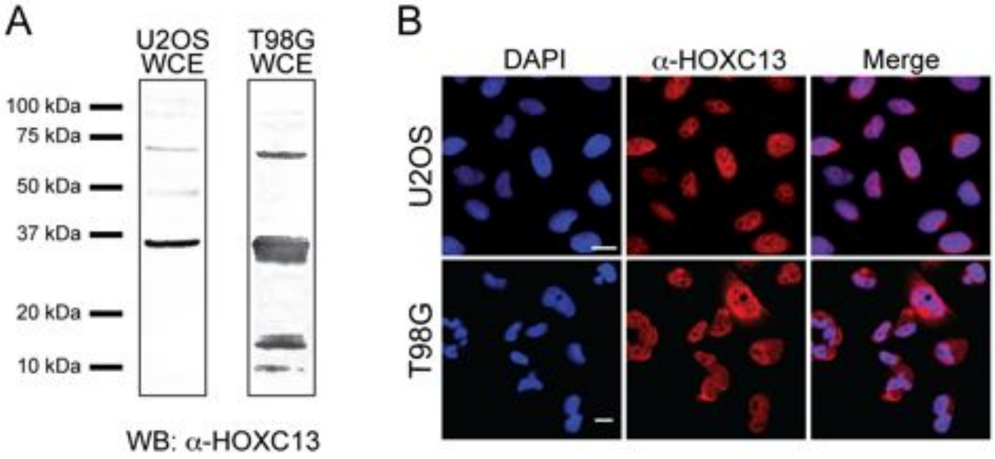


Figure 2.1 Expression and subcellular localization of endogenous HOXC13 in U2OS and T98G cells. A) Western blot detection of endogenous HOXC13 in whole cell extracts (WCE) of U2OS and T98G cells using a specific α-HOXC13 antibody: the most intense band runs at ≈ 37 kDa, consistently with the expected MW of 35.4 kDa of the protein. B) Immunofluorescence detection of endogenous

HOXC13 in U2OS and T98G cells using the same primary antibody as in panel A, subsequently detected with an Alexa647-conjugated secondary antibody. DAPI staining highlights cell nuclei. Scale bar: 20 μm.

We further characterized endogenous HOXC13 nuclear distribution. To this purpose, we performed, prior to cell fixation and subsequent HOXC13 immunofluorescence, a detergent pre-extraction by subjecting cells to a short incubation with 0.5% Triton-X100 with PBS. This treatment resulted in the complete removal of cytoplasm (as verified by microscope analysis) and of the nucleoplasm soluble proteins from the chromatin fraction, as already reported [21, 56, 206]. Under these conditions, no HOXC13 should be detected in the nuclei if the protein were exclusively a soluble component. Instead, a significant amount of the protein was still retained in U2OS pre-extracted nuclei (Figure 2.2A), providing a first indication that HOXC13 is a chromatin component; notably, this treatment unveiled a peculiar speckled-like, inhomogeneous distribution for chromatin-bound endogenous HOXC13 in both U2OS and T98G pre-extracted nuclei (Figure 2.2B).

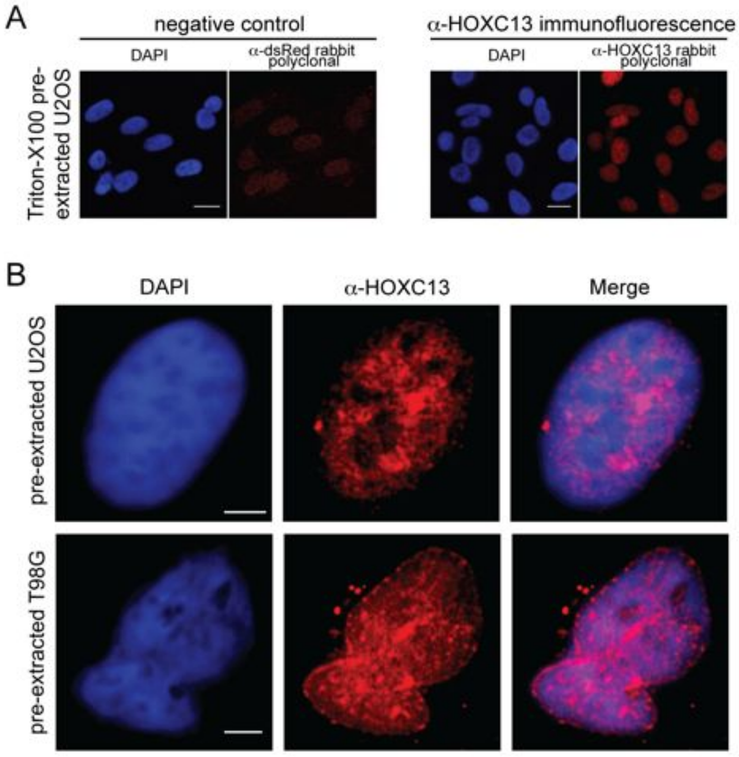

A
negative control
DAPI
α-dsRed rabbit polyclonal
α-HOXC13 immunofluorescence
DAPI
α-HOXC13 rabbit polyclonal
Triton-X100 pre-extracted U2OS
B
DAPI
α-HOXC13
Merge
pre-extracted U2OS
pre-extracted T98G

Figure 2.2 Nuclear distribution of endogenous chromatin-bound HOXC13 in U2OS and T98G cells. A) Validation of the pre-extraction procedure for the detection of chromatin-bound proteins. U2OS cells were extracted with 0.5% Triton-X100 in PBS before fixation and subsequently subjected to immunofluorescence using primary antibodies against a non-nuclear protein (negative control, α-dsRed rabbit polyclonal) or against HOXC13 (α-HOXC13 rabbit polyclonal). Images were acquired using the same set up parameters. A significant amount (> twice the negative control) of HOXC13 was detected in

U2OS nuclei. Scale bar: 20 μm. B) Magnification of pre-extracted U2OS and T98G nuclei showing the inhomogeneous, speckled-like, distribution of endogenous chromatin-bound HOXC13. Primary antibodies were detected with an Alexa647-conjugated secondary antibody. DAPI staining highlights cell nuclei. Scale bar: 5 μm.

In order to elucidate whether the detergent-extraction resistance was due to an actual chromatin affinity by the protein, we performed a biochemical *in vitro* fractionation analysis in U2OS cells. We first separated the cytoplasmic fraction of asynchronous cells from intact nuclei by low-speed centrifugation [21, 207]; the nuclear envelope was then broken with 0.5% Nonidet-P40 and the nucleoplasm was separated again by centrifugation [205]. The residual chromatin was sequentially extracted with increasing NaCl concentrations (ranging from 150 mM to 2 M), performing the first extraction step both in the presence and in the absence of DNase I. We then determined the presence of endogenous HOXC13 protein in the various fractions comparing them with the protein detected in a whole cell extract; as shown in Figure 2.3 (first row), we found almost no detectable HOXC13 in the cytoplasm, while both nucleoplasm and chromatin fractions contained a significant amount of HOXC13. The bulk of HOXC13 could be extracted from the chromatin fractions only at salt concentrations ≥300 mM; in particular its presence in the 600 mM and 2 M fractions confirmed that this protein is tightly bound to the nuclear structure.

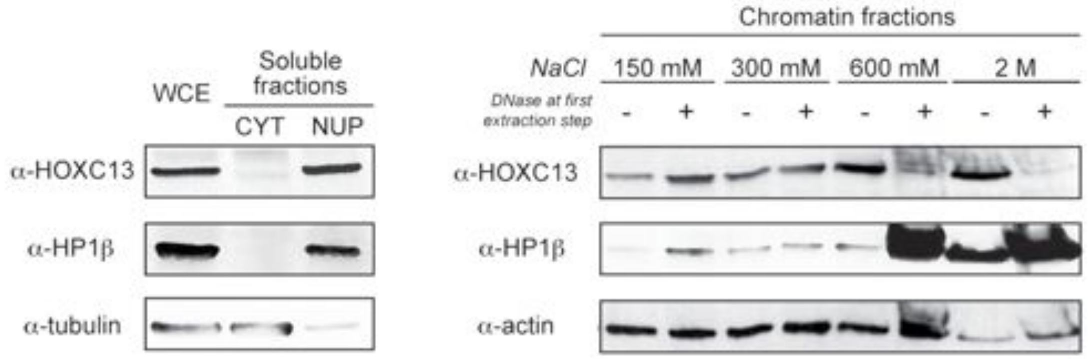


Figure 2.3 Western Blot detection of endogenous HOXC13 in comparison with HP1β in biochemically fractionated asynchronous U2OS cells. The soluble (left panel) and chromatin fractions (right panel) of asynchronous U2OS cells were investigated for the presence of endogenous HOXC13 and HP1β by Western Blot (tubulin is the control of actual cytoplasm/nucleus separation; actin is the loading control of chromatin fractions). The HOXC13 and HP1β protein levels detected in all fractions were compared to those of a whole-cell extract (WCE). CYT and NUP are cytoplasmic and nucleoplasmic fractions, respectively. The chromatin fractions are identified by the NaCl concentration used for the extraction. The first extraction step (150mM) was performed in the absence (-) or presence (+) of DNase I.

DNase treatment changed significantly the salt extraction profile: in this case, most of the protein was eluted at 150-300 mM NaCl, while the 600 mM and 2 M fractions were almost empty, providing evidence that HOXC13 is stably bound to the nuclear structure through its interaction with DNA. This behaviour was found to be specific for the HOXC13 protein; infact, the HP1β protein was investigated as a control, owing to its known chromatin-binding properties [207], and displayed a different fractionation profile when compared to that of HOXC13 in U2OS cells (Figure 2.3, second row).

2.2 Expression of GFP fusions of HOXC13 protein in human cells

In order to study HOXC13 in living cells, its sequence was fused to different variants of fluorescent proteins (FPs), spanning over almost the entire visible light spectrum. As schematically depicted in Figure 2.4A, the fluorophores were either fused at the N- or C-terminus of the protein. A longer linker was inserted between HOXC13 and FP in the C-terminal fusion construct with respect to the N-terminal one, to avoid steric hindrance of the protein homeodomain by the FP structure. The two constructs were transiently expressed in several human cell lines, namely U2OS (Figure 2.4B), HeLa (Figure 2.4C), T98G and MCF7 cells (data not shown); when expressed in all tested cell lines, both N- and C-terminal constructs showed a speckled-like nuclear localization very similar to that found for chromatin-bound endogenous HOXC13 (Figure 2.2B).

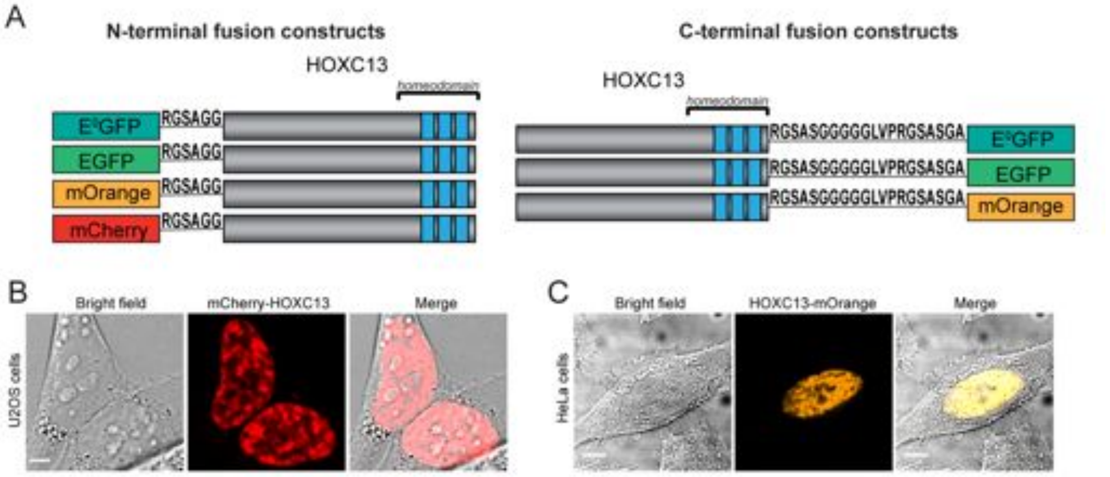

A
N-terminal fusion constructs
HOXC13
homeodomain
E²GFP
EGFP
mOrange
mCherry
RGSAGG
RGSAGG
RGSAGG
RGSAGG
C-terminal fusion constructs
HOXC13
homeodomain
RGSASGGGGGLVPRGSASGA
RGSASGGGGGLVPRGSASGA
RGSASGGGGGLVPRGSASGA
E²GFP
EGFP
mOrange
B
Bright field
mCherry-HOXC13
Merge
U2OS cells
C
Bright field
HOXC13-mOrange
Merge
HeLa cells

Figure 2.4 Fluorolabelled HOXC13: design of fusion constructs and expression in human cells. A) Shematic picture of the prepared fluorolabelled HOXC13 constructs. The FP sequence was either fused at the N- (left panel) or at the C-terminus (right panel) of HOXC13 sequence. The linker peptide designed for the fusion is indicated by its aminoacidic sequence. The three blue bars at the end of HOXC13 sequence represent the helix-turn-helix motif in the homeodomain (see Figure 1.11). B) Localization of mCherry-HOXC13 expressed in living U2OS cells. C) Localization of HOXC13-mOrange expressed in fixed HeLa cells. Scale bar: 5 μm.

We thus concluded that fusion of HOXC13 to FP variants does not cause mislocalization with respect to the endogenous protein. Although the N- or C- terminus positioning of the fluorophore did not seem to affect protein localization, in all subsequent experiments we chose to use the N-terminus fusion construct, in order to leave the DNA binding moiety free from the FP structure.

Further experiments were performed to test the correct expression and functionality of fluorolabelled HOXC13 constructs. First, biochemical extracts of U2OS cells transiently expressing mCherry-HOXC13 were investigated by Western Blot. The recombinant protein showed to be properly expressed in a whole cell extract, with no evident degradation of the construct (Figure 2.5A, first lane). Notably, the recombinant protein was absent in a U2OS cytoplasmic fraction (Figure 2.5A, second lane), similarly to what found for its endogenous counterpart (Figure 2.3). Conversely, a nuclear fraction was prepared by sequentially extracting U2OS nuclei with DNase I and 600 mM NaCl. Indeed this condition was proven to extract all endogenous HOXC13 from chromatin (see Figure 2.3). This nuclear fraction was found to be greatly enriched in mCherry-HOXC13 content when compared to the whole cell extract (Figure 2.5A, third and first lanes respectively). Furthermore, immunofluorescence experiments confirmed that recombinant E^0GFP-HOXC13 is recognized by the anti-HOXC13 antibody above described (Figure 2.5B) and, equally to the endogenous counterpart, retains detergent pre-extraction resistance possibly owing to its chromatin affinity (Figure 2.5C).

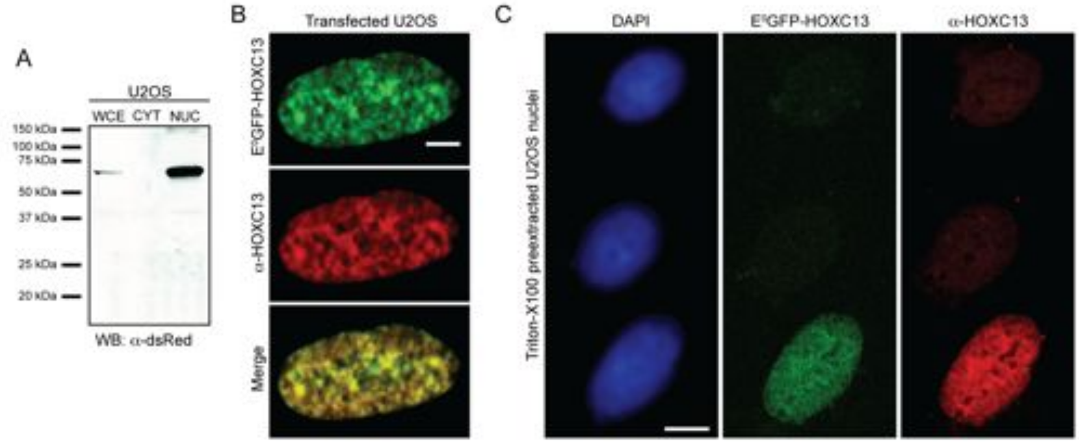


Figure 2.5 Check of correct expression and functionality of recombinant fluorolabelled HOXC13. A) U2OS cells transiently transfected with mCherry-HOXC13 were biochemically extracted to obtain whole cell extract (WCE), cytoplasmic (CYT) and nuclear (NUC) fractions, and 20 μg of each fraction were loaded on a gel and blotted using α-dsRed antibody to detect only recombinant mCherry-HOXC13. The theoretical MW of the fusion construct is ~ 63kDa. B) Immunofluorescence of U2OS cells expressing E^0GFP-HOXC13 using α-HOXC13 antibody, further detected with an Alexa 647 conjugated secondary antibody. The Merge image highlights the exact recognition of fluorolabelled HOXC13 by α-

HOXC13. Scale bar: 5 μm. C) Same immunofluorescence as described in B), but performed in U2OS cells subjected to detergent pre-extraction with 0.5% Triton-X100. An image containing both untransfected and transfected U2OS is reported to visualize both endogenous and recombinant HOXC13 detected with α-HOXC13 antibody (further detected with Alexa 647 conjugated secondary antibody). Scale bar: 10 μm.

Taken together, these data demonstrate that fluorolabelled HOXC13 retains, besides a similar nuclear distribution pattern, also the same properties of its endogenous counterpart, as far as chromatin affinity is concerned. Indeed, the recombinant protein is present only in the nucleus, where it is likely bound to chromatin, as proven by two parallel evidences: at first, a combination of high salt concentrations and chromatin DNA degradation greatly increases the amount of recombinant protein which can be recovered from U2OS nuclei, when compared to the protein detected in a whole cell extract (Figure 2.5A and Figure 2.3); in the second place, the protein is resistant to a detergent pre-extraction performed prior to cell fixation (Figure 2.5C).

Once the functionality of fluorolabelled HOXC13 was tested, we also investigated whether its over-expression, typical of all experiments performed by transient transfection of the construct, caused relevant cell-cycle alterations. HeLa cells transfected with E^0GFP-HOXC13 (same as Figure 2.4) were fixed either when asynchronous or after 1 h release from a double overnight thymidine block, which results in a cell-synchronization at the G1/S border.

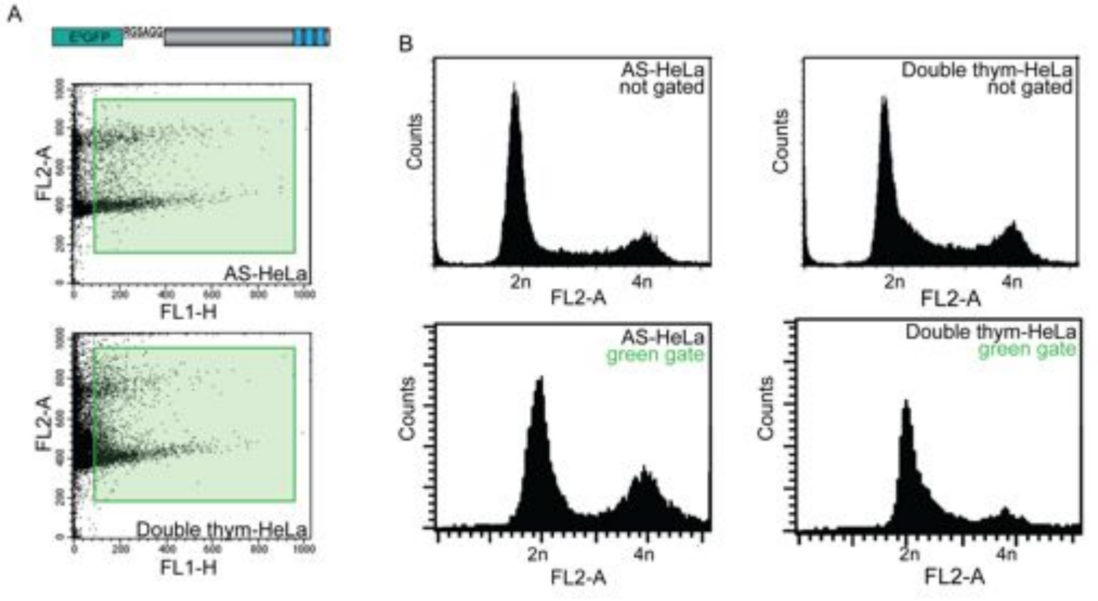


Figure 2.6 Impact of HOXC13 over-expression on HeLa cell-cycle progression. A) Top: schematic picture of the fluorolabelled HOXC13 construct (same as Figure 2.4). Bottom: plot of DNA content detected by propidium iodide staining (FL2-A) vs. green fluorescence (FL1-H) for the asynchronous (AS-HeLa, top panel) and the synchronized (Double thym-HeLa, bottom panel) samples. The green square corresponds to the selected cell population displaying green fluorescence at least 100-fold higher than

non-transfected cells. B) Cell-cycle profile of AS (left) or Double thym (right) HeLa cells after transfection without (top) or with (bottom) green-gate selection.

We measured by FACS the cell-cycle profile of transfected cells in the asynchronous and synchronized samples. In detail, we restricted this analysis only in the cell sub-population displaying a significant amount of transfected construct (i.e. we selected for green fluorescence, see Figure 2.6A). No significant difference in the cell-cycle profiles were detected for the selected green-fluorescent subpopulation of either asynchronous or synchronized HeLa cells, with respect to non-gated transfected HeLa cultures (Figure 2.6B).

We therefore concluded that transient expression of recombinant fluorolabelled HOXC13 in human cells -further used in all subsequent experiments below reported- does not alter significantly cell-cycle progression.

2.3 Co-localization of fluorolabelled HOXC13 with replication foci

An involvement of HOXC13 protein with the regulation of origin function would require its cellular localization in the nuclear compartments where DNA is replicated. We addressed this issue by investigating whether the localization observed for fluorolabelled HOXC13 correlates with the sites of DNA synthesis. These were visualized by fluorolabeling of the replication foci (RF) in cells in S phase, the moment of the cell cycle when DNA replication occurs (see **paragraph 1.1**). Two different approaches can be used to tag RF in intact cell nuclei. One is the pulsed incubation of the living cells with modified nucleotide-analogs, which are secondarily detected with a fluorescent probe after an appropriate permeabilization of the cells [194, 208, 209]. Tipically, alogenated nucleotide analogues, like BrdUrd, are detected by immunofluorescence [202]. Another approach, more recently developed, is the expression in living cells of FP fusions of replication factors, such as PCNA, Dnmt1 and DNA Ligase I, to visualize the dynamics of replication factories *in vivo* [195-197] (see **paragraph 1.3.2** for more details).

In our studies, both these strategies were used, allowing us to obtain complementary information, which will be discussed separately in the next **paragraphs 2.3.1** and **2.3.2**, respectively. Finally, in **paragraph 2.3.3** we shall present results concerning the role of HOXC13 homeodomain in the definition of the co-localization with RF.

2.3.1 Pulsed BrdUrd immunofluorescence

U2OS cells were transiently transfected with E^{0}GFP-HOXC13. Thereafter, they were pulse-labelled with BrdUrd for 10-15 min, fixed and processed for BrdUrd immunofluorescence, similarly to what already described [197]. Briefly, sites of BrdUrd incorporation, i.e. RF, were stained with a specific anti-BrdUrd antibody in the presence of DNase I and subsequently detected using an Alexa647-conjugated secondary antibody. In this way, a positive staining with red fluorescence allowed us to identify cells in S phase and to classify them in three categories depending on the spatial organization of BrdUrd foci. These were assigned to three distinct moments of S phase progression [202], as shown in Figure 2.7: early S (numerous and ubiquitous foci in the nucleus), mid S (perinuclear and perinucleolar foci), late S (some perinuclear foci plus bigger central clusters of foci) phases. These three distinct patterns of RF, although more evidently marked in mouse cells, are conserved throughout the S phase of all mammalian cells (primary, immortalized, transformed [210]).

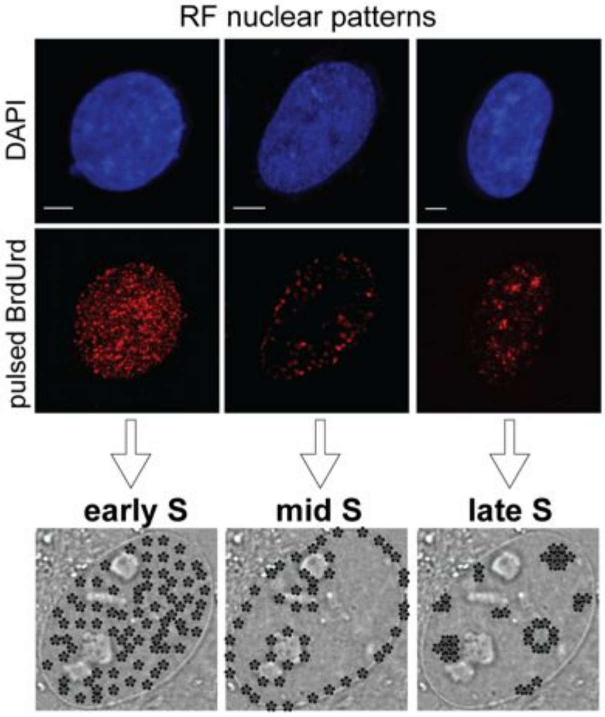


Figure 2.7 Definition of replication foci (RF) nuclear patterns in U2OS cells. Efficiency of BrdUrd immunofluorescence, as described in the main text, was checked by simultaneous DAPI staining of the nucleus. Three different patterns of RF distribution could be distinguished, based on the BrdUrd staining, which are schematically depicted on the bottom panel and described in the main text. Scale bar: 5 μm.

Moreover, the green fluorescence of transfected HOXC13 was related with the red one of RF, by performing a 3D co-localization analysis of the two channels [211, 212] at the three distinct moments of S phase above described. In detail, Z-stacks of BrdUrd and HOXC13 fluorescence images were sequentially acquired, comprising all nuclear thickness; images were restored by deconvolution and the Pearson's coefficient of co-localization (P) between the two channels was calculated (see **paragraph 4.6.2** for calculation details). This approach allowed both to increase the resolution of the fine BrdUrd focal structure and to monitor the co-localization in the whole nuclear volume. Results are summarized in Table 2.1 and Figure 2.8. As shown by the graphs in Figure 2.8, measurements performed in both U2OS (panel A) and HeLa cells (panel B) lead to analogous results and revealed a marked co-localization between HOXC13 and BrdUrd in early S phase cells, while the co-localization became progressively lower in mid S and it was almost lost in late S phase cells. Representative images of U2OS cells analyzed in these experiments, reported in panel C of the same figure, strengthen the finding that the co-localization of HOXC13 with RF is restricted to early S phase, while mid and late RF are gradually excluded from HOXC13 nuclear compartments.

These results could be reproducibly observed in two different human cell lines; nevertheless this approach suffers of the relatively invasive procedure used. Indeed, BrdUrd detection by the primary antibody requires destruction of the sorrounding nuclear DNA; this is usually achieved by acidic DNA denaturation in the presence of HCl before the antibody incubation [195] or, like we did in this study, by DNase treatment concomitant with antibody incubation [197]. We cannot exclude that such procedures, although performed in fixed cells, could result in some alteration of HOXC13 localization and nuclear solubility, as these were found to rely on its chromatin DNA affinity (Figures 2.2, 2.3 and 2.5).

In order to confirm our co-localization data in an intact chromatin context, we decided to use an alternative method to mark replication foci in living cells, as described in the following paragraph.

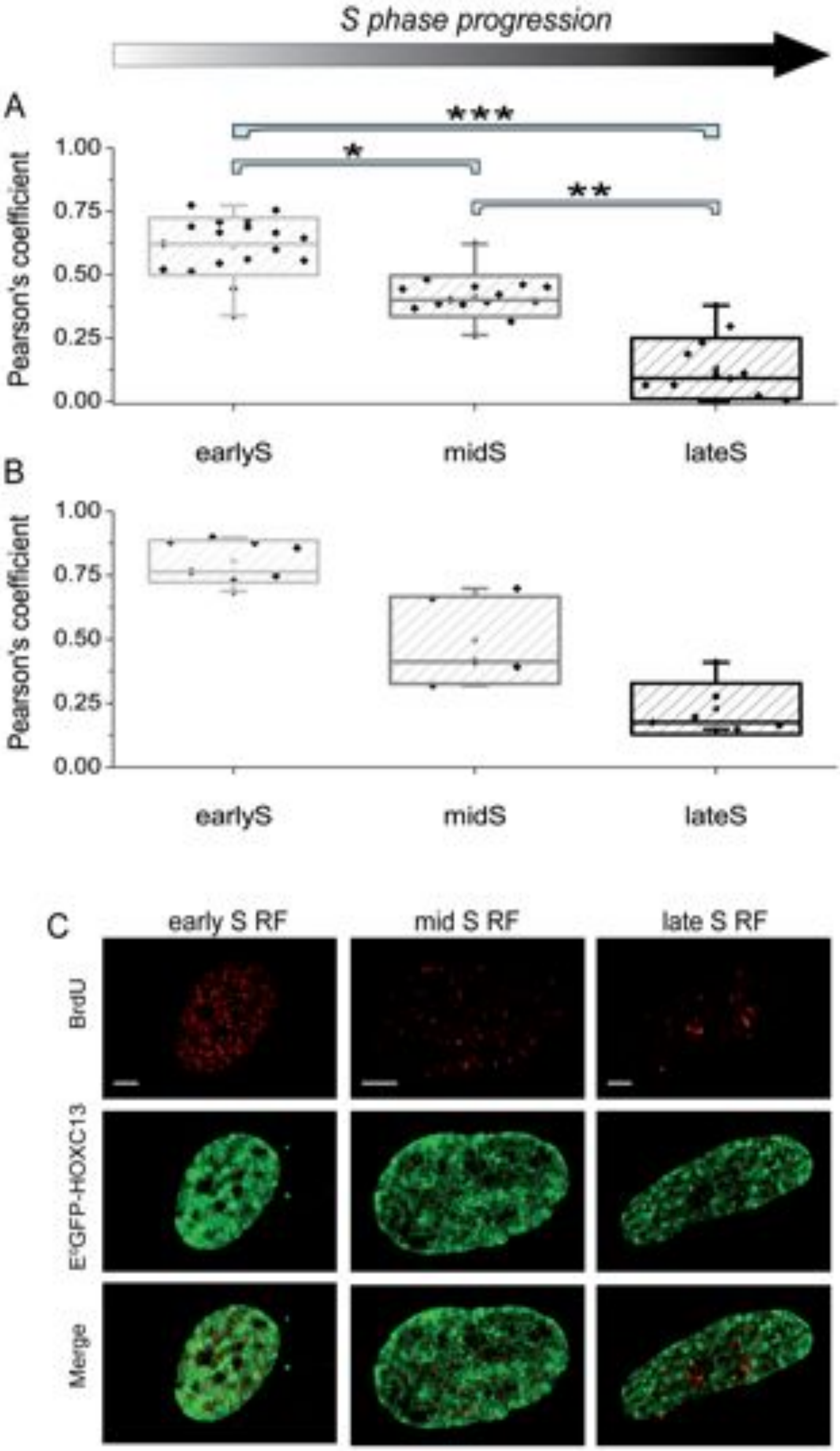

S phase progression
A
***
*
**
1.00
0.75
0.50
0.25
0.00
Pearson's coefficient
earlyS
midS
lateS
B
1.00
0.75
0.50
0.25
0.00
Pearson's coefficient
earlyS
midS
lateS
C
early S RF
mid S RF
late S RF
BrdU
E¹GFP-HOXC13
Merge

Figure 2.8. Co-localization analysis of fluorolabelled HOXC13 with RF throughout S phase progression (see previous page). Co-localization of HOXC13 with RF is quantified in each S phase moment by calculating the Pearson's co-localization coefficient (P) between the two fluorescence channels. Distributions of obtained P values from early, mid, late S phase cells are plotted for U2OS (A) and HeLa cells (B): the mean P is reported as a sphere enclosed in a box, whose vertical length represents the SD; the 5–95 percentile distribution is reported as a vertical line at the middle of each box. Gray-scale box colors correspond to S phase progression, according to the top arrow. Individual P values are reported as black dots behind each box. Statistical significance, as reported in **paragraph 4.6.2**, was calculated by the Mann-Whitney test (α=0.01): * ~0.0001, ** ~0.0001, *** <0.0001. C) Nuclear z-projections of representative U2OS cells displaying early, mid, late S BrdUrd patterns and expressing E^0GFP-HOXC13. The co-localization is highlighted by the Merge of the two channels. Scale bar: 5 µm.

Table 2.1 Co-localization analysis of fluorolabelled HOXC13 with RF throughout S phase progression. Mean P values (± SD) obtained from BrdU immunofluorescence analysis of both U2OS cells (second column) and HeLa cells (third column) at three distinct moment of S phase (each for one row). n refers to the number of cells analyzed for each sample.

Table 2.1	**U2OS**		**HeLa**	
Early S phase	P = 0.61±0.11	n=18	P = 0.81±0.08	n=8
Mid S phase	P = 0.42±0.08	n=15	P = 0.50±0.17	n=5
Late S phase	P = 0.13±0.12	n=12	P = 0.23±0.10	n=6

2.3.2 Replication factories visualized in real-time in living cells

In order to investigate whether the already described selective co-localization of HOXC13 with early S RF occurs also in living cells, human EGFP-HOXC13 was co-transfected with RFP-PCNA, a known marker of RF for live cell imaging [195, 199]. In detail, the two proteins were expressed in NIH3T3 cells, as mouse cells display more pronounced changes in the nuclear pattern of RF throughout all S phase.

The localization of RFP-PCNA allowed straightforward identification of cells in S phase: indeed, this protein displays homogeneous nucleoplasmic staining in non-S cells, while in S cells it selectively targets the RF and undergoes the same dynamic changes of pattern throughout S phase, as those reported in Figure 2.7 [196]. When selectively looking at cells in S phase in asynchronous NIH3T3 cultures, again we found a marked co-localization of HOXC13 with early S RF; this was less evident in mid S, and was virtually completely absent in the late-S phase (Figure 2.9).

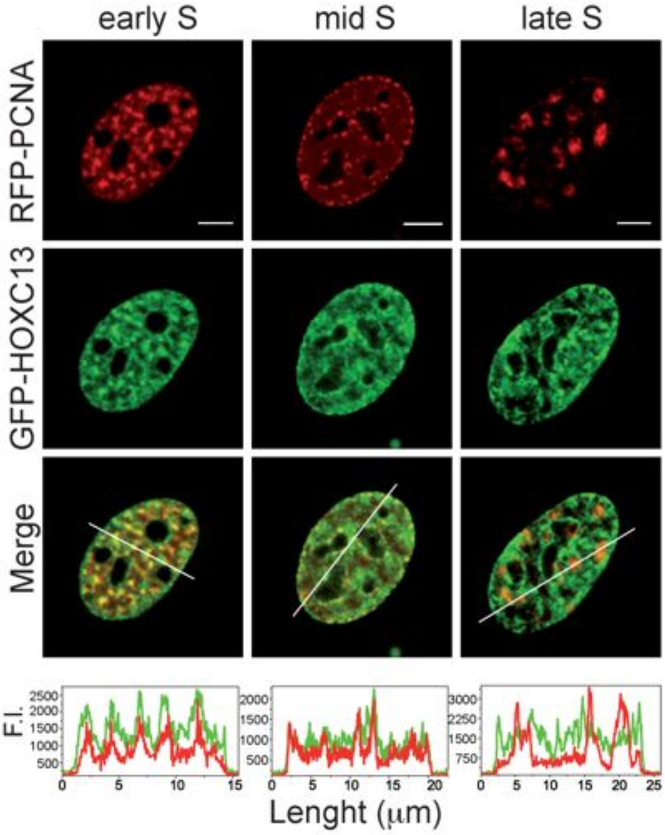


Figure 2.9 Cell-cycle modulated presence of HOXC13 at RF. EGFP-HOXC13 was co-transfected with RFP-PCNA [199] in NIH-3T3 cells. The co-localization of the two constructs during S phase can be evaluated by the Merge image. The intensity profile of green and red channels in a nuclear section (depicted as white bar across the nuclei of Merge images) is plotted below each Merge image: the different correlation between green and red signals in the three moments of S phase represents the differential co-localization of HOXC13 with RF throughout S phase. Scale bar: 5 μm.

Moreover, the availability of a marker for *in vivo* application allowed for time-lapse experiments throughout S phase in individual living cells expressing both constructs. Representative images of a time-lapse experiment of one analyzed cell are reported in Figure 2.10. These show an initial overlap of the green and red channels, i.e. HOXC13 and PCNA respectively, when RF display an early S pattern; the two channels become progressively more separated, meaning loss of co-localization, as RF get re-organized and become less numerous in the nucleus; homogenization of PCNA nuclear staining finally marks the end of S phase, after 5 h and 30 min observation, which is consistent with the complete length of S phase in mouse fibroblast cancer cells [213].

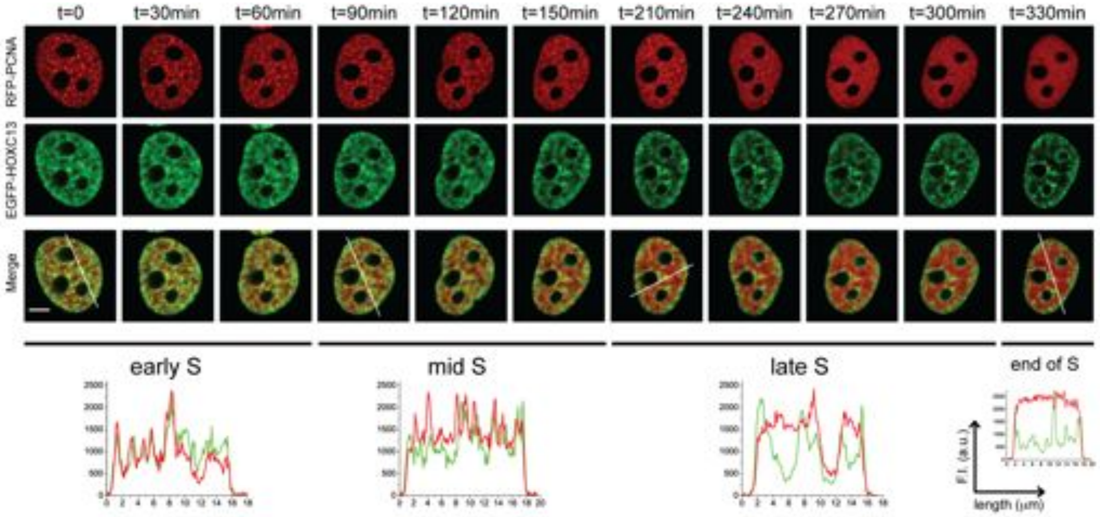


Figure 2.10 Time-lapse imaging of S phase progression in NIH3T3 cells expressing EGFP-HOXC13 and RFP-PCNA. An individual cell transfected with the two constructs was imaged at the indicated time points while constantly kept at 37°C and 5%CO_2. The start point (t=0) corresponds to a typical early S pattern of RF. The co-localization course is highlighted by the Merge image. The intensity profile of green and red channels in a nuclear section (depicted as white bar across the nuclei of selected Merge images) is plotted below the Merge images, as already explained in Figure 2.9. PCNA homogeneous nuclear staining (see the profile of the red channel below the last Merge image) marks the end of S phase. Scale bar: 5 µm.

2.3.3 Role of the homeodomain in the targeting of early replicating chromatin

Data obtained in living cells (**paragraph 2.3.2**) are in agreement with, and complement, those obtained in fixed cells (**paragraph 2.3.1**). Indeed by BrdUrd immunofluorescence (Figure 2.8), we were able to quantify the degree of co-localization between HOXC13 and RF, demonstrating that it is significantly modulated from early to late S. By live-cell imaging (Figures 2.9 and 2.10), we confirmed that this actually occurs also in an intact chromatin context, and we were able to follow in real time their mutual nuclear re-distribution during S phase.

Also, results obtained by both these approaches point to the speckled-like, chromatin-bound, HOXC13 distribution as the actor for the selective targeting of the nuclear areas where early S DNA replication occurs. In order to assess whether chromatin binding is relevant to determine this selective co-localization, a fluorolabelled deletion mutant of HOXC13 was prepared, as reported in Figure 2.11, in which aminoacids 258–330 were deleted. Deleted residues encompass the homeodomain (compare the schematic picture of the full length protein with that of the deletion mutant in panel A). This conserved helix-turn-helix motif (described in detail in **paragraph 1.2.2**, Figures 1.11-12) is known to perform two functions: it allows homeodomain proteins to bind chromatin DNA and it is a nuclear localization signal [106, 136]. Following removal of the homeodomain and in order to maintain nuclear localization properties to the mutant, a different localization sequence (NLS_{SV40}) was added to the deletion mutant. When compared to the fluorolabelled wild type protein (see Figure 2.4), the deletion mutant displayed a homogeneous, diffused nuclear staining.

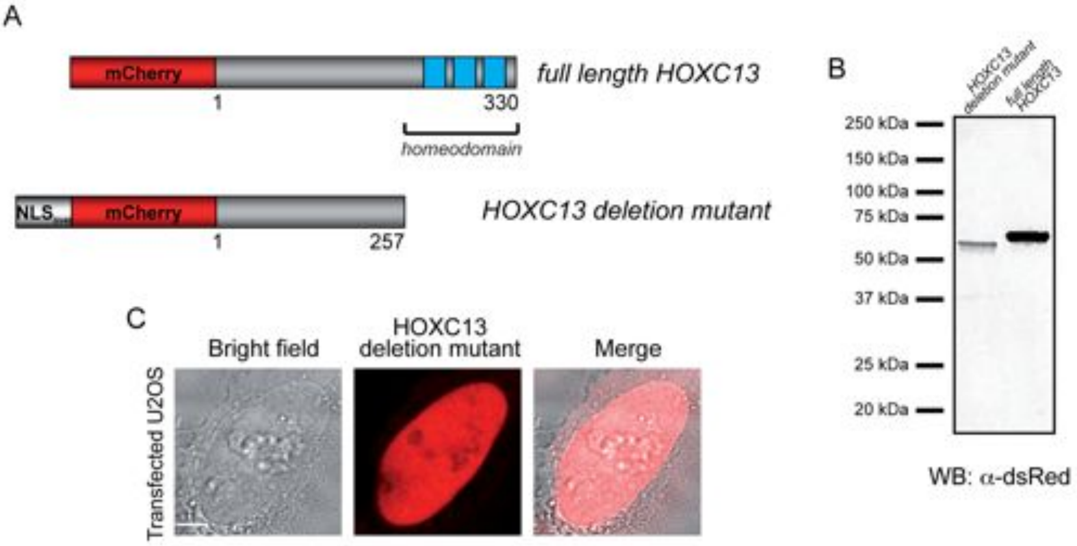

A
mCherry
1
330
full length HOXC13
homeodomain
NLS
mCherry
1
257
HOXC13 deletion mutant
B
HOXC13 deletion mutant
full length HOXC13
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa
WB: α-dsRed
C
Bright field
HOXC13 deletion mutant
Merge
Transfected U2OS

Figure 2.11 Fluorolabelled deletion mutant of HOXC13. A) Schematic picture representing both the mCherry fusion of full-length HOXC13 protein (top) and the fluorolabelled HOXC13 deletion mutant (bottom), devoid the homeodomain (residues 258-330) and supplemented with a NLS_{SV40} at N-terminal position. B) Western Blot control of the correct expression of HOXC13 deletion mutant (left lane) in comparison with the full-length (right lane) in U2OS whole cell extracts (20 μg loaded per well). The expected MW of the fluorolabelled deletion mutant is 55.3 kDa, while that of the fluorolabelled full-length construct is ~ 63 kDa. C) Localization of HOXC13 deletion mutant expressed transiently in U2OS cells. Scale bar: 5 μm.

A co-localization analysis of the HOXC13 deletion mutant with RF would indiscriminately lead to the same result with all replication factories throughout S phase, due to its diffuse staining of the nucleus. Therefore, these findings demonstrate that the homeodomain is responsible for the speckled-like nuclear distribution of the full-length protein (see Figures 2.2 and 2.4). Importantly, the selective co-localization of HOXC13 with early-S RF (see Figures 2.8-2.10) relies on the peculiar chromatin binding ability of its DNA-binding motif.

Overall, experiments reported so far in **paragraphs 2.1-2.3** unveiled an important role for chromatin DNA affinity of HOXC13 homeodomain, not only for the definition of a precise nuclear localization of the protein, but also in the context of a possible role of the protein in DNA replication regulation. This motivated us to investigate HOXC13 nuclear dynamics properties: a description of the obtained results is reported in the following paragraph.

2.4 HOXC13 nuclear dynamics

After we demonstrated the homeodomain-dependent presence of fluorolabelled HOXC13 at replication factories located in early replicating chromatin, we next investigated its dynamics within these nuclear regions in living cells using Fluorescence Recovery After Photobleaching (FRAP).

To the best of our knowledge, no FRAP experiments have been reported so far to analyze the nuclear dynamics of HOX proteins. Therefore, we decided to use a FRAP procedure, which was previously applied to study the dynamics of many different nuclear proteins, namely half-FRAP [188]. Briefly, this technique consists in the bleaching of half a nucleus of a cell expressing a GFP-tagged nuclear protein, and in the subsequent monitoring of signal recovery in the bleached area (see Figure 1.16C). Accordingly, experiments were performed, and obtained results will be presented here, following the logical scheme reported below:

First. Define the correct experimental procedure in order to perform reliable and consistent half-FRAP experiments. This was achieved by using a nuclear protein, with known properties in terms of nuclear dynamics, as a standard to optimize our microscope set up for half-FRAP acquisitions (**paragraph 2.4.1**);

Second. Choose a fluorophore, which could be ideally suited for half-FRAP procedures as well as long-term imaging in living cells, in order to tag HOXC13 protein (**paragraph 2.4.2**);

Third. Perform half-FRAP experiments on fluorolabelled HOXC13 protein. The aim here was to understand which are the key factors determining its dynamics properties (expression level of the construct? Chromatin binding by the construct? DNA binding ability of the construct? - **paragraph 2.4.3**);

Fourth. Get quantitative, absolute parameters describing HOXC13 dynamics properties, so that a comparison with the same parameters reported for other nuclear proteins is possible (**paragraph 2.4.4**);

Fifth. Compare HOXC13 dynamics to those of the numerous proteins investigated so far using the same procedure, such as transcription factors and coactivators, structural, and remodeling proteins [188, 191, 214], as well as replication proteins [196, 199, 215] (**paragraph 2.4.5**).

2.4.1 Half-FRAP experiments using the EGFP-HP1α protein as a standard

Our half-FRAP procedure was first tested for reliability by the analysis of a nuclear protein with known properties in terms of dynamics. To this purpose, we chose HP1α protein, as the dynamics of HP1 proteins has been extensively characterized by several biophysical tools, including half-FRAP [183, 184, 188, 189, 216, 217]. U2OS were transiently transfected with EGFP-HP1α construct; half nucleus of transfected cells was bleached, and fluorescence redistribution monitored in a subsequent time series (Figure 2.12A); the resulting normalized recovery of fluorescence intensity in the bleached area was finally plotted vs. time (Figure 2.12B; see **paragraph 4.7.2** for the details of normalization).

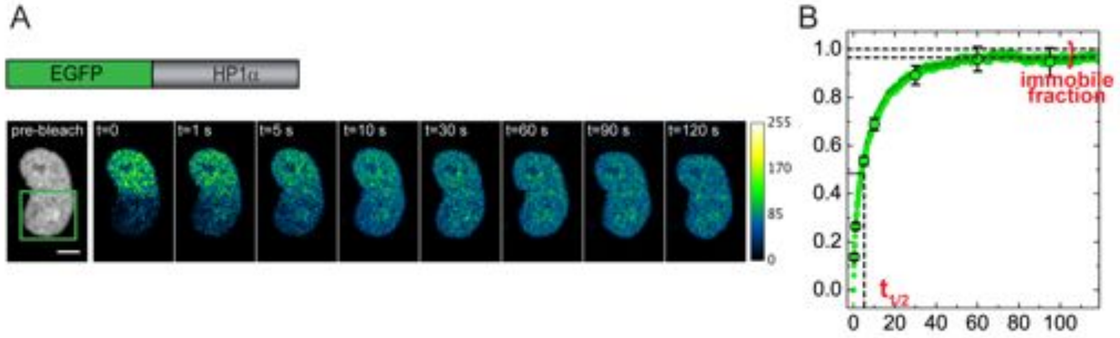


Figure 2.12 Half-FRAP experiments performed with the EGFP-HP1α protein in U2OS cells. A) Schematic picture of the construct used (top) and representative half-FRAP experiment (bottom) with EGFP-HP1α. The pre-bleach image (gray look up table) is an average of 10 frames acquired before bleaching; the post-bleach images (blue-green look up table) are single frames corresponding to the indicated times. The green square in the pre-bleach image represents the bleached area. Scale bar: 5 μm. B) Normalized fluorescence-intensity recovery vs. time for EGFP-HP1α. The curve is averaged from data of all analyzed cells (n=9). Standard errors values are reported on selected points. $t_{1/2}$ and immobile fraction values are highlighted in red color.

From the recovery curve, we derived two parameters (highlighted in the same graph), which can give a general estimate of the dynamic properties of the protein: $t_{1/2}$ is the recovery halftime and quantifies the rate at which molecules in the bleached area are replaced by molecules from the unbleached portion of the nucleus; the immobile fraction represents the amount of bleached protein which does not diffuse to the unbleached nucleus in the observed time-range, and is calculated as the difference between the total fluorescence recovery and the obtained plateau recovery.

For the EGFP-HP1α construct, we obtained $t_{1/2}$ = 3.5-4 s and immobile fraction ≈ 4%. Moreover, plateau fluorescence recovery was achieved within 60 s. These data are definitely in agreement with those already published for the protein, with typical values of $t_{1/2} \leq 10$ s and immobile (or very slow) fractions ≈ 5% [183, 184, 188]. Thus, this preliminary experiment demonstrated the reliability of our microscope set up for half-FRAP experiments so that we could proceed with the analysis of HOXC13 nuclear dynamics.

2.4.2 Choice of a fluorophore suitable for the half-FRAP procedure

Most of FRAP experiments reported so far for proteins tagged with FPs make use of the EGFP, due to the optimal combination of brightness, photostability, and possibility to achieve photobleaching by short, high laser pulses, offered by the green FP variant. Recently, red-emitting derivatives that equal the advanced properties of EGFP have been developed [168-172].

In order to exploit the advantages offered by red FPs (in particular, the reduced phototoxicity, due to their red-shifted laser excitation, during long-term imaging), we performed our half-FRAP experiments using the mCherry fluorophore, one of the most promising monomeric red FPs [170]. This choice was preliminary based on the observation that fusion of red rather than green FPs to HOXC13 did not result in localization changes nor did it affect expression of the fusion construct (Figure 2.4). Moreover, red FPs have been already used in dual-color FRAP experiments, in combination with EGFP [218-220]. Finally, we compared by half-FRAP the nuclear mobility of two equally designed EGFP- and mCherry- HOXC13 fusion constructs expressed in U2OS cells. We selected in both cases cells displaying various expression levels of the construct.

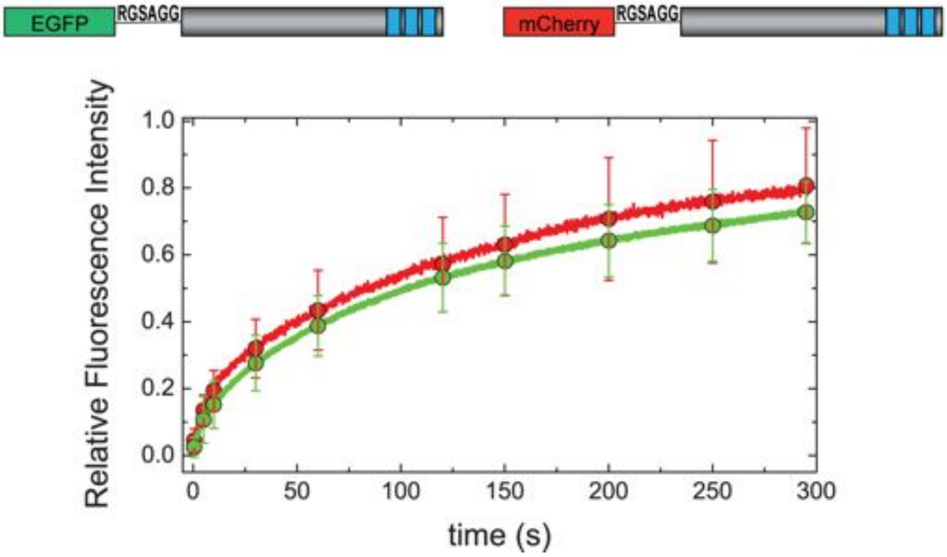


Figure 2.13 FRAP curves for EGFP- and mCherry- HOXC13 fusion constructs. Top: the two equally designed constructs (same as Figure 2.4). Bottom: mean recovery curves obtained from half-FRAP experiments performed in U2OS cells expressing EGFP-HOXC13 (n=14, green curve) or mCherry-HOXC13 (n=12, red curve). Representative SD values are reported for selected points of the curves.

We found for the two constructs similar mean recovery curves (Figure 2.13) and, considering the spread of data around the mean value, we concluded that there is no relevant difference between the two FRAP curves. Together, the documentation in the literature and our experimental data demonstrate that mCherry is a suitable alternative fluorophore to EGFP in FRAP experiments, so that it could be used in the subsequent analysis of HOXC13 nuclear dynamics.

2.4.3 Half-FRAP analysis of the nuclear dynamics of wt mCherry-HOXC13

The first issue that we addressed in the study of HOXC13 dynamics was the identification of the useful range of construct expression level ensuring reproducible results. The general rule of thumb, when performing FRAP experiments, is to consider only cells with medium-to-low expression profiles of the construct, to avoid over-expression artifacts. However, in our case the variability observed for FRAP curves was unexpectedly high even in response to slight variations of the construct expression level, both for EGFP- and mCherry-HOXC13 (Figure 2.14A, B; see also the relatively high spread of data in Figure 2.13). This motivated us to investigate deeper the dependence of HOXC13 dynamics on its expression level.

To this purpose, half-FRAP measurements were performed in cells expressing various amounts of mCherry-HOXC13, keeping constant all set-up acquisition parameters, but the fluorophore-excitation laser power (543 nm). In this way, laser power was directly linked to the expression level (high expression levels need low laser power to be visualized, vice versa for low expression levels). Mean $t_{1/2}$ and immobile fraction were estimated from the recovery curves and were compared to the values obtained for a stable, very low expression of the same construct in U2OS cells. We found that high expression levels resulted in uncertain $t_{1/2}$ overestimates (Figure 2.14C) and dramatically increased the immobile fraction (Figure 2.14D, E). On the contrary, as the protein concentration decreased, both parameters reached stable values; in detail, only cells acquired at 10-20 μW of laser power displayed $t_{1/2}$ and immobile fraction stable values comparable to those obtained for a stable expression of the construct.

Therefore, based on these results, we decided to perform all subsequent half-FRAP experiments on transiently transfected cells, displaying expression levels within the 10-20 μW laser power acquisition range.

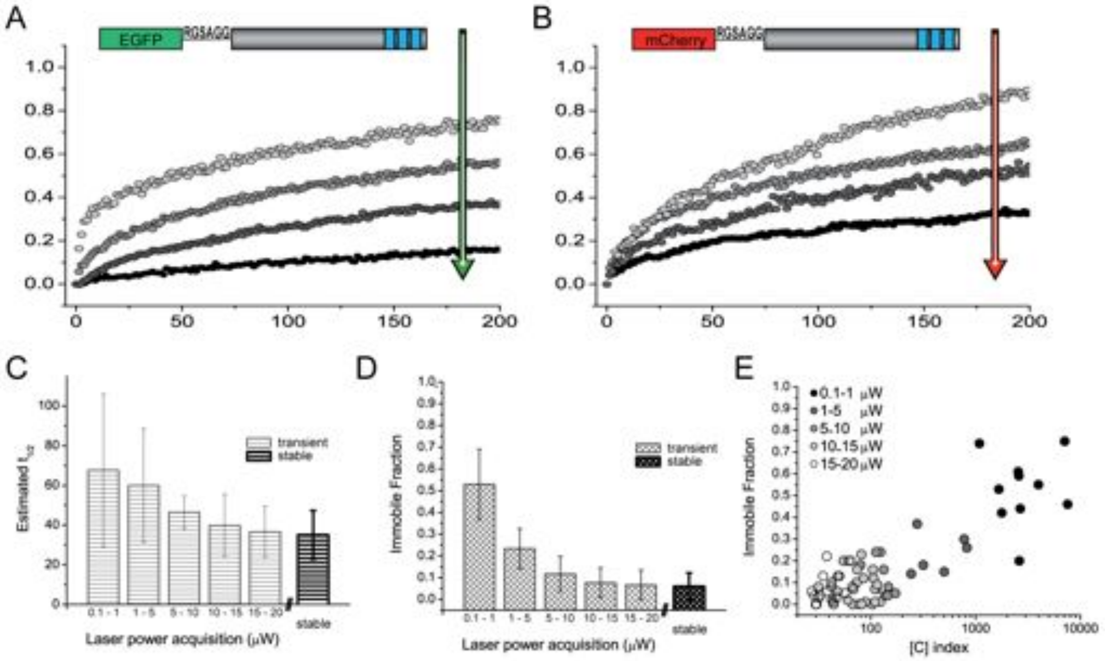


Figure 2.14 Dependence of wt HOXC13 nuclear dynamics on expression levels. FRAP curves of 4 representative cells displaying different expression levels of EGFP-HOXC13 (panel A) and mCherry-HOXC13 (panel B). The arrows represent protein concentration increase. Mean estimated $t_{1/2}$ (panel C) and immobile fraction (panel D) of all analyzed mCherry-HOXC13-expressing cells (n=67) plotted vs. ranges of acquisition (from 0.1-1 to 15-20 μW) of 543 nm laser power (error bars are SD). Mean $t_{1/2}$ and immobile fraction of a stable expression of the construct are reported as bold histogram bars. E) Dependence of the immobile fraction on protein concentration, expressed as [C] index: this was calculated as a ratio between mean pre-bleach nuclear fluorescence and acquisition laser power. Different gray-scale tones correspond to different laser power ranges of acquisition, as indicated; a similar gray-scale is used in the graphs of panels A and B.

We next investigated the nuclear dynamics of mCherry-HOXC13. To this purpose, three mCherry-tagged constructs (Figure 2.15A: NLS_{SV40}-mCherry, mCherry-HOXC13, mCherry-Deletion mutant) were in turn expressed in U2OS cells, and analyzed by half-FRAP as explained before (Figure 2.15B). Under these conditions, wt mCherry-HOXC13 displayed the slowest average recovery curve of all constructs (Figure 2.15C). Its mean halftime of recovery was $t_{1/2} = 35\text{-}40$ s; within 300 s of post-bleach observation signal recovery reached a plateau corresponding to about 10% immobile fraction. The slow dynamics of wt HOXC13 was compared to the recovery rate under the same bleaching conditions of the NLS_{SV40}-mCherry construct. The latter was chosen as a control to monitor unspecific binding events: we obtained $t_{1/2} \approx 0.75$ s, full recovery in ~ 6.5 s (similarly to what previously reported for NLS_{SV40}-GFP [221]). In order to investigate whether this difference was caused by the anchoring of the homeodomain to the nuclear structure, we measured FRAP for the mCherry-HOXC13 deletion mutant devoid of the homeodomain, already described in Figure 2.11. We found that upon deletion of the homeodomain and insertion of NLS_{SV40} to rescue protein nuclear localization, the dynamics was very similar to that of the NLS_{SV40}-mCherry control protein ($t_{1/2} \approx 1$ s, full recovery in ~ 10.5 s).

Thus, we can conclude that wt HOXC13 displays a peculiar nuclear mobility, about 50-fold slower than a non-interacting control, which relies on the chromatin affinity of its homeodomain. This observation is in agreement with the properties of chromatin interaction previously demonstrated for both endogenous and recombinant HOXC13 (Figures 2.2, 2.3, 2.5). In particular, the presence of intact DNA was found to be relevant for the maintenance of endogenous HOXC13 at chromatin subnuclear-fractions (Figure 2.3). With regard to this finding, we further investigated HOXC13 nuclear mobility in response to changes in the homeodomain-DNA affinity. To this end, we mutated to alanine the homeodomain residues that create crucial contacts with DNA, as reported both by crystal structure [129, 130] and *in vitro* affinity studies [139].

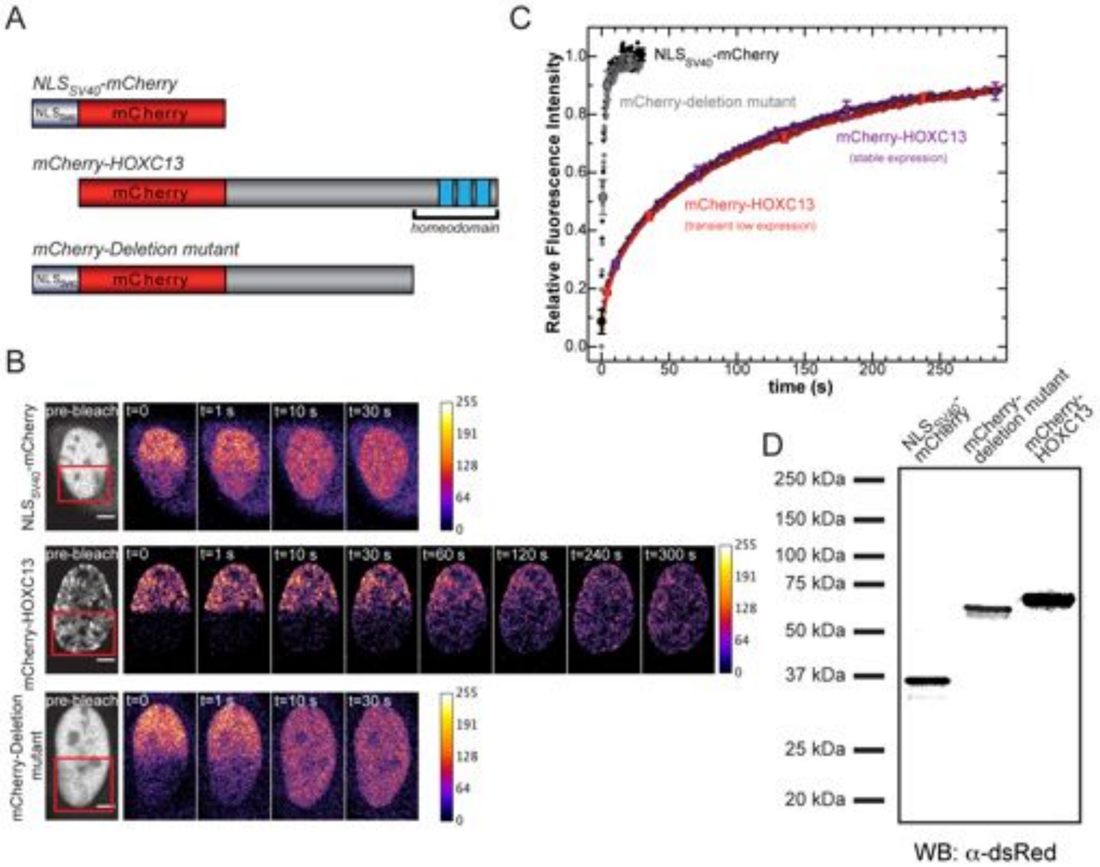


Figure 2.15 Half-FRAP analysis of wt mCherry-HOXC13. A) Schematic picture of the constructs used: (top, non-binding control) NLS_{SV40}-mCherry, (middle) wt mCherry-HOXC13, and (bottom) mCherry-Deletion mutant of HOXC13 devoid of the homeodomain. B) Representative FRAP experiments with the three constructs. The pre-bleach image (gray look-up table) is an average of 5-10 frames acquired before bleaching (red square in the same image); the post-bleach images (colored look-up table) are single frames corresponding to the indicated times. Scale bar: 5 μm. C) Normalized fluorescence-intensity recovery vs. time for the three analyzed constructs (reported in different colors). The wt mCherry-HOXC13 construct was expressed either transiently (red dots) or stably (purple dots). Curves are averaged from data of all analyzed cells (NLS_{SV40}-mCherry: n=11, mCherry-Deletion mutant: n=12, transient wt mCherry-HOXC13: n=40, stable wt mCherry-HOXC13: n=18). Standard error values are reported on selected points. D) Western Blot control of the correct expression of the three constructs.

Figure 2.16A schematically shows the mutated constructs prepared. The “HBX-helix mutant” (top construct), was obtained by mutating three residues of the third helix of the homeobox (Ile47, Gln50, Asn51, numbers correspond to the classical homeobox numeration, as indicated in Figure 1.12 at **paragraph 1.2.2**). This is also known as “recognition helix”: it inserts into the DNA major groove, where the cited residues provide a contact with DNA bases (Figure 2.16C, top panel).

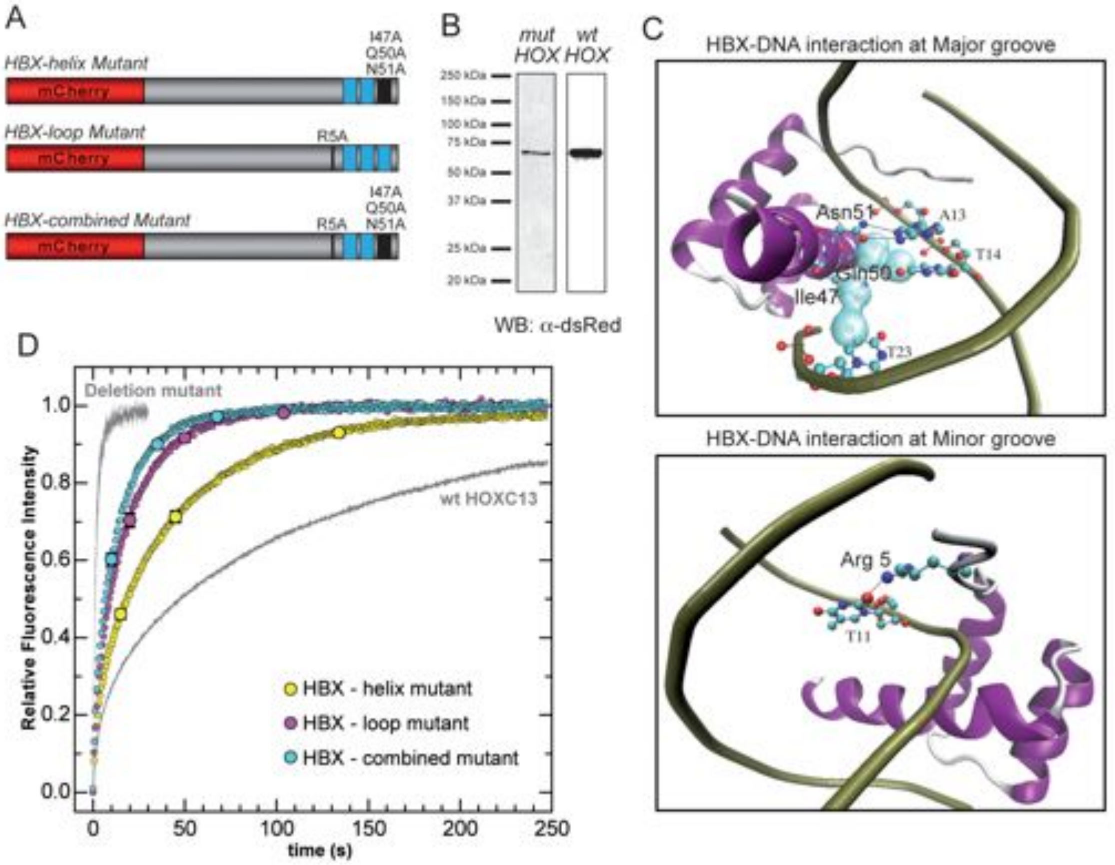


Figure 2.16 FRAP analysis of three homeodomain (HBX) mutants of mCherry-HOXC13. A) HBX mutant constructs: the HBX-helix mutant (top, three mutations in the third HBX recognition helix); the HBX-loop mutant (middle, single mutation at the HBX N-terminal loop); the HBX-combined mutant (bottom, combining both helix and loop mutations). B) Detection of the HBX-helix mutant (*mut HOX*) in comparison to the wt mCherry-HOXC13 (*wt HOX*) by Western Blot. C) The homeobox-DNA co-crystal structure is the same as Figure 1.12: the two panels correspond to a magnification of the interactions occurring at DNA major (top) and minor (bottom) groove. Only interactions mediated by the mutated residues described in A) are highlighted in the two panels. Top panel: HBX-major groove interaction is mainly mediated by two hydrogen bonds (depicted with black dashed lines) between residue Asn51 and adenine-13 base and Van der Waals contacts (light cyan surface contour) between Ile47, adenine-13 and thymine-14 and between Gln50 and thymine-23. Bottom panel: HBX-minor groove interaction is mediated, among others, by a hydrogen bond (black dashed line) between Arg5 and thymine-11. D) Normalized fluorescence-intensity recovery vs. time of the three mutants (colored dot-curves). The mCherry-Deletion mutant and the wt mCherry-HOXC13 curves (same as Figure 2.15) are also reported in gray color, as a reference for 0 and 100% DNA-binding activity, respectively. All data represent mean recovery curves (±SE, reported at selected time points) obtained from analyzed cells (HBX-helix mutant: n=38, HBX-loop mutant: n=20, HBX-combined mutant: n=37).

Mutation of these to alanine causes abolition of DNA binding capacity for residues 50 and 51 [139]. The HBX-helix mutant was investigated by FRAP with the same protocol used for wt HOXC13; the observed nuclear mobility of this mutant was increased more than twice ($t_{1/2}\approx$ 17.5 s, 98% recovery in ~ 160s) with respect to the wt protein (Figure 2.16D). In the “HBX-loop mutant” (Figure 2.16A, middle construct)

Arg5 was mutated to alanine; Arg5 belongs to the N-terminal loop of the homeodomain and makes contact with the DNA minor groove (Figure 2.16C, top panel). Crystal structure shows that among all DNA-homeobox contacts, Arg5 forms the strongest hydrogen bond by interacting with a thymine base of the minor groove [129, 130]. Arg5 mutated to alanine causes the total abolition of DNA-binding capacity *in vitro* [139]. Our FRAP data revealed that this mutant displays more than 4-fold increased mobility with respect to the wt protein ($t_{1/2}\approx$ 9s, full recovery in ~ 100s). We also combined the three mutations of HBX-helix with the HBX-loop one to get the "HBX-combined mutant" (Figure 2.16A, bottom construct), which yielded a 5-fold mobility increase ($t_{1/2}\approx$ 7s and full recovery in ~ 80s). Thus, these data provide evidence that DNA binding ability is responsible for the slow nuclear dynamics of the protein, and its progressive impairment causes an up to 5-fold increased mobility as well as the loss of immobile fraction in the observed time range.

2.4.4 Quantitative analysis of wt mCherry-HOXC13 nuclear dynamics

A comparison between HOXC13 and other nuclear proteins, as far as nuclear dynamics is concerned, requires the definition of quantitative parameters describing the kinetics of chromatin binding by the protein. This approach was indeed often used to describe the nuclear dynamics of chromatin-binding proteins investigated by FRAP [182, 185, 186, 188, 214, 217]. To this purpose, we applied a simplified diffusion-reaction mathematical model to FRAP curves obtained from experiments in cells expressing mCherry-HOXC13. Although a detailed description of the model will be presented in Appendix A of this thesis, it is useful to define here the parameters that we can derive from the model, quantitatively defining HOXC13 chromatin-binding kinetics.

Briefly, the model describes the redistribution of unbleached protein within the bleached area, assuming that it is governed both by free diffusion and by immobilization due to chromatin binding, as shown for other proteins [186, 214, 217]. Fitting our FRAP curves to a function combining diffusion and interaction kinetics allowed to estimate three parameters, namely:

- D, the diffusion coefficient of free protein. Actually, the estimate for D was too slow to be explained by free diffusion; therefore it is further referred to as effective D, or D_{eff}. This is to underline that it probably reflects the process of effective diffusion, whereby fast and unspecific binding interactions combine with free diffusion to mimic a quite slowed diffusion [186, 214, 222].
- $K=k_{on}^{*}/k_{off}$, the equilibrium constant of binding of chromatin-bound protein. For a direct comparison with published data [214], this parameter was further converted into F_{free} (% free protein), using the following equation:

$$F_{\text{free}} = \frac{[F]_{eq}}{[F]_{eq} + [C]_{eq}} = \frac{1}{1+K}; \qquad \text{(Eq.2.1)}$$

where [F] and [C] represent the concentration of free and bound proteins, respectively.

- k_{off}, the dissociation rate of the protein interaction with chromatin. Again, in accordance with published data [188], this parameter was inverted to obtain the mean residence time of the protein in the chromatin-bound state, $t_{\text{mean}}=1/k_{\text{off}}$.

We used two different approaches to apply this model, as reported in Figure 2.17. In the first case (see panel A), the fluorescence recovered in the whole bleached area was fitted, assuming that it was covering exactly half of the nucleus (dimensionally approximated by a parallelepiped defined by its nuclear length, *l*). Under these conditions, we obtained $D_{\text{eff}} = 5.03 \pm 2.08\ \mu\text{m}^2/\text{s}$, a value consistent with other D values calculated by FRAP for transcription factors with similar molecular weight [214, 217]. The obtained average F_{free} reveals that HOXC13 is in equilibrium between the chromatin-bound (≈60%) and the freely diffusing/transiently interacting forms (≈40%), similarly to most of nuclear factors investigated by FRAP [188]. A relatively low value for K_{off} ($0.009 \pm 0.003\ \text{s}^{-1}$) was found that corresponds to a mean residence time on chromatin of 110±40 s.

For a protein with dynamics similar to the one found for HOXC13, characterized by both a high percentage of free protein and a slow dissociation from chromatin when compared to diffusion, Beaudoin *et al.* observed that the results of the half-nucleus fit, in a 2D approximation, are not optimal [214]. In particular, they propose to apply the model to fit only the average fluorescence of the 2nd and 5th slices of a series of 6 in which the nucleus is virtually dissected, in order to estimate the three binding parameters with better precision. Accordingly, we obtained the average recovery curve in the 5th slice (and loss in the 2nd slice) of the previously described analyzed nuclei, and applied our simplified model to calculate the results in this case as well. As shown in Figure 2.17B, we did not gain any improvement from this analysis: the three estimated binding parameters are in agreement with those obtained from the half-nucleus fit, but their spread is larger, in particular for the case of D_{eff} and F_{free}.

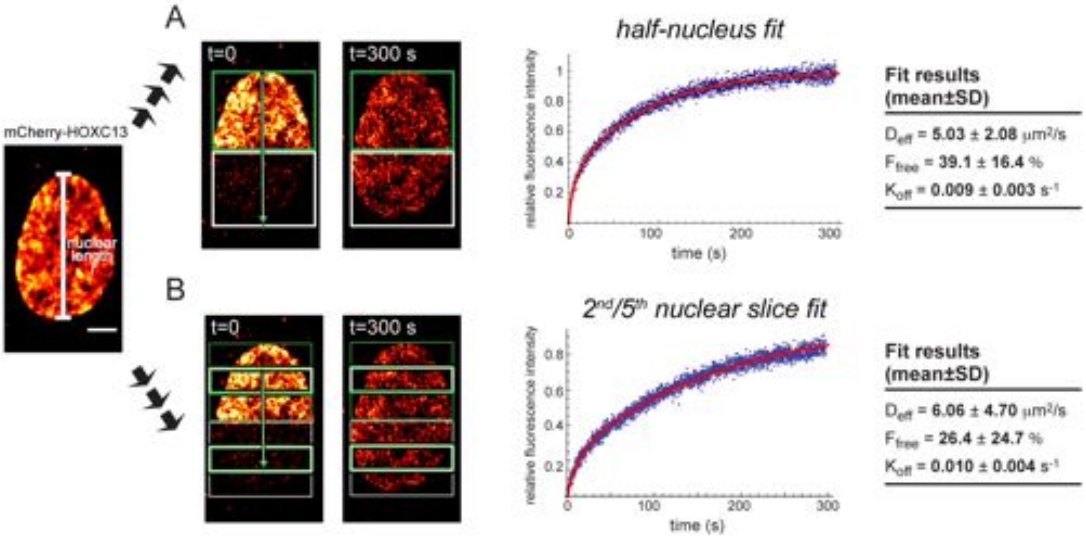


Figure 2.17 Quantitative analysis of wt mCherry-HOXC13 nuclear dynamics. FRAP curves obtained from cells expressing mCherry-HOXC13 were analyzed using a reaction-diffusion model, in which the nucleus is approximated to a parallelepiped with a defined nuclear length (left image). The average fluorescence in the whole bleached area (A) or in the 2^{nd} and 5^{th} of 6 slices dissecting the nucleus length (B) were fitted using the model. The two fits (red curve), superimposed to raw data (blue dots) of a representative cell, are reported in the two graphs. The fit results of all analyzed cells, for both cases, are reported on the Tables on the right. Scale bar: 5 µm.

Notably, both approaches revealed a clear dependence of the fit quality on the similarity between the real nuclear geometry of each analyzed cell and the approximated model geometry. Results reported here were obtained from a selection of the experiments which best fitted the model geometry (a total of 16 out of 40 analyzed cells). Selection criteria are reported in Appendix A of this thesis, together with the summary of the results on all cells which, in any case, does not change the conclusions obtained here, in particular concerning the HOXC13 residence time onto chromatin of the order of minutes.

2.4.5 Concluding remarks about HOXC13 nuclear dynamics

Although detailed considerations about this FRAP study will be presented in the Discussion section, some comments need to be done here to understand results reported in the following. Based on our FRAP data, a new scenario emerges for the role of HOXC13 that veers significantly from the commonly accepted dynamic vision of nuclear architecture and compartmentalization [187, 189], which has been presented in **paragraph 1.3.1**. In fact, we found that a large amount of HOXC13 is actually bound to chromatin very stably, with a mean residence time of the order of minutes. This means that, while the nucleoplasmic diffusion of the protein cannot be neglected, once the protein binds chromatin, it forms stable complexes and dissociates very slowly.

These considerations make HOXC13 different from most nuclear proteins, especially transcription factors [182, 188]. In the context of a possible role for HOXC13 in DNA replication regulation, it was interesting to find out that its nuclear dynamics is very different also from those reported for RC proteins. Indeed, ORC1, ORC4, and Cdt1 proteins were found to be very dynamically exchanging with chromatin, when investigated by FRAP [203, 221]. It is possible that this derives from the low sequence-specificity reported to characterize the DNA binding of these proteins [13, 17, 87]. Therefore, proteins like HOXC13, rather stably bound to chromatin, could be useful in the recruitment and stabilization of RC proteins at replication origin DNA. This would require a direct interaction between HOXC13 and proteins regulating replication origin activation. Accordingly, we explored the possible interactions of this homeotic protein with other proteins of the RC in living cells, which led to the results reported in the following **paragraphs 2.5-2.6**.

2.5 Cell-cycle related changes of sub-cellular localization of RC proteins in living cells

After we characterized the behaviour of fluorolabelled HOXC13 at early replicating chromatin in living cells, we applied a similar approach to study other selected RC proteins in comparison with HOXC13. Accordingly, RC proteins were tagged with FPs, similarly to what previously showed for HOXC13 (Figure 2.4). We created fusion constructs of four different RC proteins, namely ORC2, Cdc6, MCM3, and ORC1 with a green FP variant (Figure 2.18A) and checked by Western Blot their proper expression in living U2OS cells after immunoprecipitation with α-GFP antibody (Figure 2.18B). We found that ORC2 and MCM3 fusion constructs are expressed entirely; conversely, CDC6 and ORC1 fusion constructs display some degradation. The latter finding is consistent with the regulative degradation which the two respective endogenous proteins undergo throughout cell-cycle progression [22, 23, 49, 64, 223, 224]. However, in the case of ORC1-E^{0}GFP construct, the abundant bands at lower MW than the full-length construct could derive both from degradation of the construct and impurity of the immunoprecipitated sample.

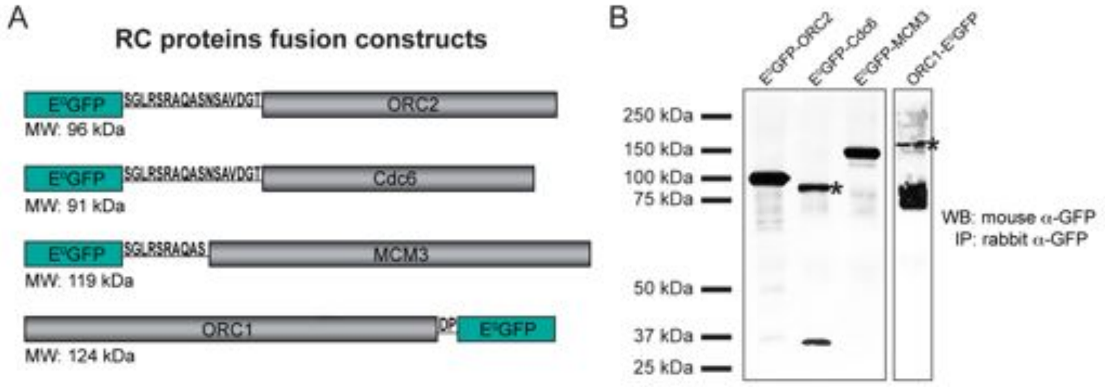

A
RC proteins fusion constructs
E⁰GFP
SGLRSRAQASNSAVDGT
ORC2
MW: 96 kDa
E⁰GFP
SGLRSRAQASNSAVDGT
Cdc6
MW: 91 kDa
E⁰GFP
SGLRSRAQAS
MCM3
MW: 119 kDa
ORC1
DP
E⁰GFP
MW: 124 kDa
B
E⁰GFP-ORC2
E⁰GFP-Cdc6
E⁰GFP-MCM3
ORC1-E⁰GFP
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
*
*
WB: mouse α-GFP
IP: rabbit α-GFP

Figure 2.18 Fluorolabelled RC proteins: design of fusion constructs and expression in human cells. A) Shematic picture of the prepared fluorolabelled RC proteins. The FP sequence was either fused at the N- (Cdc6, ORC2, MCM3) or at the C-terminus (ORC1) of the proteins. The linker peptide designed for the fusion is indicated by its aminoacidic sequence. B) Cell lysates from U2OS cells in turn transfected with the four constructs were immunoprecipitated using rabbit polyclonal α-GFP antibody. Protein was immunodetected using mouse monoclonal α-GFP antibody. The asterisks indicate the bands of full-length Cdc6 and ORC1 fusion constructs.

Thereafter we analyzed, by means of live imaging techniques, the dynamic changes of intracellular distribution of these proteins along the cell cycle, focusing in particular at the G1/S transition. This is indeed a fundamental step at which the cycle of origin activation/deactivation is regulated [23, 49, 53, 221, 225, 226], as reviewed in detail in **paragraph 1.1.3**.

Results obtained for each RC protein are reported separately in the four following **paragraphs 2.5.1-2.5.4**. The final **paragraph 2.5.5** deals with a comparison between results presented here and those concerning HOXC13 localization presented in the previous paragraphs.

2.5.1 ORC2

E^0GFP-ORC2 was expressed in HeLa and U2OS cells; in the latter cell line, it was expressed transiently as well as stably. In all these cases, the protein displayed a predominantly nuclear localization; interestingly, it exhibited various subnuclear configurations in asynchronous cells, spanning from highly concentrated at few aggregated structures to homogeneously diffused nuclear ORC2. Many intermediate distributions between these two were also represented (Figure 2.19).

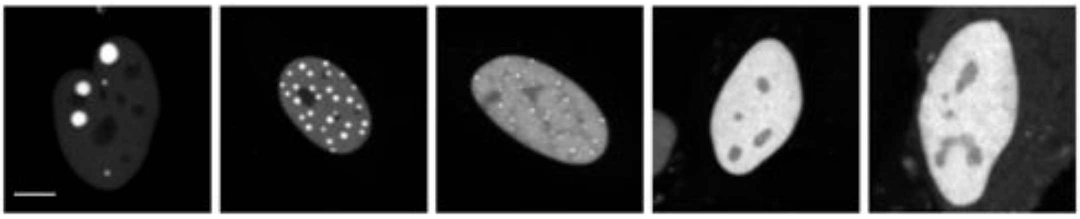


Figure 2.19 Subnuclear distributions of E^{0}GFP-ORC2 in human living cells. Representative images of the typical E^{0}GFP-ORC2 nuclear localizations in asynchronous U2OS cells. Scale bar: 5 μm.

Immunofluorescence analysis revealed that the focal, aggregated ORC2 organization relies on the association with heterochromatin (Figure 2.20), as revealed by a marked co-localization with HP1α (panel A) and another heterochromatin marker like PML [227] (panel B). Notably, both these immunofluorescence analyses required the use of detergent in the experimental procedures (see **paragraph 4.4**). This resulted in the almost complete removal of diffused nuclear mCherry-ORC2 (red panels of Figure 2.20). Thus, the diffused ORC2 likely corresponds to nucleoplasmic protein.

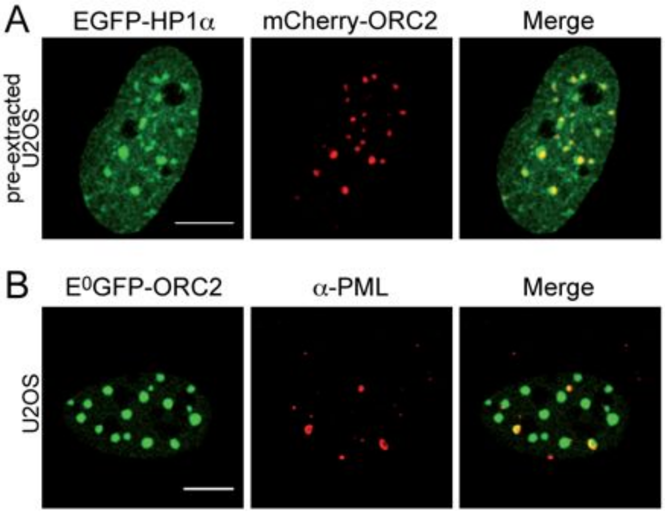


Figure 2.20 Heterochromatic nature of E^{0}GFP-ORC2 nuclear aggregation. The E^{0}GFP-ORC2 focal structure displayed in G1 phase (Figure 2.21), relies on the association with heterochromatin. A) GFP-HP1α and mCherry-ORC2 co-localize in the same foci (see the Merge image) in U2OS cells subjected to 0.5% TritonX-100 pre-extraction prior to fixation. B) E^{0}GFP-ORC2 foci partially co-localize with endogenous PML bodies (detected with a specific primary antibody in the presence of 0.1% Tween-20, further stained with an Alexa647-conjugated secondary antibody). Scale bar: 10 μm.

BrdUrd immunofluorescence experiments revealed that cells in S phase display only homogeneously diffused nuclear E^{0}GFP-ORC2, whereas the ORC2 focal distribution is displayed only by non-S cells (Figure 2.21A). Time-lapse imaging analysis of U2OS cells, transiently or stably expressing E^{0}GFP-ORC2, showed that the diffused distribution is actually a temporal consecution of the aggregated one: within 13 hours, few big foci are disassembled into smaller ones, which then dissolve until a final homogeneous nucleoplasmic distribution is reached (Figure 2.21B).

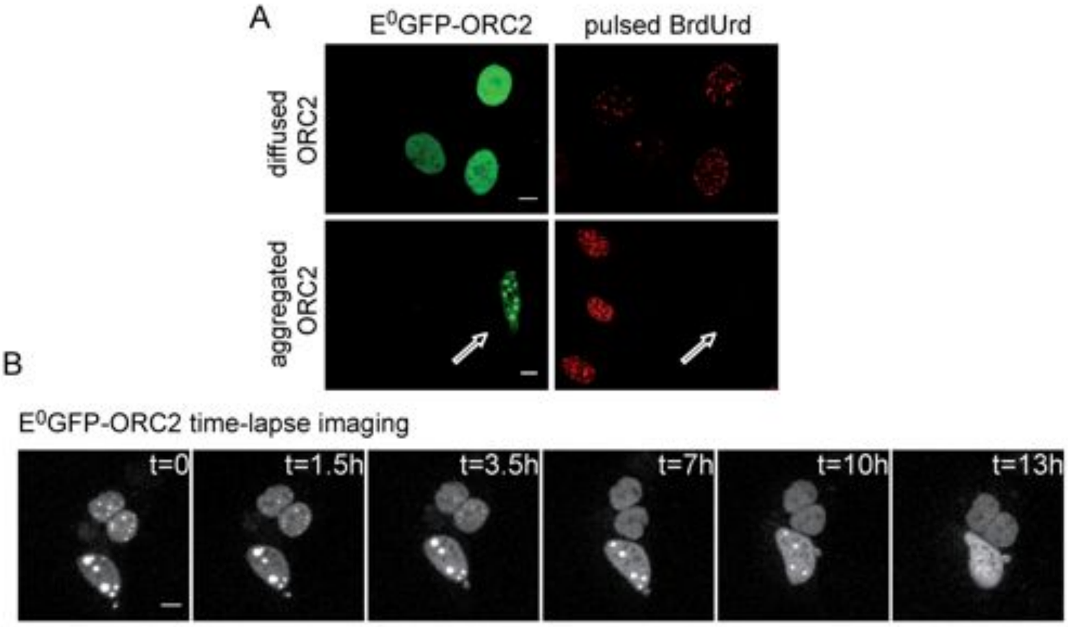


Figure 2.21 Dynamics of E^0GFP-ORC2 localization at the G1/S transition. A) Pulsed BrdUrd immunofluorescence shows that cells having a diffused nuclear E^0GFP-ORC2 localization are positively stained with BrdUrd, while cells displaying a nuclear focal structure for E^0GFP-ORC2 do not incorporate BrdUrd (white arrow). B) Time-lapse imaging reveals that the big E^0GFP-ORC2 nuclear foci are first disassembled into smaller ones, until a final nuclear diffused localization is reached within 13 hours (typical time observed with cells stably expressing the construct). Scale bar: 10 μm.

Taken together, these data depict the following spatio-temporal nuclear organization for ORC2 protein: in G1 phase, fluorolabelled ORC2 is tightly associated with heterochromatin, via a possibly direct interaction with HP1α protein occurring in defined nuclear foci. This interaction gradually fades, as the homogenization of ORC2 nuclear distribution is necessary for the cells to undergo S phase. These observations are definitely in agreement with data reported for the localization of endogenous ORC2 protein [53, 228, 229], thus providing evidence that the recombinant construct is functional and well-tolerated when expressed in human cells. The recent finding that human cells, when ORC2 is ectopically expressed, down-regulate the expression levels of endogenous ORC2 by a compensation mechanism [230], further support our results and encouraged the use of this fluorolabelled ORC2 construct.

2.5.2 Cdc6

E^0GFP-Cdc6 was expressed in U2OS cells. Similarly to ORC2, when investigated in asynchronous cells, also Cdc6 displayed various localizations, in this case spanning from nuclear to cytoplasmic, with various intermeditate distributions (Figure 2.22). Notably, all these subcellular localizations shared an intense staining of nucleoli, which likely results from a regulatory mechanism to avoid accumulation of the overexpressed protein at replicating chromatin [231].

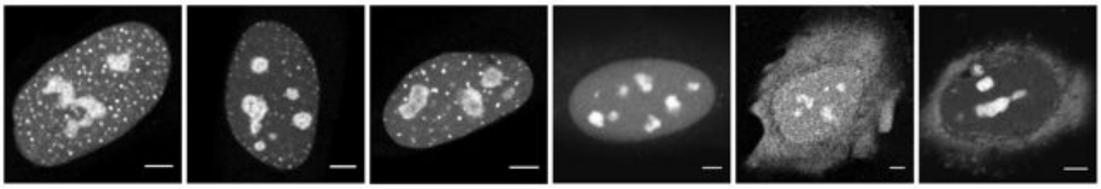


Figure 2.22 Subcellular distributions of E^0GFP-Cdc6 in human living cells. Representative images of the typical E^0GFP-Cdc6 nuclear localizations in asynchronous U2OS cells. Scale bar: 5 μm.

BrdUrd immunofluorescence experiments revealed that cells in S phase display only the cytoplasmic E^0-Cdc6 localization (Figure 2.23A); instead, E^0-Cdc6 is nuclear in non-S phase cells. Time-lapse imaging of U2OS cells expressing E^0-Cdc6 showed that the nuclear and cytoplasmic localizations are temporally related: nuclear E^0-Cdc6 is indeed gradually excluded to cytoplasm, typically within 15-20h observation, depending on the expression level of the construct (Figure 2.23B).

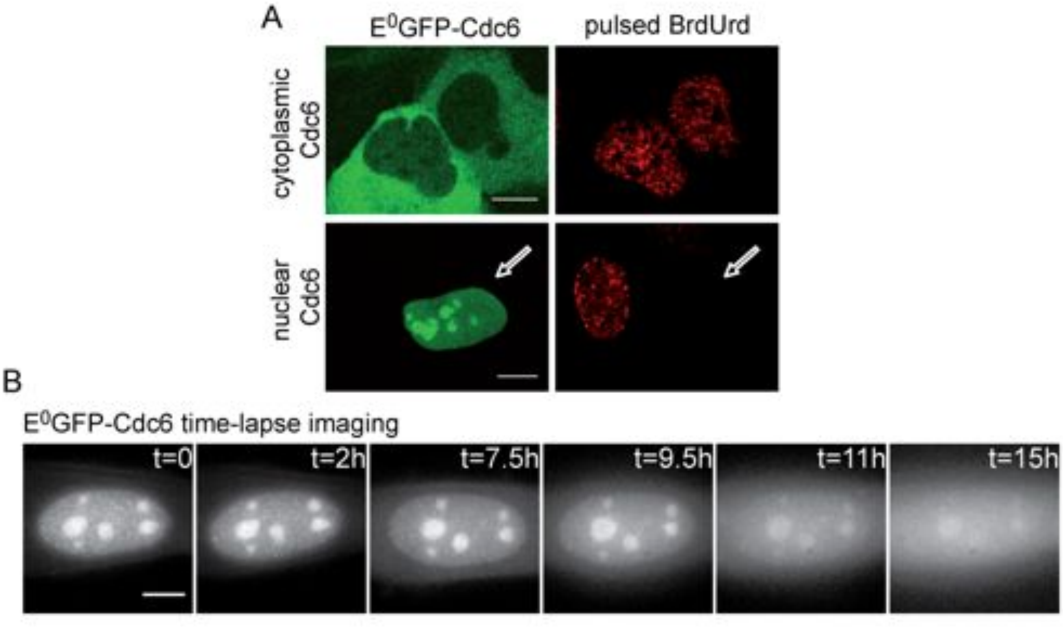


Figure 2.23 Dynamics of E^0GFP-Cdc6 localization at the G1/S transition. A) Pulsed BrdUrd immunofluorescence shows that cells having a diffused cytoplasmic E^0GFP-Cdc6 localization are

positively stained with BrdUrd, while cells displaying nuclear E^0GFP-Cdc6 do not incorporate BrdUrd (white arrow). B) Time-lapse imaging reveals that nuclear E^0GFP-Cdc6 is gradually exported to cytoplasm, typically within 15 hours (mean time of several experiments performed with cells with low-expression profile of the construct). Scale bar: 10 μm.

The combination of BrdUrd immunofluorescence and time-lapse imaging data allow to define the following spatio-temporal dynamic distribution for Cdc6 protein: in G1 phase, fluorolabelled Cdc6 is in the nucleus to exert its function in replication origin activation. Once origins are ready to fire, which corresponds to the switch to S phase, Cdc6 is excluded from nucleus to cytoplasm. This is certainly a regulatory mechanism of the G1 to S transition to avoid re-replication, i.e. the occurrence of multiple rounds of replication per cell-cycle (see Figure 1.4 at **paragraph 1.1.3**). These observations agree with data reported for the localization of endogenous Cdc6 protein [64, 224, 231, 232]. Infact, endogenous Cdc6 has been reported (albeit not univocally [233]) to undergo concerted events of phosphorylation [223] and acetylation [225], that determine its export to cytoplasm before the start of S phase.

2.5.3 ORC1

In order to study the localization of ORC1-E^0GFP construct in human cells, no BrdUrd experiments were needed, as the ORC1 protein is degraded before S phase entry [23, 48, 49], so that cells positively stained with BrdUrd or PCNA are negatively stained with ORC1 protein [234]. Therefore, we limited the analysis of this protein to the G1 phase of the cell-cycle. Unfortunately, when transfected in U2OS cells, ORC1-E^0GFP construct not only displayed low expression efficiency and probable degradation (see Figure 2.18), but also caused relevant cell-toxicity and death. This hampered a clear analysis of its localization patterns. Nevertheless, two distinct nuclear localizations were observed for ORC1-E^0GFP construct (Figure 2.24): one displayed densely spotted ORC1 throughout the nucleus (left panel), the other one showed less concentrated ORC1 on nuclear periphery and in nucleoli (two panels on the right); intermediate phenotypes were also observed.

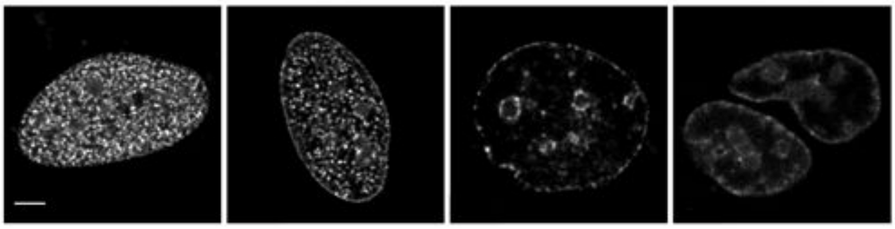


Figure 2.24 Subnuclear distributions of E^0GFP-ORC1 in human living cells. Representative images of the typical E^0GFP-ORC1 nuclear localizations in asynchronous U2OS cells. Scale bar: 5 μm.

Although we argued that these two nuclear patterns could be temporally related in the context of cell-cycle progression, no clear interdependence between them was found by time-lapse imaging of cells expressing ORC1-E^{0}GFP (i.e. in G1): within 20-24h observation, no evident subnuclear re-distribution of the protein could be detected, and often apoptosis occurred in the observed cells (data not shown). This could be explained by the fact that overexpressed ORC1 is not well tolerated by the cells and displays altered chromatin-binding properties when compared to its endogenous counterpart, as already observed for this protein [235].

2.5.4 MCM3

E^{0}GFP-MCM3 was expressed in U2OS cells. Differently from what observed for ORC2 (paragraph 2.5.2) and Cdc6 (paragraph 2.5.3), when investigated in asynchronous cells, fluorolabelled MCM3 displayed a unique, unchanging, homogeneous nuclear staining (Figure 2.25).

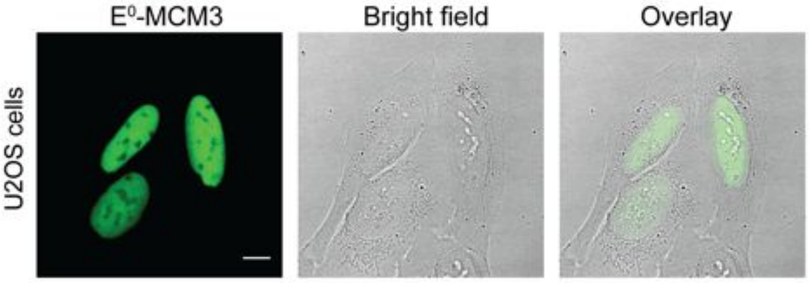


Figure 2.25 Subnuclear distribution of E^{0}GFP-MCM3 in human living cells. E^{0}GFP-MCM3 localization in asynchronous U2OS cells was diffused in the nucleus, as evident from the overlay of the fluorescence channel with the bright field image. Scale bar: 10 µm.

These data agree with previous observations in *S. cerevisiae* that MCM proteins are present in excess over the effective number of replication origins: only a small fraction of MCM tightly associates with chromatin, from late M phase to S phase, while the redundant rest of them are distributed to both nucleoplasm and cytoplasm in relatively constant levels throughout the cell cycle [236]. The same cell-cycle dependent chromatin loading of MCM proteins has been reported from studies in human cells [21]. Actually only pre-extraction procedures were found to successfully unveil the chromatin-bound localization patterns of MCM proteins hidden under the nucleoplasmic excess of protein in human cells [56].

2.5.5 Concluding remarks about RC proteins vs. HOXC13

The data reported in the previous **paragraphs 2.1-2-3** show that HOXC13, both in the endogenous and recombinant form, has an exclusively nuclear localization (Figures 2.1, 2.3, 2.4) and does not undergo significant changes of intracellular distribution throughout cell cycle (data not shown); this is probably a result of the protein stable association with chromatin demonstrated by FRAP (Figures 2.15 and 2.16). Strikingly, as shown in the last four paragraphs, we found that this was not the case for other RC proteins, at least for ORC2, Cdc6 and ORC1. We indeed found distinct distributions of these proteins throughout the nucleus and also cytoplasm; for ORC2 and Cdc6, these changes of localization were clearly related to the G1/S transition.

The differential localizations of these RC proteins throughout the cell-cycle imply a differential co-localization of them with HOXC13, which is instead always nuclear. As concerns ORC2, we found that its aggregated structure typical of G1 phase (see Figure 2.21) occurs in nuclear areas excluded from HOXC13 nuclear compartments; therefore these two proteins preferentially co-localize at the end of G1 and beginning of S phase, when ORC2 protein diffuses out of the heterochromatic foci to the nuclear areas where HOXC13 is (Figure 2.26).

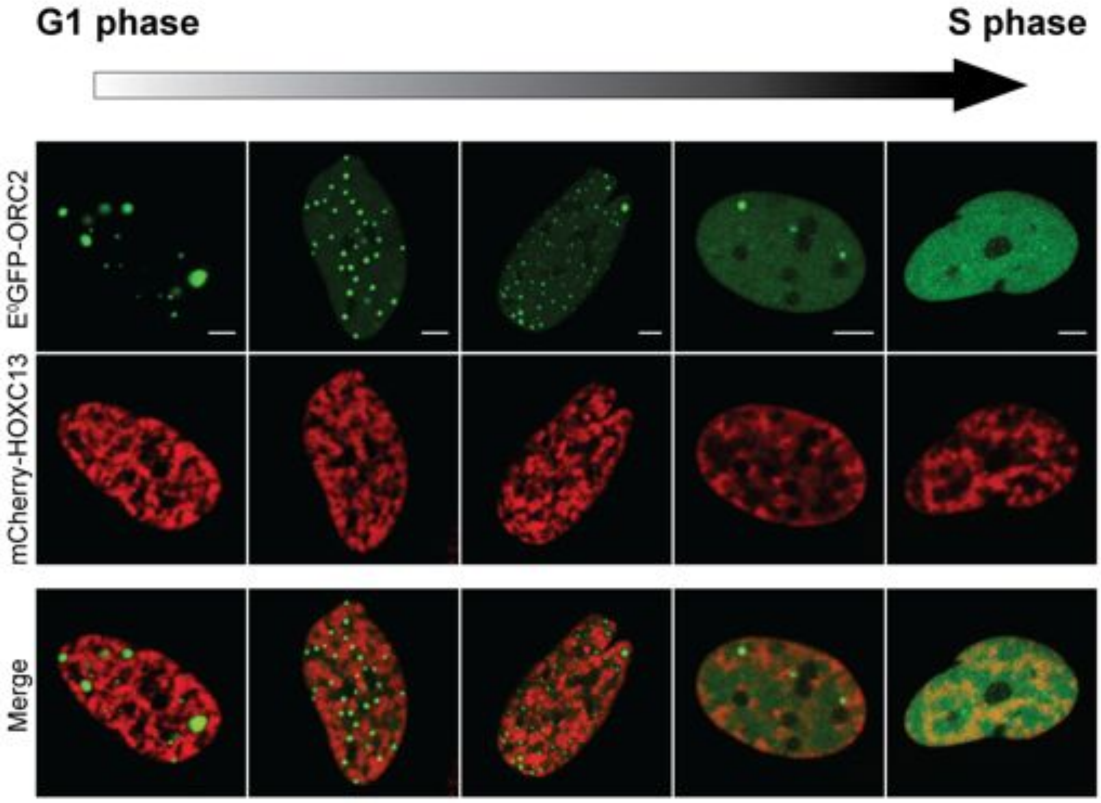


Figure 2.26 ORC2/HOXC13 co-localization in human living cells. Representative images of the mutual distributions of E^0-ORC2 (green channel) and mCherry-HOXC13 (red channel) proteins in asynchronous U2OS cells. Direction of cell-cycle progression, based on E^0-ORC2 localization (Figure 2.21), is represented by the top arrow. Scale bar: 5 μm.

As for Cdc6, the co-localization of this protein with HOXC13 is restricted to G1, the only phase when these proteins are in the nucleus, before Cdc6 is exported to cytoplasm in S phase (Figure 2.27). Notably the same Figure shows that nuclear E^0-Cdc6, besides nucleolar addensation, displayed a peculiar focal distribution resembling that observed for replication foci in early S phase (see Figure 2.7). These Cdc6 foci are present in the nucleus during G1 phase, and therefore are not replication foci. Nevertheless, similarly to what already shown for replication foci (see paragraph 2.3), HOXC13 displayed a strong co-localization with Cdc6 nuclear foci, when these were ubiquitously distributed in the nucleus.

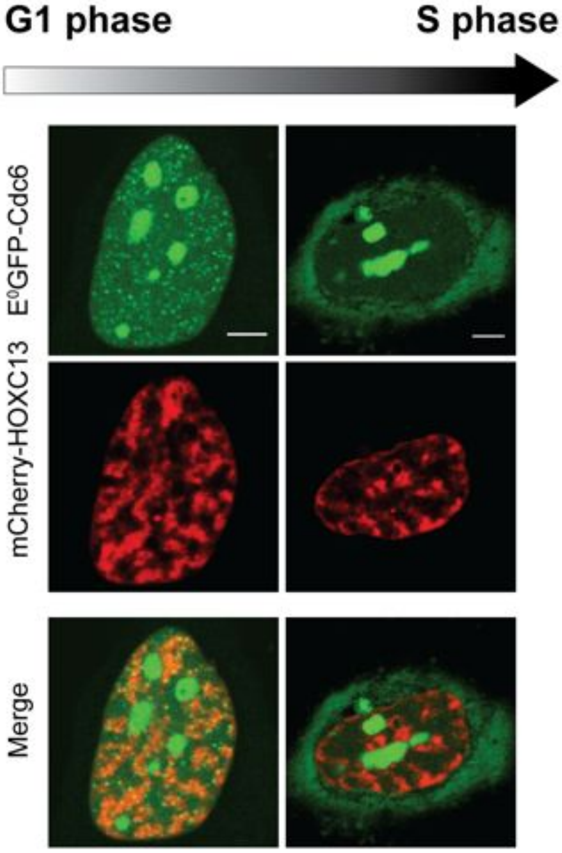


Figure 2.27 Cdc6/HOXC13 co-localization in human living cells. Representative images of the mutual distributions of E^0-Cdc6 (green channel) and mCherry-HOXC13 (red channel) proteins in asynchronous U2OS cells. Direction of cell-cycle progression, based on E^0-Cdc6 localization (Figure 2.23), is represented by the top arrow. Scale bar: 5 µm.

As concerns ORC1, neither nuclear localization reported for the fluorolabelled protein when investigated in asynchronous U2OS cell population (see Figure 2.24) displayed a significant co-localization with mCherry-HOXC13 (Figure 2.28).

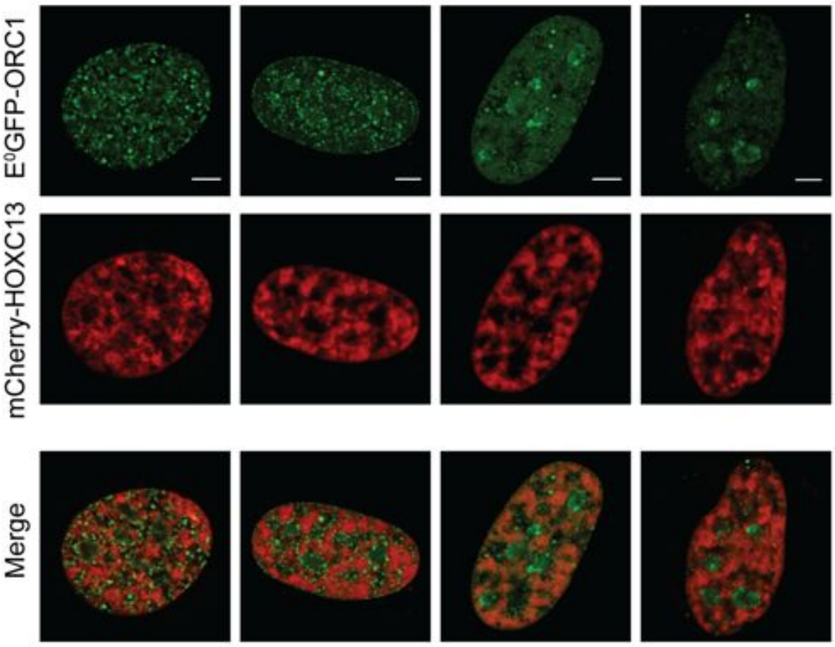


Figure 2.28 ORC1/HOXC13 co-localization in human living cells. Representative images of the mutual distributions of ORC1-E^0 (green channel) and mCherry-HOXC13 (red channel) proteins in asynchronous U2OS cells. Scale bar: 5 μm.

Finally, with regard to MCM3 protein, as this constantly displayed homogeneous nuclear staining, it was always found to co-localize with HOXC13 independently from cell-cycle progression (Figure 2.29).

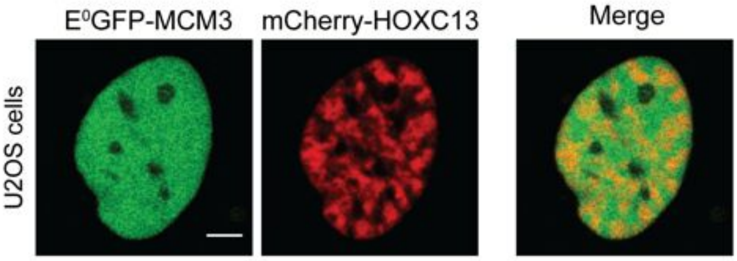


Figure 2.29 MCM3/HOXC13 co-localization in human living cells. Representative images of the mutual distributions of E^0-MCM3 (green channel) and mCherry-HOXC13 (red channel) proteins in asynchronous U2OS cells. Scale bar: 5 μm.

We wondered at this point whether the above described evidence of co-localization between HOXC13 and some RC proteins corresponds to an actual vicinity of the proteins in living cells (see Figure 1.17A). In order to sharpen the *in vivo* protein-protein interaction analysis in the very crowded context of the nucleus, we utilized a

Fluorescence Resonance Energy Transfer (FRET) approach to detect interactions within the nanometre scale (≤10nm), as will be presented in detail in the following paragraph.

2.6 *In vivo* detection of HOXC13 interaction with RC proteins

The possible interactions of HOXC13 with RC proteins were investigated in living cells by using Fluorescence Resonance Energy Transfer (FRET). This approach allows to detect in living cells only short-range interactions between two proteins of interest once they are tagged with an appropriate couple of FP variants, one acting as donor and one as acceptor (see **paragraph 1.3.1** and Appendix B of this thesis). FRET was detected by monitoring the reduction of the donor fluorescence lifetime in the presence of a close (≤10nm) acceptor, as can be achieved on a pixel-by-pixel basis by Fluorescence Lifetime Imaging Microscopy (FLIM) [193].

We performed accordingly the FLIM measurements by using E^0GFP [237, 238] and mCherry [170] as donor and acceptor, respectively. A summary of FRET/FLIM microscopy is reported in Appendix B, together with the description of the photophysic properties of the FRET-pair chosen in our study. Here it is just worth to recall that the optimized E^0GFP/mCherry FRET-pair shows a negligible donor/acceptor cross-talk when compared to other GFP-based FRET pairs; moreover, it features a rather long, mono-exponential donor fluorescence decay when excited at 403nm (3.01 ns *in vivo*). Therefore, this choice allowed us to minimize the spectral bleed-through issue and to simplify the fitting procedures in FLIM analysis.

In our experiments, the E^0GFP donor was fused to Cdc6, ORC2, ORC1 and MCM3 proteins (see Figure 2.18): these were in turn co-expressed in U2OS cells with HOXC13 fused to the mCherry acceptor. FLIM measurements were performed and the results so obtained will be presented in the next paragraphs according to the following order:

First. From the comparison between the mean donor lifetime of E^0GFP-RC protein alone or expressed with mCherry-HOXC13, we present which of the probed RC proteins does interact with HOXC13 in living cells (**paragraph 2.6.1**);

Second. By analyzing the donor lifetime maps obtained by FLIM analysis, we identify the nuclear compartments were the above mentioned protein-protein interactions occur (**paragraph 2.6.2**);

Third. Considering the cell-cycle dependent localizations of the probed proteins previously described (**paragraphs 2.3-2.5**), we define the cell-cycle intervals when these interactions occur (**paragraph 2.6.3**);

Fourth. We finally provide a biochemical test of the affinity between the proteins involved in the above mentioned *in vivo* interactions (**paragraph 2.6.4**).

2.6.1 FLIM detection of HOXC13-RC interactions in living cells

Figure 2.30 reports, on the left side of each panel, four sets of representative images in which mCherry-HOXC13 was co-expressed with E^0GFP fusions of nuclear Cdc6 (A), ORC2 (B), MCM3 (C) and ORC1 (D).

To obtain lifetime values, E^0GFP was excited with a pulsed 403 nm laser, fluorescence decay was recorded and fitted in all nuclear pixels; the result was a distribution curve of the frequency of nuclear pixels vs. mean donor lifetime (τ_m). Cumulative τ_m distribution curves were obtained based on all analyzed cells expressing in turn E^0GFP-Cdc6, E^0GFP-ORC2, E^0GFP-MCM3, ORC1-E^0GFP with mCherry-HOXC13 (red curves in the graphs of Figure 2.30 A, B, C and D respectively). In the same graphs, the blue curve is the τ_m distribution of cells expressing the respective E^0GFP-RC protein alone; the black dashed area represents the negative control, i.e. an estimate of unspecific FRET signal between donor and acceptor when they are not supposed to interact. This was obtained either tranfecting E^0GFP-NLS$_{SV40}$ with mCherry-HOXC13 or E^0GFP-pre-RC proteins with untagged mCherry.

The graphs show that, when expressed alone, all E^0GFP-RC proteins display a similar average lifetime in the nucleus (from 2.97 to 3.03 ns, see also the first row of Table 2.2 at the end of the paragraph): these values are very close to those reported for untagged E^0GFP [237]. Conversely, when co-expressed with mCherry-HOXC13, E^0GFP-Cdc6 and E^0GFP-ORC2, but not E^0GFP-MCM3 and E^0GFP-ORC1, display a significant τ_m reduction when compared either to the donor alone, or to the negative control curves (Figure 2.30 and Table 2.2, second and third rows). Thus, we can conclude that Cdc6 and ORC2 do interact with HOXC13 by analysis in living cells, while this does not seem to be the case for MCM3 and ORC1 proteins. In detail, by calculation of the FRET efficiency for the former samples (explained in Appendix B), we finally estimated that the actual distance *in vivo* between ORC2 and HOXC13 is 6.40±0.04 nm, while that between Cdc6 and HOXC13 is 6.72±0.07 nm.

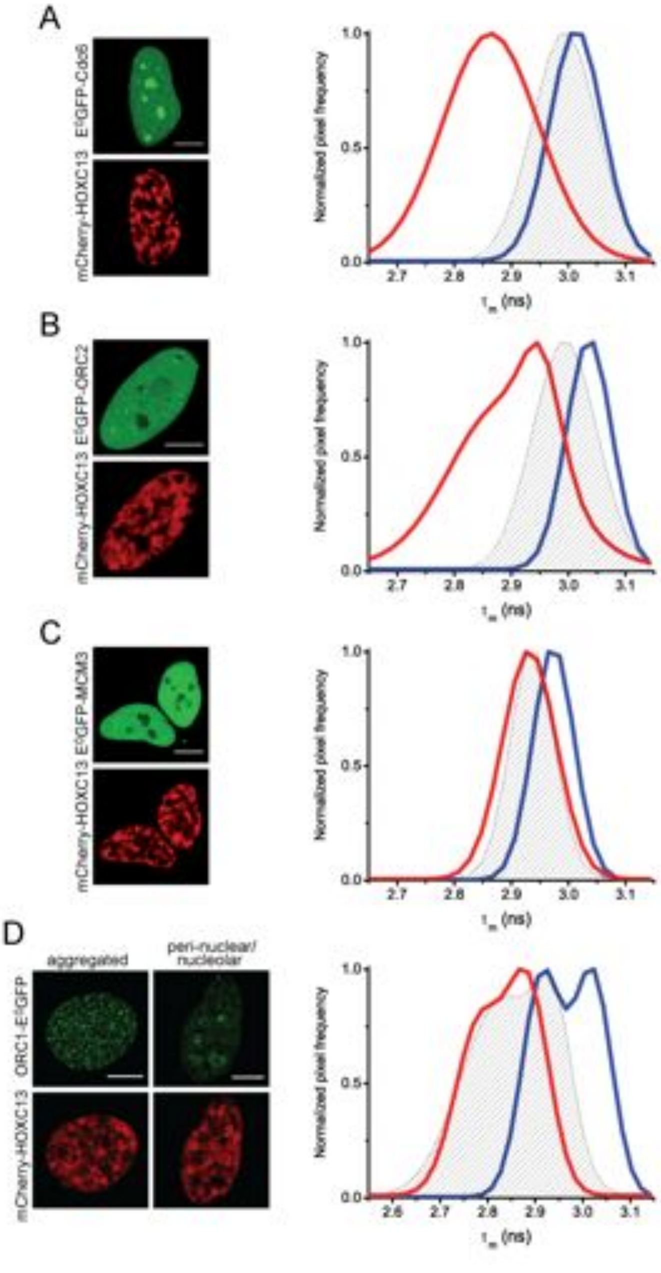


Figure 2.30 FLIM analysis of HOXC13 interaction with RC proteins. mCherry-HOXC13 was co-expressed with E^{0}GFP-fusions of Cdc6 (A), ORC2 (B), MCM3 (C) and ORC1 (D) in living cells. Left side of each panel: representative fluorescence-intensity images of the cells used in the FLIM study. In the graphs of each panel: plot of the frequency of nuclear pixels vs. the mean donor lifetime (τ_m) calculated for each pixel. Blue curve: τ_m distribution of cells expressing the respective RC protein alone. Dashed-black area: negative control, coming from the average of the two negative controls indicated in the main text in A, B, C panels. In panel D, only ORC1-E^{0}GFP expressed with untagged mCherry was considered as negative control. Red curve: τ_m distribution for cells co-expressing the respective RC protein with HOXC13. All reported curves are cumulative sum-distribution data relative to all analyzed cells (see Table 2.2); the pixel frequency is normalized to the peak of each distribution curve. Scale bar: 10 μm.

A particular comment needs to be done for the case of ORC1/HOXC13 interaction (refer to panel D in Figure 2.30). In this case, both the two ORC1 localizations, aggregated and perinuclear (see Figure 2.24) were tested for the interaction with HOXC13. FLIM analysis of an equal amount of the two phenotypes resulted in a double peaked lifetime distribution for the donor alone (blue curve) as well as for the donor with acceptor (dashed area and red curve). In all three curves, the peaks at lower lifetime correspond to the more aggregated ORC1 phenotype only (left image of panel D). It is possible that, under the same acquisition set-up, the aggregated phenotype can lead *per se* to a lower lifetime with respect to the other perinuclear/nucleolar one, which is actually less concentrated. This artifact could also be responsible of the 2-fold increased unspecific FRET signal detected for the negative control of ORC1-E^0GFP, when compared to the other probed proteins (second row of Table 2.2). It must be underlined that this wide range of unspecific interaction compromises any further consideration on the lifetime distribution curve obtained for ORC1 in the presence of HOXC13.

Table 2.2 Mean lifetime values obtained with FLIM analysis. Each column corresponds to the RC protein indicated at the top. Reported values are τ_m (±SEM) obtained for the respective E^0GFP-RC protein alone (first row), the negative control (second row) or the E^0GFP-RC protein expressed with mCherry-HOXC13 (third row). The row of the negative control comes from the average of the two negative controls indicated in the main text for Cdc6, ORC2 and MCM3 columns. For ORC1 protein, only ORC1-E^0GFP expressed with untagged mCherry was considered as negative control. The number of cells (n) used for the statistical analysis is reported below each corresponding τ_m value.

Table 2.2	Cdc6	ORC2	MCM3	ORC1
E^0GFP-RC protein alone	3.00±0.04 ns n=49	3.03±0.04 ns n=51	2.97±0.04 ns n=46	2.96±0.05 ns n=62
Negative control	2.97±0.04 ns n=85	2.97±0.04 ns n=116	2.95±0.04 ns n=76	2.84±0.05 ns n=48
E^0GFP-RC protein + mCherry-HOXC13	2.86±0.04 ns n=106	2.86±0.06 ns n=64	2.91±0.05 ns n=36	2.83±0.04 ns n=72

2.6.2 Spatial definition of HOXC13-RC interactions in the nucleus

A lifetime value was calculated for all nuclear pixels of the analyzed cells. In this way we generated donor-lifetime maps at sub-cellular level (Figure 2.31). These τ_m maps are superimposed onto the corresponding donor-intensity image (gray-scale image on the left side of each lifetime map). Pseudocolours represent lifetime values: as indicated in the colour calibration bar, red-shifted colours correspond to lower lifetime values and indicate FRET occurrence. Indeed, E^0GFP fused to Cdc6 and ORC2 displays a red-shifted lifetime map in presence of HOXC13 (panel B), when compared to the maps of the corresponding donor alone (panel A). Conversely, the lifetime map of MCM3 co-expressed with HOXC13 (panel B) retains a blue-shifted colour similar to that of the donor alone (panel A).

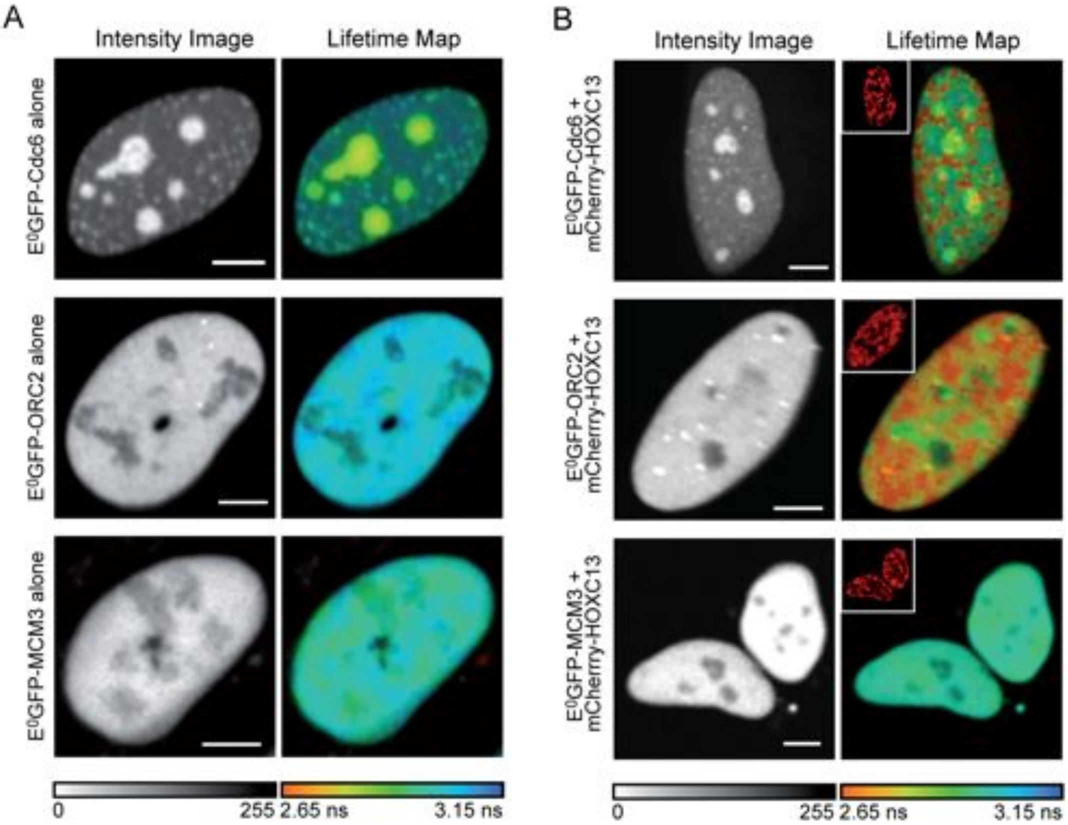

A
Intensity Image
Lifetime Map
E0GFP-Cdc6 alone
E0GFP-ORC2 alone
E0GFP-MCM3 alone
0
255
2.65 ns
3.15 ns
B
Intensity Image
Lifetime Map
E0GFP-Cdc6 + mCherry-HOXC13
E0GFP-ORC2 + mCherry-HOXC13
E0GFP-MCM3 + mCherry-HOXC13
0
255
2.65 ns
3.15 ns

Figure 2.31 Analysis of the donor lifetime maps obtained by FLIM measurements. Representative donor lifetime maps (right) and corresponding intensity image (left), obtained for Cdc6 (up), ORC2 (middle), and MCM3 (bottom) in the absence (panel A) or presence (panel B) of HOXC13. There are two calibration bars: that in gray-scale refers to fluorescence intensity images, that in pseudocolour refers to lifetime maps. Insets in the lifetime maps of panel B represent the fluorescence intensity image of mCherry-HOXC13 (same as Figure 2.30, panel A, B, C, respectively). Scale bar: 10 µm.

Notably, the red areas in Cdc6 and ORC2 lifetime maps are super-imposable to the areas of HOXC13 localization. Thus, we can conclude that the interaction of both Cdc6 and ORC2 with HOXC13 occurs in the same nuclear regions where the homeodomain binds to chromatin (Figures 2.11, 2.15 and 2.16). Therefore, the interactions of

HOXC13 with the RC do not occur limitedly at the one or few origins to which the homeotic protein was demonstrated to bind *in vitro* [101], but rather appear to have a general relevance in the context of the nucleus.

It must be mentioned that, particularly for the case of Cdc6, some reduction of donor lifetime could be detected in nucleoli (see the green/yellow colour of these subcellular structures in the lifetime maps of E^0-Cdc6 in Figure 2.31 A and B). This does not significantly affect the final lifetime distribution, however, since these areas are relatively small, when compared to the whole nuclear areas, and excluded from HOXC13 characteristic nuclear compartments and therefore from the FRET efficiency calculation (see Figures 2.27 in this paragraph and Figure B.3 in Appendix B).

2.6.3 Temporal definition of HOXC13-RC interactions along cell-cycle progression

The localization of fluorolabelled HOXC13 and RC proteins, as well as their dynamics within cell-cycle progression, have been previously shown in detail (paragraphs 2.3, 2.4 and 2.5). These observations were particularly useful to establish the specific cell-cycle intervals in which Cdc6 and ORC2 interactions with HOXC13 occur *in vivo*.

The interaction of HOXC13 with Cdc6 is restricted to G1, the only phase when these proteins co-localize in the nucleus (see Figure 2.27). Indeed, as specified at the beginning of paragraph 2.6.2, only cells displaying nuclear E^0GFP-Cdc6, i.e. in G1 phase, were used for the FRET-FLIM study.

As for the interaction with ORC2, the lifetime distribution of E^0GFP-ORC2 co-expressed with mCherry-HOXC13 (Figure 2.30B, enlarged in Figure 2.32) displays two distinct peaks, one close to the negative control (~2.95 ns), the other one shifted towards lower-lifetime values (~2.82 ns). We argue that this stems from the different behaviour of cells in different moments of the cell-cycle that are simultaneously present when the interaction is investigated in asynchronous population. This could be expected from the fact that many different E^0GFP-ORC2 configurations were tested, ranging from highly aggregated to completely homogeneous nuclear ORC2, and not all of them displayed the same co-localization pattern with mCherry-HOXC13 (see Figure 2.26).

In order to verify this hypothesis, cells were synchronized at the G1/S border. Resulting FLIM data are shown as a green, dashed curve in Figure 2.32. Upon aphidicolin block, an inversion of the amplitude of the two lifetime peaks was observed, as a consequence of an increase of the cell population in which ORC2 interacts with HOXC13. Thus, the maximum of direct interaction between ORC2 and HOXC13 proteins occurs in late G1, before the start of S.

Considering that the homeotic protein is present at replication foci of early S (paragraph 2.3), it is not surprising that its interaction with ORC2 and Cdc6 turns out to be maximized at this cell-cycle stage.

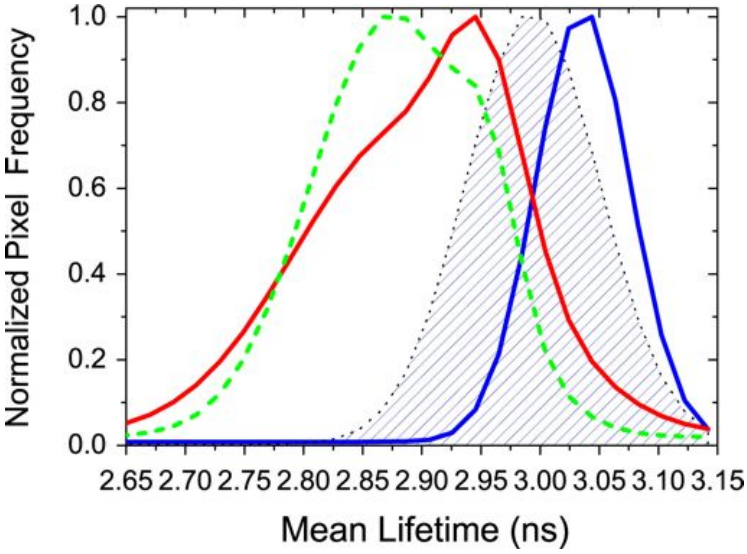


Figure 2.32 FLIM analysis of ORC2/HOXC13 interaction resolved within cell-cycle progression. The graph is a magnification of the graph reported in Figure 2.30B, to which a supplemental green curve is added. The blue, dashed-black, red curves are the same as Figure 2.30. Green, dashed curve: cumulative τ_m distribution for cells (n=43) co-expressing E^0GFP-ORC2 with HOXC13 after 24h aphidicolin treatment.

2.6.5 Biochemical test of HOXC13-RC proteins affinity

Results reported in the previous three paragraphs show that HOXC13 interacts *in vivo* with both Cdc6 and ORC2 in G1, before the start of S phase. In order to verify if such interactions rely on a specific affinity between the analyzed proteins, we performed co-immunoprecipitation (co-IP) assays from lysates of cells expressing the probed recombinant proteins. Immunoprecipitations with an anti-GFP antibody confirmed the affinity of mCherry-HOXC13 for E^0GFP-Cdc6 and E^0GFP-ORC2, but not for E^0GFP-NLS$_{SV40}$ (Figure 2.33A). Experiments with an antibody against GFP-HOXC13 demonstrated also an affinity of the fluorolabelled homeotic protein for endogenous ORC2 protein (Figure 2.33B). These experiments were performed in asynchronous cell cultures. Synchronization of transfected cells in G1 phase was also performed before the co-immunoprecipiation assays, but no relevant difference between the presented results were obtained (data not shown). Indeed, both cases revealed a clear dependence of the co-IP yield on the quantity of immunoprecipitated material, so that the variability observed experiment-by-experiment was greater than the difference between the experiments performed in asynchronous and synchronized cells.

Overall, we were able to assess that the interactions detected *in vivo* as described in **paragraph 2.6.1** correspond to a specific affinity between HOXC13 and RC components detected by canonical *in vitro* biochemical assays.

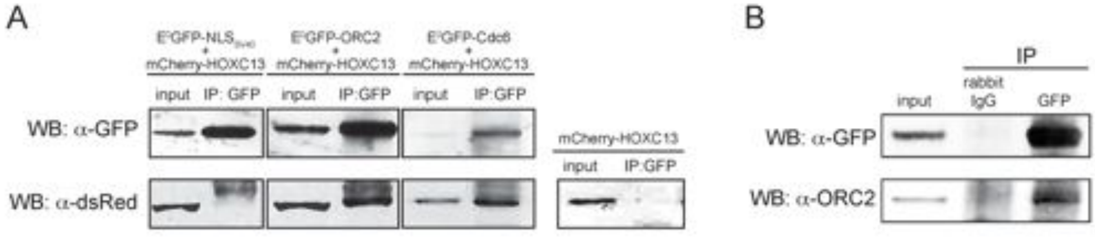

A
E⁰GFP-NLS
+
mCherry-HOXC13
E⁰GFP-ORC2
+
mCherry-HOXC13
E⁰GFP-Cdc6
+
mCherry-HOXC13
input
IP: GFP
input
IP:GFP
input
IP:GFP
WB: α-GFP
WB: α-dsRed
mCherry-HOXC13
input
IP:GFP
B
IP
input
rabbit IgG
GFP
WB: α-GFP
WB: α-ORC2

Figure 2.33 *In vitro* HOXC13 affinity for Cdc6 and ORC2 detected by co-immunoprecipitation. A) First three columns on the left: co-immunoprecipitation experiments performed with lysates from asynchronous U2OS cells (input) transfected with the constructs indicated on the top of the lanes. The IP was performed with an α-GFP antibody; both E^0GFP and mCherry constructs were immunodetected (top and bottom lanes, respectively). The fourth column refers to a cross-reactivity control experiment: in lysates from cells expressing only mCherry-HOXC13, no protein could be detected by α-dsRed antibody after immunoprecipitation with α-GFP antibody. Thus, α-GFP antibody used in the IP does not recognize mCherry protein. C) Immunodetection of endogenous ORC2 after co-immunoprecipitation with exogenous GFP-tagged HOXC13 in transiently transfected asynchronous U2OS cells (input). Each co-IP was repeated at least twice.

3. DISCUSSION

In this work we have explored, by means of live-cell imaging techniques, the spatial and temporal dynamics of the interaction of HOXC13 with the replication factories as well as with the proteins of the RCs. The results indicate that HOXC13 is a rather stable component of chromatin, confined by the homeodomain in nuclear areas where early S DNA replication occurs; in addition it interacts with members of the RC in coincidence with origin activation and the interaction appears to be of general nature in the context of DNA replication regulation. We therefore propose that HOXC13 participates in origin specification, as will be discussed in detail in the following paragraphs.

3.1 HOXC13 is a stable chromatin component

This work provides an unprecedented in-depth study of the nuclear localization and of the chromatin binding properties of a homeotic protein, namely HOXC13, in human cells. We have obtained unambiguous evidence that HOXC13 is a stable chromatin component. This statement stems from several observations. As expected for a transcription factor, HOXC13 displays an almost exclusively nuclear localization in human cells, either as endogenous protein (Figure 2.1) or as fluorolabelled recombinant construct (Figure 2.4). Its peculiar speckled-like nuclear distribution relies on the presence of the homeodomain: indeed, a deletion mutant devoid of the homeodomain displays homogeneous nuclear staining only after it is forced to achieve nuclear localization, as this property is lost upon homeodomain removal [136] (Figure 2.11). Interestingly, HOXC13 detection in the nucleus is resistant to detergent pre-extraction procedures (Figures 2.2 and 2.5), which are typically used to uncover the components of chromatin and/or nuclear matrix from the excess of nucleoplasmic proteins [21, 56, 206, 239, 240]. Based on our U2OS cells biochemical fractionation experiment (Figure 2.3), we can assert that HOXC13 is actually a chromatin, and not a nuclear matrix, component. Indeed, nuclear matrix is biochemically defined as the nuclear non-chromatin structure that is resistant to nuclease digestion followed by high-salt treatment [241, 242]. HOXC13, although resistant to high salt extraction (it is detected in chromatin extractions performed with up to 2 M salt concentration), is not resistant to

DNase treatment (this indeed changes the protein solubility, so that it can be extracted at salt concentrations ≤300 mM).

However, most of our understanding of the chromatin-binding properties of HOXC13 comes from the FRAP data obtained in living cells. These confirm that HOXC13 is actually a stable component of intact chromatin because, at least in interphase, over half of it is chromatin-bound, with a mean residence time of the order of minutes (Figures 2.15 and 2.17). The remaining protein is instead not chromatin-bound, consistently with the detection of endogenous HOXC13 both in nucleoplasm and chromatin fractions of biochemically fractionated U2OS cells (Figure 2.3). Notably, such slow nuclear dynamics relies completely on the presence of the homeodomain (Figure 2.15), and in particular on its DNA binding activity (Figure 2.16).

By comparing HOXC13 nuclear dynamics to that of the other numerous nuclear proteins so far investigated by similar biophysical approaches, we can finally classify HOXC13 as a stably chromatin-bound protein, with properties intermediate between the rapidly exchanging transcription factors [188] and the almost immobile structural proteins like histones [190], cohesin [191], CENP at centromeres [192] or PCNA at replication foci [28, 199]. Crucially, HOXC13 deviates from the dynamic nuclear organization typical of all gene transcription regulators [180, 187, 189] (reviewed in **paragraph 1.3.1**). This suggests that its nuclear mobility properties could be exploited in other functions, besides transcription regulation.

3.2 HOXC13 is present at early S replication foci due to its homeodomain

Our data demonstrate that HOX13 is a stable component selectively targeted to the chromatin that is replicated first during S phase progression. This claim is supported by our study of the spatio-temporal correlation of HOXC13 with replication factories (paragraph 2.3), achieved by using two different, complementary approaches. By BrdUrd immunofluorescence of cells expressing E^0GFP-HOXC13 (Figure 2.8), we quantify the degree of co-localization between HOXC13 and RF, demonstrating that it is significantly modulated from high in early S to almost null in late S. In this case, working in fixed cells was crucial, because it allowed to freeze cells at three distinct points of S phase (Figure 2.7); the fine BrdUrd focal structure could be resolved for each of the three moments by deconvolution microscopy and the co-localization analysis was thus more accurate [211, 212]. Alternatively, the live-cell imaging of EGFP-HOXC13 co-expressed with RFP-PCNA (Figure 2.9) confirms that the selective co-localization of HOXC13 with early S replication foci occurs also in an intact chromatin context, and this can be monitored in real time by time-lapse imaging of a nucleus expressing the two constructs (Figure 2.10). Furthermore, our data show that the selective co-localization of HOXC13 with early-S RF relies on the peculiar speckled-like distribution set by its DNA-binding motif (Figure 2.11). Thus, the homeodomain is not only responsible of the stable binding of chromatin by the protein (Figures 2.15-2.17), but also of the targeting of the protein to the chromatin which is earliest replicated during S phase progression.

It is certainly not surprising to find a transcription factor at early replicating chromatin: indeed, early DNA replication has been reported to occur at open, transcriptionally-active chromatin by several groups [14, 34, 243-246]. HOXC13 must be located in these regions if it plays a role in transcriptional activity. Furthermore, early replicating foci could be those in closest proximity with replication origins, where the homeotic protein could be involved in origin function. This observation is consistent with the reported affinity of HOXC13 protein, detected both *in vitro* and *in vivo*, for the DNA sequence of the Lamin B2 origin, that is an early replicating origin [101]. Moreover, our results concerning HOXC13 are in agreement with those reported for the c-Myc transcription factor: this was also found to co-localize with early-S RF, to bind early-S replication origins and to be involved in origin activity in human cells [100].

In this context, it was very interesting to find out that nuclear E^0GFP-Cdc6 displays a peculiar focal distribution resembling the one observed for RF in the three different moments of S phase: early (numerous and ubiquitous foci in the nucleus), mid (perinuclear and perinucleolar foci), late (few bigger central clusters of foci) phases (first three images on the left of Figure 2.22; see also the progressive disappearance of these foci in the course of cell-cycle progression in Figure 2.23B). These Cdc6 foci are present in the nucleus during G1 phase (Figure 2.23), and therefore are not RF, but still they display a spatio-temporal organization comparable to that of RF. These observations support a model (Figure 3.1) in which pre-RC proteins, among them Cdc6, associate with chromatin during the G1 phase probably defining the sites of RC assembly in a precise spatio-temporal fashion that anticipates the replication fork movement during S phase. Thus, pre-RC assembly sites anticipate and determine the patterns of RF during replication timing. A similar model was already proposed to explain the G1 localization patterns of early- and late- replicating chromosomal domains [200] as well as of endogenous MCM proteins [56].

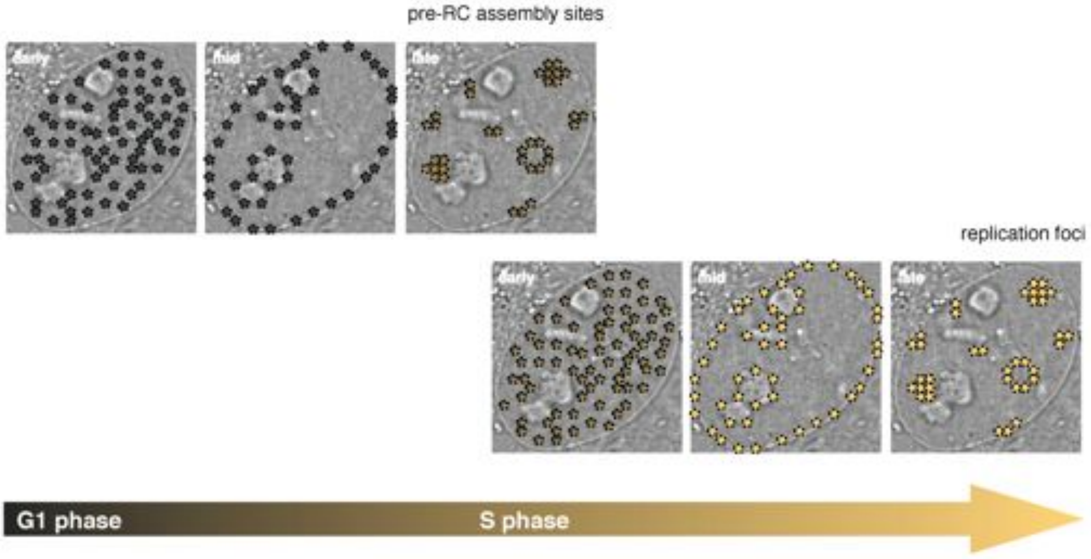


Figure 3.1 Model for the spatio-temporal determination of RF by pre-RC sites assembled during the G1 phase of the cell-cycle. Details of this model are discussed in the text.

In this view, it is worth to underline that, similarly to what already reported for RF (Figures 2.8-2.10), HOXC13 displayed a significant co-localization with Cdc6 nuclear foci when these were numerous and ubiquitous in the nucleus (Figure 2.27). Thus, under the assumptions of this model, it is likely that HOXC13 is present at early replicating chromatin already in G1.

These data, together with those previously discussed, depict a possible role for HOXC13 in origin function, that will be presented in the next paragraph.

3.3 HOXC13 contributes to origin specification in human DNA replication

The mechanism that governs the selection of replication origins in human (and, more generally, metazoan) genomes, in the course of G1 phase of the cell-cycle, is still poorly understood (reviewed in **paragraph 1.1.5**). The lack of DNA-sequence consensus among well-characterized replication origins [84-86], together with the little binding-specificity displayed by ORC [13, 17, 87], suggest that the binding of ORC at sequences selected as origins might rather be determined by local chromatin structures and/or accessory targeting factors. HOXC13 represents an ideal candidate for a role in origin specification, because it combines an affinity for ORC and RC proteins, with structural chromatin-binding properties relevant in the context of DNA replication initiation.

HOXC13 displays affinity for the RC, as demonstrated by the FLIM study of the interaction of this protein with Cdc6 and ORC2 detected in living cells (paragraph 2.6). These interactions correlate with origin function: indeed, HOXC13 interacts with Cdc6 during G1 phase (Figures 2.27 and 2.30), and with ORC2 at the end of G1 phase, before the start of S (Figures 2.26, 2.30 and 2.32). Moreover, these interactions occur in the chromatin compartments defined by the homeodomain (Figure 2.31), i.e. in the nuclear areas where early replication origins are collected in G1 phase (Figure 3.1). With these considerations, it is tempting to suppose that HOXC13 contributes to the specification of early origins by recruitment and stabilization of the RC onto replication origin sequences. This hypothesis stems not only from the above-cited direct interactions of HOXC13 with the RC. Indeed, it has to be remembered that HOXC3 is a stable component of early replicating chromatin (see the two previous paragraphs). Both its nuclear localization and dynamics are almost constant throughout cell-cycle progression. Thus, HOXC13 does not undergo the specific chromatin loading at the beginning of S phase, which was reported to be crucial for MCM [21] or PCNA [28] proteins to exert their role during elongation. This, again, points to a role of the homeotic protein at the stage of replication initiation, rather than elongation. The stable chromatin binding of HOXC13 could be usefully exploited by the RC (via direct interaction with ORC2 and Cdc6), which instead undergoes a regulatory dynamic exchange with chromatin and often subcellular re-localization of its components during cell-cycle [22, 23, 49, 64, 224, 225, 231, 232]. Indeed, an accepted model for the RC proposes that it is a stable complex consisting of dynamic subunits [203] (reported in Figure 1.18). The dynamic model of the RC subunits is also supported by our observation that four selected RC proteins, namely Cdc6, ORC2, ORC1, MCM3, although supposed to be part of the same complex, displayed four different localizations when their fluorescent fusion constructs were expressed in U2OS cells (**paragraph 2.5**).

Nevertheless, the hypothesis of HOXC13 as a key protein for origin specification suffers of the relatively low sequence-specificity displayed by the HOX family, when considered for its role in transcription [247] (reviewed in **paragraph 1.2.3**). HOXC13, like all homeotic proteins, has a preference for the binding of AT-rich DNA sequences [136]: the presence of asymmetric A/T stretches is up to date the only relatively conserved feature among origin sequences studied by genome-wide approaches [15]. Our FRAP data indicate that most of the HOXC13 stabilization onto chromatin relies on the anchoring of the homeodomain to DNA (Figures 2.15-2.16). Notably, not all mutations that did show to completely abolish the DNA-binding of the homeodomain *in vitro* [139] had the same effect in our *in vivo* analysis. In particular, we have indication that most of the stabilization onto chromatin depends on the interaction of the Arg-5 residue of the homeodomain with the DNA minor groove. Interestingly, N-terminal arginine residues of the homeobox were recently indicated as the only residues necessary to confer specificity in the DNA binding by HOX and other homeotic proteins (*paragraph 1.2.3.2*) [146]. Such recognition property may provide the relatively loose sequence specificity contribution of HOXC13 that could lead to precise origin specification, in combination with other partial but converging specificities. Indeed, minor groove anchoring by arginine residues seems to constitute the basis for a new, general DNA-recognition mechanism used by many families of DNA-binding proteins [152]. This suggests that other proteins, with DNA-binding properties similar to HOXC13, could play a role as structural "chromatin-marker" of early replication origins. In this context, it is worth mentioning that also other homeotic proteins (HOXA13, HOXC10, HOXD11 and HOXD13) were reported to have affinity for the lamin B2 origin [101, 103], suggesting that the interaction of HOXC13 with the replication origins might occur in the context of a multi-protein homeotic effector. This is indeed the rule when HOX proteins exert their function in transcription, by forming a multi-protein complex named "Hoxasome", including HOX proteins, homeobox cofactors and possibly other transcription factors [248, 249] (Figure 1.13). In this perspective, we should expect that, besides HOXC13, other homeotic proteins could display a stable interaction with chromatin based on the conserved Arg-5 at DNA minor groove. Moreover, it should be stressed that among proteins taking part to the pre-RC, HOX proteins share some similarity with *S. pombe* ORC4 and with HMG1 protein; both proteins were shown to bind to the minor groove of AT-rich stretches of the origin sequence and thus contribute to origin specification [99, 250].

With all these considerations, we might propose a "network/combinatorial" approach as a possible guide for solving the puzzle of metazoan origin selection: according to this model, constitutively open-chromatin sequences may be bound preferentially by topoisomerases [92], assuring a local negatively-supercoiled status, or by other proteins recognizing yet to be defined features of the sequence (asymmetric AT stretches, bent DNA, narrow minor grooves?), and by the homeoprotein complex that, thanks to the interaction with the RC, facilitates the assembly of this complex onto the origin. In this view, the property of replication origin might emerge from a limited number of combinations of the structural features mentioned above, in an appropriate cellular context, in which certain protein categories are available.

Finally, it appears certainly stimulating to discover a connection between an element of the DNA replication regulation and an actor of development and differentiation [104]: actually, the HOXC13 protein may be only the first identified of possibly several go-betweens of the two processes. It would be interesting to elucidate whether the involvement in origin specification is a specific feature of Abd-B Hox paralogs (Figure 1.10), since, so far, all HOX proteins found to recognize origin sequences belong to this class [104]. If this were the case, the specific structural elements of Abd-B HOX proteins conferring this property, and their interaction modes with the elements of the replicative complexes should be determined. This would allow the description of the features of this HOX family in terms of protein-DNA and protein-protein interaction that are required for origin function, also in relation with their function in differentiation processes. Moreover, more detailed functional studies should be performed to investigate the effect of HOXC13 (and other related interacting homeoproteins) depletion on origin function. This would finally lead to uncover the precise moment, in the stepwise build-up of the pre-RC and pre-IC complexes, in which HOX proteins are involved.

4. MATERIALS AND METHODS

4.1 Expression constructs

All constructs used in this study were prepared using standard cloning procedures. Primers for PCR amplification were from Sigma; restriction enzymes and ligases were from New England Biolabs; DNA polymerases from Invitrogen and Stratagene; dNTPs were from Eppendorf. The complete sequence of all obtained constructs was always verified by nucleotide sequencing prior to use.

4.1.1 HOXC13 protein constructs

To generate the N-terminal fluorolabelled HOXC13, the sequence of full length HOXC13 (1–330) was BamHI/XbaI amplified and cloned downstream a HindIII/BamHI amplified mCherry, mOrange (both are kind gifts of R.Y. Tsien), EGFP or E^0GFP [238] sequence in a pcDNA3.1(+) vector (Invitrogen), thus obtainining the mCherry-HOXC13, mOrange-HOXC13, EGFP-HOXC13, and E^0GFP-HOXC13 fusion constructs, respectively. For the preparation of C-terminal fluorolabelled HOXC13, a HindIII/BamHI amplified HOXC13 was cloned upstream BamHI/XbaI amplifications of the same fluorophores above mentioned in a pcDNA3.1(+) vector. In this case, a 20-aminoacids linker (RGSASGGGGGLVPRGSASGA) was interposed between protein and fluorophore, to avoid hampering of the homeodomain by the FP structure.

HOXC13 N-terminal portion (1–257), devoid of the homeodomain, was subcloned into the BamHI/EcoRI sites of a pcDNA3.1(+) vector in which a KpnI/BamHI NLS_{SV40}-mCherry fusion was previously inserted, to get the mCherry-Deletion mutant.

The mCherry-HOXC13 construct was mutated (Stratagene Site-directed Mutagenesis Kit) in three residues of the third homeobox helix (I47A, Q50A, N51A, where numbers correspond to the homeodomain numeration) to generate the “HBX-helix mutant”. Site-directed mutagenesis was also performed to convert Arg5 of the homeodomain N-terminal loop into alanine to generate the “HBX-loop mutant”. Both mutations of the third helix and of the N-terminal loop were inserted in the “HBX-combined mutant”.

4.1.2 Control constructs

NLS_{SV40}-mCherry sequence was cloned in the Kpn/BamHI sites of a pcDNA3.1(+) vector and used as control in FRAP experiments. EGFP-NLS_{SV40} similarly cloned in pcDNA3.1(+) vector (a kind gift of Dr. Michela Serresi) was subjected to EGFP-point mutation to generate E^0GFP fluorophore and used as control in FRET experiments. mCherry was also cloned alone in the HindIII/BamHI sites of a pcDNA3.1/Hygro vector (Invitrogen) and used as control in FRET experiments.

4.1.3 RC proteins constructs

RFP-PCNA [199] was a kind gift of M.C. Cardoso and H. Leonhardt. All Cdc6, ORC2, ORC1 and MCM3 sequences were cloned into pEGFP-C1 or pEGFP-N1 vectors (Clontech). The cDNA of ORC2 and Cdc6 proteins (KpnI/SmaI and KpnI/BamHI amplified, respectively) were cloned into pEGFP-C1 vector. EGFP-MCM3 and ORC1-EGFP constructs were a kind gift of Dr. Roberta Paolinelli and Dr. Ramiro Mendoza-Maldonado. All these four EGFP constructs were subjected to EGFP-point mutation to generate E^0GFP fluorophore.

4.1.4 Other nuclear proteins constructs

GFP-HP1α was a kind gift of Dr. Roberta Paolinelli and Dr. Ramiro Mendoza-Maldonado.

4.2 Cell culture, transient and stable transfection, synchronization

U2OS, MCF7, HeLa cells (ATCC: HTB96, HTB22, CCL2, respectively) were grown in DMEM, supplemented with 2 mM glutamine, 1 mM sodium pyruvate, 10 U/L penicillin, 10 μg/L streptomycin and 10% fetal bovine serum (Gibco). NIH3T3 cells (ATCC: CRL1658) were grown in the same conditions, but using 10% calf serum (ATCC). Sub-confluent U2OS, MCF7, HeLa cells were transfected using Effectene reagent (Qiagen), while confluent NIH3T3 cells with Lipofectamine 2000 (Invitrogen), following the manufacturer's protocol. 24-48h after transfection cells plated onto 35-mm glass-bottom dishes (WillCo-dish GWSt-3522) were usually imaged at 37°C, 5% CO_2, otherwise they were processed for fixed-cell microscopy or for biochemical analysis. U2OS cell lines stably expressing the desired recombinant fluorescent construct were generated by selecting transfected cells with 700 μg/ml Neomycin

(Gibco); typically, after one week selection, they were pooled, expanded and usually kept in culture for 2-3 weeks in the constant presence of antibiotic.

For synchronization, U2OS cells were synchronized at the G1/S border by 24 hours incubation with 5 μg/ml aphidicolin (Sigma). HeLa cells were synchronized at the G1/S border by subjecting cells to a double 16 h block with 2.5 mM thymidine (Sigma), spaced out by 9 h release in fresh DMEM. To evaluate the synchronization, an aliquot ($\geq 5 \cdot 10^4$) of treated cells was washed twice in ice-cold PBS, fixed in frozen absolute ethanol, and further stained with Propidium Iodide (Sigma), in PBS buffer supplemented with RNase (Sigma) and 0.1% Nonidet-P40, before FACS analysis of the DNA content.

4.3 Antibodies

The following primary antibodies were used for either immunofluorescence (IF) or Western Blot (WB) experiments: rabbit polyclonal α–HOXC13 (IF: 1:50 for detection of endogenous protein, 1:100 for detection of recombinant protein; WB: 1:500); mouse monoclonal α–tubulin, clone B-512 Sigma (WB: 1:5000); mouse monoclonal α–actin, clone AC-40 Sigma (WB: 1:2000); goat polyclonal α–ORC2, clone B-18 Santa Cruz Biotechnology (WB: 1:200); mouse monoclonal and rabbit polyclonal α–GFP, ab1218 and ab290 Abcam (WB: 1:1000); mouse monoclonal α–dsRed, 6323392 Clontech (WB: 1:300/500); rabbit polyclonal α–dsRed, 632496 Clontech (IF: 1:100); mouse monoclonal α–HP1β, MAB3448 Chemicon (WB: 1:1000); mouse monoclonal α–PG-M3 antibody, sc-966 Santa Cruz Biotechnolgy (IF: 1:100); mouse monoclonal α–BrdUrd antibody, sc-32323 Santa Cruz Biotechnology (IF: 1:100). Secondary antibodies, for WB (1:2500) and IF (1:100) applications, were from Biorad and Jackson Immunoresearch, respectively.

4.4 Immunofluorescence

Cells were plated at sub-confluence on coverslips and, if needed, transfected with the desired construct; 48-72 h later, they were rinsed with ice-cold PBS, fixed with 4% paraformaldehyde for 10 minutes, permeabilized in 0.25% Triton X-100 in 0.5-1% BSA/PBS for 10 minutes. When detergent pre-extraction was required to detect only chromatin-bound proteins, cells were incubated with 0.5% Triton X-100 in PBS (plus protease inhibitors) on ice prior to fixation. In this case no further permeabilization was needed. Cells were blocked in 1-2% BSA/PBS before incubation with the primary antibody diluted in 0.5-1% BSA/PBS. 1-3 h incubations at room temperature were typically used. Coverslips were washed 3 times with PBS before incubation with the secondary antibody, diluted in 0.5% BSA/PBS, for 1h at room temperature.

Secondary antibodies were usually conjugated to Alexa647. Slides were washed 3 times with PBS and once with H_2O and finally mounted with Vectashield/DAPI mounting medium (Vector Labs). For BrdUrd immunofluorescence, cells were pulse-labelled with 100mM BrdUrd (Sigma) in DMEM for 15 minutes, washed with PBS, fixed and permeabilized as above described. The anti-BrdUrd antibody was 1:100 diluted in 33 mM Tris-HCl pH 8, 3 mM $MgCl_2$, 0.5 mM β-mercaptoethanol, 0.5% BSA, 40 U/ml DNase (Roche, 10 U/μl), similarly to what already reported [197]. For PML protein immunofluorescence, permeabilized cells were incubated with α–PGM3 antibody 1:100 diluted in 0.5% BSA/PBS supplemented with 0.1%Tween-20, as already reported [251].

4.5 Biochemical cell extraction, fractionation and co-immunoprecipitation

Cell extracts were required to check endogenous HOXC13 and its fluorolabelled constructs for correct expression and possible degradation. In this case, total cell extracts were usually obtained scraping 3-5×10^6 cells in RIPA buffer (Sigma), supplemented with protease inhibitors tablets (Roche).

Cell fractionation was required to analyze the subcellular localization and chromatin affinity of endogenous HOXC13. To this purpose, typically 3-5×10^7 cells were used for each fractionation: these were harvested by trypsinization, counted and washed twice with ice-cold PBS. 5×10^6 cells were lysed in RIPA buffer and further referred to as whole cell extract. The remaining cells were sequentially fractionated. First cytoplasm was extracted, similarly to what already reported [21, 207], resuspending cells (at 4×10^7cells/ml) in ice-cold 15 mM Tris-HCl pH 7.5, 10 mM KCl, 10 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 300 mM sucrose, 10% glycerol and 0.1% Triton-X100 for 7’. Low-speed centrifugation (1300g, 5’) allowed separating cytoplasm (supernatant) from intact nuclei. These were washed three times before performing the nucleoplasm extraction (at 1.2×10^8cells/ml) in ice cold 25 mM Tris-HCl pH 8, 20 mM NaCl, 5 mM $MgCl_2$, 1 mM EDTA, 10% glycerol and 0.5% NP-40 for 20’. Low-speed centrifugation (1500g, 5’) allowed separating nucleoplasm (supernatant) from chromatin. Chromatin was washed once and then salt-extracted in ice-cold 25 mM Tris-HCl pH 8, 0.5 mM $MgCl_2$, 300 mM sucrose, 10% glycerol and sequentially increasing NaCl concentrations (150, 300, 600 mM, and 2M NaCl). The 150 mM extraction step was either performed without or with 200U DNase I (Roche) and 6 mM $MgCl_2$ (in this case, extraction was performed at 22°C). Each extraction step lasted 30' under rotation and was followed by centrifugation at 14000g for 5-15' before subsequent pellet extraction with increasing NaCl. All obtained fractions were clarified (at 16000g for 30'). The described buffers were supplemented with protease inhibitors (leupeptin, aprotinin, PMSF), phosphatase inhibitors (NaF, Sodium orthovanadate), 1 mM DTT and 1 mM ATP prior to use. All centrifugation steps were performed, if not differently indicated, at 4°C.

The co-immunoprecipitation assays were performed using transfected U2OS cell nuclei, extracted using a sequential combination of DNase I + 600 mM NaCl extractions. 250 μg of nuclear extract (1:10 diluted in PBS) were incubated for 2 h at 4°C with Dynabeads-protein A (Invitrogen), previously functionalized either with rabbit α–GFP, or with control rabbit IgG. Beads were washed twice with a phosphate buffer supplemented with 300 mM NaCl before investigation of the immunoprecipitated proteins by Western Blot.

Lysate protein concentration was always estimated by Bradford assay (Pierce) and usually ~10-30 μg were analyzed by SDS-PAGE and transferred onto nitrocellulose filters, before subsequent protein detection by Western Blot.

4.6 Co-localization analysis

4.6.1 Image acquisition

All co-localization analyses were performed between sequentially acquired, separated fluorescence channels. Laser line excitation and acquisition ranges were optimized to the following values:

- DAPI/Alexa647 co-localization:

  blue channel (DAPI): 403nm laser line excitation with [405-480nm] acquisition range;

  far-red channel (Alexa 647): 633nm laser line excitation with [650-750nm] acquisition range;

- EGFP(or E^0GFP)/mCherry co-localization:

  green channel (EGFP/E^0GFP): 488nm laser line excitation with [495-530nm] acquisition range;

  red channel (mCherry): 543 or 561nm laser line excitation with [585-650nm] acquisition range;

- EGFP(or E^0GFP)/Alexa647 co-localization:

  green channel (EGFP/E^0GFP): 476nm laser line excitation with [485-550nm] acquisition range;

  far-red channel (Alexa 647): 633nm laser line excitation with [650-750nm] acquisition range.

For the co-localization of E^0GFP-HOXC13 with pulsed BrdUrd foci labeled with Alexa647, of DAPI with endogenous HOXC13 labeled with Alexa647, images were acquired with a Leica TCS SP2 inverted confocal microscope (Leica Microsystems) using a 40X (NA 1.25) oil immersion objective and pinhole set to 1 AU. DAPI was detected at the Leica microscope by a 403 nm pulsed diode laser (M8903-01; Hamamatsu), tuned at 50 MHz repetition rate. To evaluate the co-localization in the whole nuclear volume, an optimized number of z-sections of cell nuclei were acquired in the separated fluorescence channels. For the co-localization of EGFP-HOXC13 with RFP-PCNA, images were acquired with an Olympus FluoView 1000-ASW-2.0 confocal microscope (Olympus), using a 60X (NA 1.35) oil immersion objective and pinhole set to 1 AU. In both cases, typically, frame size was 512×512 pixels and frame physical length ranged from 25 to 50 μm, depending on the nucleus size.

4.6.2 Image data analysis

Usually, co-localization was estimated by the Merge image of the two fluorescence channels. In the case of the co-localization of HOXC13 with BrdUrd (**paragraph 2.3.1**), a quantitative co-localization analysis was performed. The z-sections of cell nuclei were first deconvolved using the Huygens Essential Deconvolution Software, as previously described [212]. Deconvolved 3D Z-stacks were finally analyzed with the “colocalization analyzer tool” of this software to get the Pearson’s coefficient (P). Mean P values (±SD) were obtained out of all analyzed cells and used to compare the co-localization obtained from cells at different moments of S phase (see Figure 2.7). Statistical significance was determined by Mann-Whitney test, with $\alpha < 0.01$ considered as significant. Under these conditions, P values obtained for early-, mid- and late-S U2OS cells were found to be statistically different one from each other (see Figure 2.8). For the same experiment performed with HeLa cells, the cell sampling is not sufficient to determine statistical significance. The analysis of HeLa cells is only intended to prove that the modulation of co-localization between fluorolabelled HOXC13 and RF does not specifically occur in U2OS cells only.

4.7 FRAP

4.7.1 Image acquisition

FRAP experiments were performed with an Olympus FluoView 1000-ASW-2.0 confocal laser scanning microscope, equipped with an incubator chamber set to 37°C and 5% CO_2. All images were acquired using a 60X/1.42NA oil immersion objective and a frame size of 256×256 pixels (0.138 μm constant pixel size). mCherry-tagged constructs were imaged with the 543nm laser line (acquisition range: 555-655 nm), at variable laser power (optimal: 10 -20 μW), constant 760 V PMT voltage and pinhole set to 200 μm (≈ 2AU). mCherry photobleaching was achieved using a single pulse with all 543, 514, 488, 458nm laser lines set to full power for the necessary time to bleach a ROI covering half of the nucleus (typically: 250-600 msec, depending on the nucleus size). To verify the suitability of mCherry fluorophore for FRAP experiments, a GFP-tagged version of HOXC13 was used in control FRAP measurements, together with the reference GFP-HP1α construct. These were imaged with the 488nm laser line (acquisition range: 495-595nm), at variable laser power, constant 680V PMT voltage and pinhole set to 200 μm. GFP photobleaching was achieved with the same procedure above described for mCherry, but using the 405, 458, 488nm laser lines set to full power for the GFP bleaching. Time series were recorded for 5 minutes at 0.129 ms/frame sampling rate with 20 pre-bleach frames for slow proteins; for 30 seconds at 0.065ms/frame sampling rate with 5 pre-bleach frames for fast constructs (NLS_{SV40}-mCherry and HOXC13 deletion mutant). Longer time series were also acquired for the wt mCherry-HOXC13 construct. In this case, images were acquired for 30 minutes at 2s/frame sampling rate with 5 pre-bleach frames. These experiments were only intended to prove that the immobile fraction of ~10% reported for the construct within 300 s observation (Figure 2.15) is actually a very slow component of the recovery; therefore they are not reported in this thesis in the presentation of the results of the FRAP analysis.

4.7.2 Image data analysis

For each analyzed nucleus, the average fluorescence intensities of the bleached area for each time point were background substracted, normalized to the pre-bleach average value and also for total nuclear fluorescence (in order to correct for photobleaching during acquisition). Data were finally normalized for the bleach depth [221]. For each analyzed construct, the FRAP curves of 10-40 cells were averaged and the mean curve

(±SE) was used to compare the mobility of all analyzed constructs. To get quantitative information about the chromatin binding properties and the dynamics of the wt mCherry-HOXC13, a simplified diffusion-reaction mathematical model was applied to the analysis of FRAP data (see Appendix A of this thesis for detailed information). The fits where performed using Wolfram Mathematica 6.0.1.0, and as explained in Appendix A the quality of the fit was checked in selected cases, by looking if there was a clear minimum for the χ^2 (see Figure A.2 in Appendix A).

4.8 FLIM

4.8.1 Image acquisition

Measurements were performed with a Leica TCS SP2 inverted confocal microscope (Leica Microsystems), equipped with an incubator chamber set to 37°C and 5% CO_2. The microscope was interfaced with fast photon counting external detectors (Hamamatsu, H7422P-40) and time-correlated single-photon counting (TCSPC) electronics (Becker & Hickl). All images were acquired using a 40X (NA 1.25) oil immersion objective. First, reference intensity images were obtained for E^0GFP constructs and, when co-expressed, mCherry constructs, at 512×512 frame size using 488 and 561nm laser lines, respectively. Then the donor image was acquired at the frame size of 128×128 pixels using the photon counting mode: in this case, 403nm-excitation of E^0GFP was achieved using a pulsed diode laser (M8903-01; Hamamatsu) set at 10MHz repetition rate and 3-5 μW laser power. These conditions ensured neither photobleaching nor photoactivation of the donor fluorophore, as well as photon counting rates between 10^4–10^5 cps. Time of acquisition ranged from 80 to 200s (typically 120s), depending on donor expression level.

4.8.2 Image data analysis

The images of the donor, following pulsed 403nm excitation, were used to obtain lifetime values from fluorescence decays using a pixel-by-pixel fitting procedure. Usually fluorescence decays were optimally fitted after binning of 1-3. Only pixels within cell nuclei were considered: lifetimes were repeatedly fitted until all nuclear pixels displayed a $\chi^2 \leq 1.3$. First, decays of cells expressing donor alone were fitted with a monoexponential decay equation, to obtain the mean lifetime value of the donor alone, τ_D. When the donor was expressed with the acceptor, it was assumed to exist either in the unbound or in the acceptor-bound state (see also **paragraph B.2** in Appendix B); therefore data were fitted with the following biexponential equation:

$$I(t) = a_1 \cdot e^{-\frac{t}{\tau_{DA}}} + a_2 \cdot e^{-\frac{t}{\tau_D}}, \qquad \text{(Eq. 4.1)}$$

where τ_D was known from the previous monoexponential fit and therefore was fixed. τ_{DA}, the shorter lifetime of the donor involved in FRET with the acceptor, was the fitting parameter together with a_1 and a_2. The resulting τ was an average of τ_D and τ_{DA} components weighted for the respective subpopulations. In the main text, both τ derived from mono- and bi-exponential fittings are referred to as τ_m (mean lifetime). Lifetime distribution histograms were obtained from all analyzed nuclei and were normalized to the nucleus area (i.e. pixels number). The sum of all distribution histograms of each analyzed cell was used to calculate the weighted mean lifetime, τ_m, reported in Table 2.2 (**paragraph 2.6.1**). This sum histogram was also fitted with a standard Gaussian curve for presentation purposes (see graphs of Figure 2.30). The peak value of the Gaussian curves is representative of the τ_m reported in Table 2.2; moreover these graphs give an idea of the distribution of lifetime values around τ_m.

The fitting analysis was performed with SPC-Image software (Becker & Hickl). Data and images were further analyzed by Origin Pro 7.0 and ImageJ softwares.

4.9 Time lapse imaging

4.9.1 Image acquisition

The time-lapse imaging of E^0GFP-ORC2 or E^0GFP-Cdc6 expressed in U2OS cells was performed with a Leica TCS SP2 confocal microscope, using the 488nm laser line excitation, a 40X/1.25NA oil immersion objective, pinhole set to 3AU and 1024×1024 pixels frame size. Four to five z-sections encompassing all nucleus thickness were imaged every 30 minutes for maximum 16-20 hours. The time-lapse videos of EGFP-HOXC13 and RFP-PCNA co-expressed in NIH3T3 cells were acquired with an Olympus FluoView 1000-ASW-2.0 confocal microscope, with 488 and 543nm laser lines sequential excitation, 60X/1.35NA oil immersion objective, pinhole set to 1AU and 512×512 pixels frame size. The middle section of the nucleus was manually searched and imaged every 30 minutes for maximum 6-8 hours (S phase). All images were acquired at constant 37°C and 5% CO_2, at the lowest possible laser power to avoid both photobleaching and phototoxicity artifacts.

4.9.2 Time-lapse imaging presentation

For the time-lapse imaging of E^0GFP-ORC2 and E^0GFP-Cdc6, the maximum Z-projection of each time point was used to build up the final movie. For the time-lapse imaging of EGFP- HOXC13 and RFP-PCNA, either the single fluorescence channels, or the Merge of the two channels, were used to build up the final movie.

4.10 Image presentation

Images were analyzed and prepared for presentation using ImageJ, Adobe Photoshop CS3, Adobe Illustrator CS3 and CS4 softwares. All acquired images were always background subtracted before analysis. All presented images were subjected to linear contrast enhancement after image acquisition. Magnification is reported with a white scale bar in the images.

APPENDIX A

Models for the FRAP analysis of chromatin-binding by nuclear proteins

This appendix reviews the mathematical models that can be used for the FRAP analysis of nuclear proteins dynamics, in order to obtain quantitative parameters about chromatin-binding kinetics. The outline here reported is by no means complete and is only intended to provide a quick reference. More extensive and detailed information about these models can be found elsewhere [182, 185, 186, 188, 214, 217, 222, 252].

A.1 Available models to analyze the nuclear dynamics of chromatin-binding proteins

Three models are most commonly reported for the analysis of nuclear dynamics:

1. the diffusion-dominant model, assuming that proteins are freely diffusing in the nucleus;
2. the reaction-dominant model, assuming that diffusion is very fast compared to binding on the timescale of the FRAP measurement;
3. the diffusion-reaction model, which considers contributions from both binding and diffusion on similar timescales.

The first model is usually adopted for the study of nuclear proteins exchanging very rapidly within chromatin, as is the case of HP1 proteins at euchromatic nuclear regions [217]. The second one has been extensively reported in many biological studies (for example, of nuclear proteins), where it was assumed that diffusion could be neglected to simplify the analysis because the redistribution of fluorescence between bleached and unbleached areas is slow compared to the case of freely diffusing molecules [188, 253-255]. Unfortunately, this assumption is not always correct, because very transient interactions in cases when diffusion is clearly limiting, can also lead to slow fluorescence redistributions [186, 214]. For this reason, most of FRAP studies more recently reported [182, 217, 252] make use of the third method, the diffusion-reaction model, to analyze nuclear dynamics of chromatin-binding proteins. Accordingly, this was the method of our choice in the study of HOXC13 nuclear dynamics.

A.2 Equations of the diffusion-reaction model

The diffusion-reaction model assumes that proteins are immobile when bound to chromatin, because chromatin does not show large-scale movements over 1 h in mammalian cells [256-258], while free molecules can normally diffuse within the whole nucleus with a single diffusion coefficient D following Fick's second law:

$$\frac{\partial[\mathrm{free}(\vec{\mathrm{r}},\mathrm{t})]}{\partial t} = D\Delta([\mathrm{free}(\vec{\mathrm{r}},\mathrm{t})]), \qquad \text{(Eq. A1)}$$

where $[\mathrm{free}(\vec{\mathrm{r}},\mathrm{t})]$ is the local concentration of unbound fluorescent molecules.

The interaction between protein and chromatin can be written as:

$$F + S \underset{k_{on}}{\overset{k_{off}}{\rightleftarrows}} C, \qquad \text{(Eq. A2)}$$

where F represents free molecules, S vacant binding sites, C bound [FS] complexes, and k_{on} and k_{off} are the on- and off-rates, respectively [186, 214].

In order to apply the model to the analysis of half-nuclear FRAP experiments, we assume that the biological system has reached equilibrium before photobleaching, and that the number of free binding sites [S] does not fluctuate appreciably during the FRAP experiment; under these assumptions, applicable in many biological situations, one can consider a pseudo-first-order rate constant given by $k_{on}^{*} = k_{on}[S]_{eq}$, where $[\mathrm{S}]_{\mathrm{eq}}$ is the concentration of the binding sites at equilibrium; we consider $[\mathrm{S}]_{\mathrm{eq}}$ homogeneous in the nucleus as a further simplification.

Combining diffusion and interaction kinetics, changes in the local concentrations of fluorescent proteins during fluorescence redistribution can then be described by the differential equations:

$$\frac{\partial f}{\partial t} = D\nabla^2 f - k_{on}^{*} f + k_{off} c$$
$$\frac{\partial c}{\partial t} = k_{on}^{*} f - k_{off} c,$$

(Eq. A3)

where f and c represent free and bound molecule concentrations, respectively; D is the diffusion coefficient for f; k_{off} is the off-rate; k_{on}^* is the association rate at equilibrium. Boundary conditions require no flux of f cross the boundary; at equilibrium, $c=K\,f$, where $K = \frac{k_{on}^*}{k_{off}}$.

We substituted in Eq. A3 $k^*_{on}=Kk_{off}$ and calculated the solution of the Fourier-transform in $\vec{r}$ of the resulting system of differential equations with the initial condition $\tilde{c}(q,t=0)=K\tilde{f}(q,t=0)\equiv 1$ ($\tilde{c}$ and $\tilde{f}$ are the spatial Fuorier transform of c and f, respectively; $\vec{q}$ is the transform variable of $\vec{r}$ and the solutions depend only on $q=|\vec{q}|$). The final result in real space ($\vec{r}$) clearly depends on the model geometry chosen to approximate the actual geometry and on the initial conditions of the experiment, as is explained in the following two paragraphs.

A.2.1 Half-nuclear geometry

At first, the resulting system of differential equations (Eq. A3) was solved for a spatial geometry resembling the whole half-bleached nucleus. We calculated the Fourier series for the 1D solution of Eq. A3, i.e. the solution for a nucleus approximated as a rectangular parallelepiped, with initial conditions that the border between the bleached and unbleached parts is a symmetry plane of the system. The fluorescence as a function of time and space is given by $c(\vec{r},t)+f(\vec{r},t)$; an example of the solution for this function is reported in Figure A.1. Accordingly, the FRAP recovery curves were fitted to the function:

$$h(t;D,K,k_{off}) = 1 + \frac{4}{l\left(1+K^{-1}\right)}\sum_{n=1}^{N}\frac{-2l}{(2n+1)^2\pi^2}\left[\tilde{f}\left(\frac{(2n+1)\pi}{l},t\right)+\tilde{c}\left(\frac{(2n+1)\pi}{l},t\right)\right], \quad \text{(Eq. A4)}$$

where D, K and k_{off} are fitting parameters (their presence in every term of the sum on the right is not explicit), N is the number of non-zero terms considered in the Fourier series, and l is the total length of the rectangular parallelepiped best approximating the nucleus. $h(t)$ is the integral of $c(\vec{r},t)+f(\vec{r},t)$ in the bleached part, and represents the change of the average fluorescence in the bleached part normalized between 0 (at the starting time) and 1 (after total recovery).

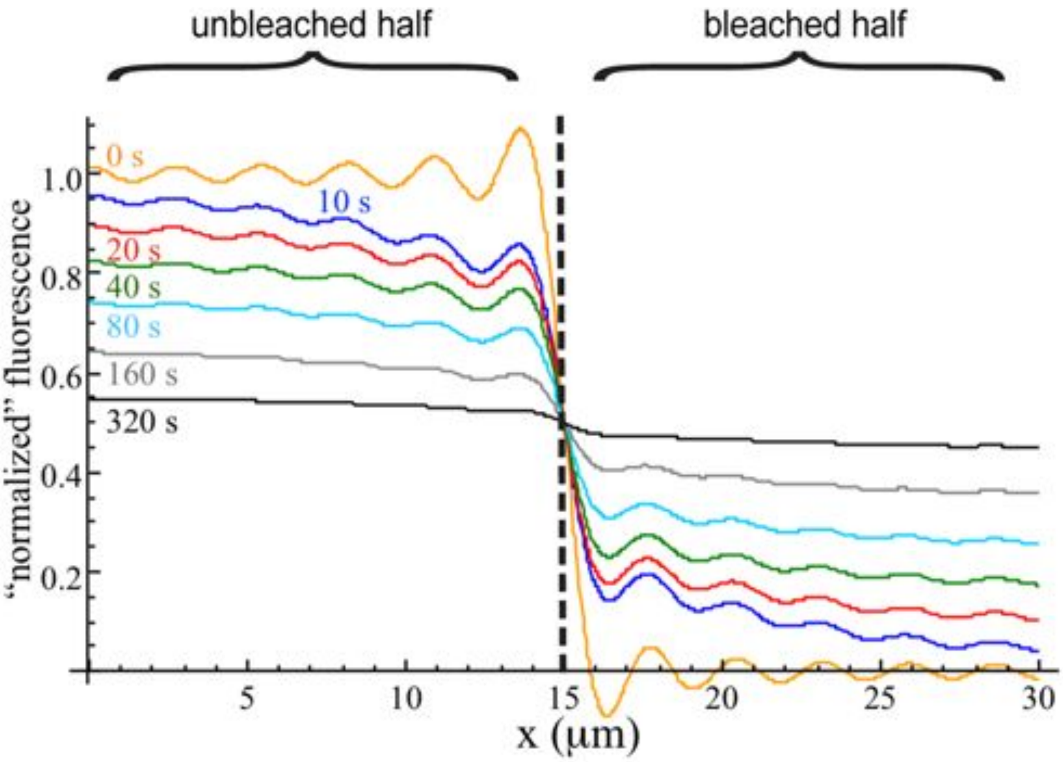


Figure A.1 Fluorescence re-distribution in a half-bleached nucleus. The different curves correspond to the fluorescence distribution in the nucleus, at the indicated different times (from 0 to 300s), according to the solution of Eq. A4 approximated with N=10 non-zero and non-trivial terms in the Fourier series for l=30μm, D=5.03μm^2/s, K=1.56, k_{off}=0.009s^{-1}.

Since the bleach depth normalization in the experimental FRAP data forces them to start at zero, in order to correct for truncation or rounding errors we actually fitted the data with the function:

$$\frac{h(t;D,K,k_{off}) - h(0)}{1 - h(0)}, \qquad \text{(Eq. A5)}$$

where $h(t$=0) is usually <<1 in absolute value, does not depend on the values of D, K and k_{off}, but depends on N. We verified that the results of the fit didn't change significantly for N between 6 and 15; for the results presented in this work, N was usually 10 (actually corresponding to a Fourier series with 22 terms, including the first one for q=0). The fits where performed using Wolfram Mathematica 6.0.1.0, and we checked in selected cases that there was a clear minimum for the χ^2 (proportional to the sum of the square of the residuals from fit); this was clearly the case considering the plane of parameters (D, k_{off}) and (K, k_{off}), whereas this analysis revealed a strong correlation in the plane (D, K), with the consequence that a bigger uncertainty is expected for D and K parameters (Figure A.2).

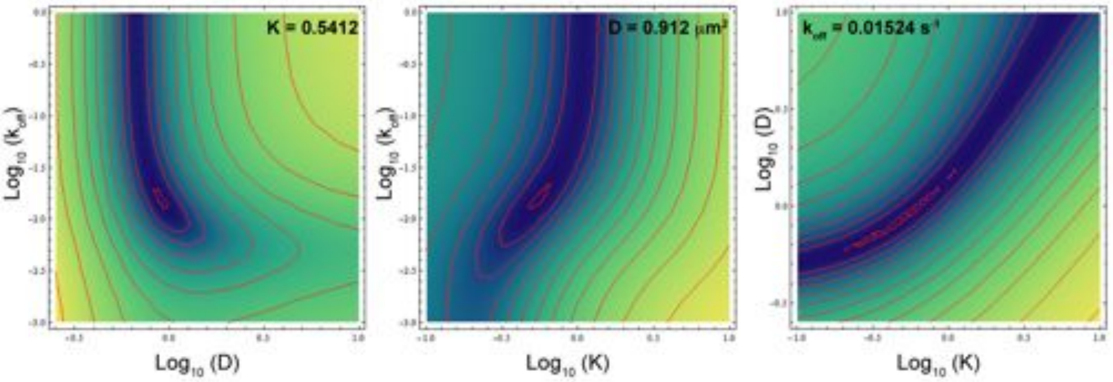


Figure A.2 Maps of Log_{10} (χ^2) obtained in the fitting procedure. The three graphs correspond to the variation of Log_{10} (χ^2) with respect to the planes of parameters (D, k_{off}), (K, k_{off}) and (K, D), respectively, with the third fitting parameter corresponding to the one of the best fit. The fit refers to a representative analyzed cell. Blue-shifted colors correspond to lower Log_{10} (χ^2) values, while yellow-shifted colors to high Log_{10} (χ^2) values; level lines every 0.2 change of Log_{10} (χ^2).

Fitting raw FRAP curves to this function allowed to obtain, for each analyzed cell, three parameters, namely D, *K*, and k_{off}, once *l* was fixed cell-by-cell. Actually, in order to give results directly comparable to those reported in literature [188, 214], the three parameters are presented here and in Chapter 2 (**paragraph 2.4.4**) as:

- $D = D_{eff}$. This is to underline that this parameter probably represents the process of effective diffusion, whereby fast and unspecific binding interactions combine with free diffusion to mimic a slowed diffusion [186, 214, 222].
- $K=k_{on}^{*}/k_{off}$, the equilibrium constant of binding of chromatin-bound protein. This parameter was further converted into F_{free} (i.e. the percentage of free protein), using the following equation:
$$F_{free} = \frac{[F]_{eq}}{[F]_{eq}+[C]_{eq}} = \frac{1}{1+K}; \qquad \text{(Eq.A6)}$$
- k_{off}, the dissociation rate of the protein interaction with chromatin. This parameter was inverted to obtain the mean residence time of the protein in the chromatin-bound state, $t_{mean}=1/k_{off}$, according to what reported by Phair *et al.* [188].

The mean value obtained for each parameter (±SD) is reported in the Results section (see Figure 2.17A).

A.2.2 $2^{nd}/5^{th}$ nuclear slices geometry

For a protein with dynamics similar to the one found for HOXC13, Beaudoin *et al.* observed that the results of the fit in a 2D approximation was better when fitting only the average fluorescence of the 2^{nd} and 5^{th} slices of a series of 6 in which the nucleus was virtually dissected [214]; we therefore applied our simplified model to calculate the results also in these case; the average recovery in the 5^{th} slice (and loss in the 2^{nd} slice) is calculated as:

$$j(t;D,K,k_{off}) = 1 + \frac{12}{l\left(1+K^{-1}\right)} \sum_{n=1}^{N} \frac{(-1)^{n+1} 2l\left\{\sin\left[(2n+1)\pi/3\right] - \sin\left[(2n+1)\pi/6\right]\right\}}{(2n+1)^2 \pi^2} \left[\tilde{f}\left(\frac{(2n+1)\pi}{l},t\right) + \tilde{c}\left(\frac{(2n+1)\pi}{l},t\right)\right], \tag{Eq. A7}$$

where the same arguments about Eq. A4, including the application of a "normalization" similar to Eq. A5, apply.

In order to decrease the noise in the fitted data and to partially correct for possible asymmetries in the physical systems but not in the model, we averaged the loss in the 2^{nd} slice with the recovery in the 5^{th} slice, and fitted the results with Eq. A7.

Again, by fitting raw FRAP curves to this function we derived the same D_{eff}, F_{free}, and k_{off} parameters described in the previous paragraph. Their mean values (±SD) are reported for each parameter in the Results section (see Figure 2.17B), finding results not significantly different from the ones obtained using the model in Eq. A4. Further comments are reported in Chapter 2, **paragraph 2.4.4**.

A.2.3 Selection criteria of FRAP experiments for the fitting with the diffusion-reaction model

When we applied the fit to the FRAP curves of all analyzed cells (n= 40), expressing transiently wt mCherry-HOXC13 at low levels, we obtained the following results: D_{eff}= 5.10±3.88 $\mu m^2/s$, F_{free}=42.0±24.7%, K_{off}=0.008± 0.003s^{-1}. Values reported in the presentation of results (Figure 2.17A and B) actually refer to a selection of the 16 experiments displaying the most similar geometry with respect to the model. The most stringent criterion in the choice of suitable cells was the length of the bleached area when compared to the whole nucleus. As the initial conditions of the model were that the border between the bleached and unbleached parts is a symmetry plane of the system, we discarded all experiments in which the bleached region was in length less than 44% (or more than 56%) of the whole nucleus. For analogous reasons, nuclei that moved significantly vs. the bleached or unbleached region during the time series were discarded as well. Finally, also nuclei presenting a visible non-homogeneity between bleached and unbleached regions were discarded. This criterion was requested because we considered $[S]_{eq}$ (the concentration of chromatin binding sites) homogeneous in the nucleus as a further simplification for the resolution of the model.

To conclude, we shall underline that the presented selection of suitable experiments was only intended to obtain more accurate estimates of the three binding parameters. Indeed, the fit performed on all acquired cells did not lead to significantly different results when compared to those reported in Figure 2.17.

APPENDIX B

FRET/FLIM Microscopy for the detection of protein-protein interactions in living cells

This appendix reviews some basis of FRET (Fluorescence Resonance Energy Transfer) and FLIM (Fluorescence Lifetime Imaging Microscopy). This outline is intended to provide a general explanation in order to better understand the results presented in **paragraph 2.6**. More extensive and detailed information about FRET and FLIM microscopy can be found elsewhere [193, 259-264].

B.1 Definition of FRET

FRET is a distance-dependent physical process whereby energy is transferred nonradiatively from an excited molecular fluorophore (donor) to another fluorophore (acceptor) by means of dipole-dipole coupling. FRET can provide an accurate measurement of the intra- or inter- molecular proximity between two properly labeled entities, on a molecular scale (10-100 Å).

We consider two different chromophores, with absorption and fluorescence emission spectra as shown in Figure B.1. The donor (D) is the chromophore with absorption at higher energy, and the other one is the acceptor (A). We consider only the ground singlet states (D_a, A_a) and first excited singlet states (D_b, A_b) of each chromophore. Upon excitation, the donor will rapidly lose energy by internal conversion until it reaches the ground vibrational level of the first excited singlet, D_b. If donor emission energies are coincident with acceptor absorption energies, the very weak coupling can permit the following resonance to take place:

$$(D_b, A_a) \xrightarrow{k_T} (D_a, A_b), \qquad \text{(Eq. B1)}$$

where k_T is the rate constant of FRET between a particular (D, A) pair in the considered geometry.

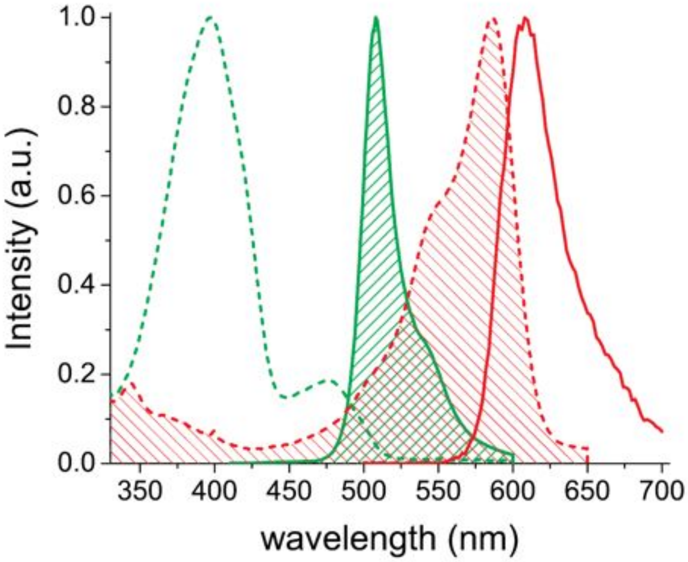


Figure B.1 Spectral properties of the E^{0}GFP/mCherry FRET pair. Normalized absorption spectra (dashed curves) and fluorescence emission spectra (continuous curves) of the donor E^{0}GFP (green) and acceptor mCherry (red) fluorescent proteins. The overlap of the green dashed area (donor emission) with the red dashed area (acceptor absorption) leads to FRET between the two fluorophores.

The energy transfer resonance shifts the relative population of excited donors and acceptors. The donor becomes quenched, the acceptor becomes excited, and if it is fluorescent it can emit, if not, the energy is dissipated as heat. A physical theory predicting the distance-dependence of FRET was proposed by Förster [265, 266]. Förster's theory predicted that energy could be transferred by resonance dipole-dipole mechanism over distances between 10 and 100 Å depending on the spectroscopic parameters of D and A. The rate of transfer is:

$$k_T = \frac{1}{\tau_D} \cdot \left(\frac{R_0}{r} \right)^6, \qquad \text{(Eq. B2)}$$

where τ_D is the fluorescence lifetime of D measured in the absence of A; r is the distance between D and A; R_0 is a distance parameter calculated from the spectroscopic parameters and mutual-dipole-orientations of D and A; when $r=R_0$, $k_T=1/\tau_D$, i.e. half the donor molecules decay by energy transfer and half decay by the usual radiative and non-radiative rates. R_0 is also commonly defined as the distance at which FRET efficiency (E) is 50%.

From a kinetic point of view, E is the ratio of k_T to the total sum of rates of all processes by which an excited D can return to its ground state, including k_T.

$$E = \frac{K_T}{K_T + \sum_i^{i \neq T} k_i} \qquad \text{(Eq.B3)}$$

where the subscript '*i*' refers to the different pathways of deactivation from the D_b state. E can be measured in several ways. Generally, at steady state, the efficiency of transfer can be calculated according to the equation:

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} = \frac{R_0^{\ 6}}{R_0^{\ 6} + r^6}. \qquad \text{(Eq. B4)}$$

Thus, E depends strongly on the D-A distance, and it decays at long distances as r^{-6}.

Because of its dependence on the distance FRET has become an important tool to verify whether labeled proteins are physically linked and to determine distances on the nanometer scale. Technological advances in confocal microscopy imaging, combined with the availability of genetically encoded fluorescent proteins, provide the tools necessary to obtain detection of protein interactions in living cells with spatial and temporal resolution. The spectroscopic properties that are carefully considered in selecting GFPs as workable FRET pairs include sufficient separation in absorption spectra to avoid cross-excitation of the two fluorophores; an overlap between the emission spectrum of the donor and the absoption spectrum of the acceptor to obtain efficient energy transfer; and reasonable separation in emission spectra between donor and acceptor GFPs to allow independent measurement of the fluorescence of each fluorophore.

B.2 FRET detection by FLIM: the ideally suited E^0GFP-mCherry FRET pair

Several techniques can be applied to the evaluation of FRET efficiency. Some of them are based either on the measurement of the donor to acceptor emission ratio or on the quantification of the recovery of donor fluorescence after acceptor photobleaching. One of the main problems of these steady-state FRET techniques is that the concentrations of donor and acceptor may change throughout the sample. In living cells expressing genetically encoded fluorophores, the concentration of fluorescent probes can not be easily controlled. The technique we chose to quantify FRET is based instead on the measurement of donor lifetime, which changes as a function of E; since within reasonable limits the lifetime does not depend on the local concentration of the fluorophore, Fluorescence Lifetime Imaging Microscopy (FLIM) has become a valuable tool to map protein associations on the molecular scale in living cells.

Consider the E^{0}GFP-mCherry FRET pair (with spectral properties reported in Figure B.1). In our FLIM measurements, the use of this FRET pair offered several advantages [237]. E^{0}GFP is the H231L-F64L mutant of wt GFP and belongs to the UV-excitable FP family. It was chosen as donor for mainly two reasons: i) it displays a large Stokes shift (i.e. separation from absorption and emission spectra, this is important to avoid cross-excitation artifacts); ii) it displays a fluorescence decay curve (Figure B.2, green curve) which is optimally fitted by a monoexponential function, of the type:

$$I(t) = e^{-\frac{t}{\tau_D}}, \qquad \text{(Eq. B5)}$$

where I is the donor fluorescence intensity, t is time, and τ_D is the fluorescence lifetime of D.

The acceptor (mCherry) is one of the most commonly used monomeric proteins derived from stony-coral DsRed [167, 170, 171], and displays: i) fast maturation, ii) large absorption cross-section, iii) high photostability. It was already reported as a good FRET acceptor [267]. In the presence of a close mCherry molecule, E^{0}GFP fluorescence lifetime becomes shorter when compared to its lifetime in absence of acceptor (Figure B.2, red and blue curves).

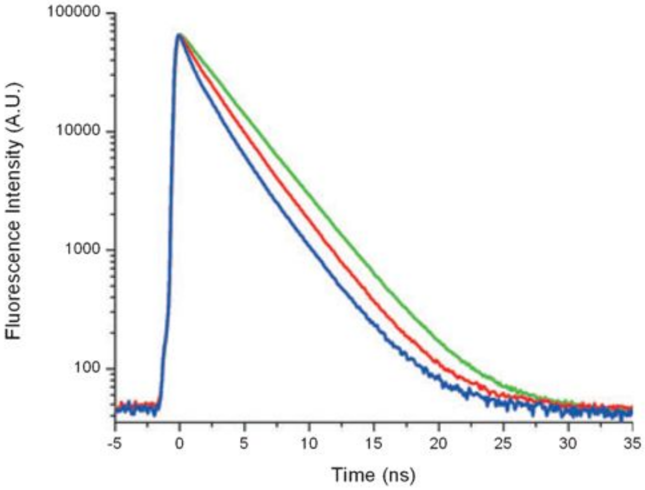


Figure B.2 Fluorescence intensity decays for E^{0}GFP in the presence of mCherry. Fluorescence decay upon pulsed excitation at 403 nm for E^{0}GFP (green line), and two FRET constructs where E^{0}GFP is directly fused to mCherry with linkers of different lengths (red and blue curves). The green donor shows a monoexponential decay while the two FRET constructs display faster components owing to FRET. For more details about this Figure, see Albertazzi *et al.* [237].

The total donor decay curve, resulting on a combination of interacting and non-interacting FRET couples, can be approximated by the following double exponential function:

$$I(t) = a_1 \cdot e^{-\frac{t}{\tau_{DA}}} + a_2 \cdot e^{-\frac{t}{\tau_D}} , \qquad \text{(Eq. B6)}$$

where τ_{DA} is a fast donor lifetime component from the interacting (quenched) donor molecules and τ_D is the lifetime of the non-interacting donor, and is assumed to be equal to that of the donor alone. If the labeling is complete, as it is expected in cells expressing fusion proteins of the GFPs, the decay components a_1 and a_2 directly represent the fractions of interacting and non-interacting proteins. Thus, FLIM double exponential decay analysis using the E^0GFP/mCherry FRET pair directly delivers the lifetimes of the interacting and non-interacting donor species, τ_{DA} and τ_D, and the intensity factors of the two decay components. This is not always the case when using other well-established FRET pairs, like the CFP/YFP one, as CFPs present *per se* a multiexponential decay that makes lifetime analyses much more complicated [268].

B.3 Distance calculation by FLIM

In FLIM experiments the FRET efficiency (E) can be calculated from the lifetimes of the interacting and non-interacting donor species, τ_{DA} and τ_D, according to:

$$E = 1 - \frac{\tau_{DA}}{\tau_D} . \qquad \text{(Eq. B7)}$$

The calculated E value can be used to estimate the mean D-A distance when the donor is interacting with the acceptor using Eq. B4. Accordingly, we used this approach to calculate the mean distance between interacting species labeled with E^0GFP and mCherry.

In our experiments, we found a significant lifetime reduction of E^0GFP fused to Cdc6 and ORC2 when these were in turn co-expressed with mCherry-HOXC13 (Figure 2.30). For these two cases, the FRET efficiency was calculated using Eq. B7 (see Figure B.3 for the image analysis leading to E calculation) and further converted in a mean distance of interaction using Eq. B4, with R_0= 5.1nm for the E^0GFP/mCherry FRET pair [237]. We finally estimated that the actual vicinity for ORC2 and HOXC13 is 6.40±0.04 nm, while that between Cdc6 and HOXC13 is 6.72±0.07 nm. A summary of obtained results is reported in Table B.1.

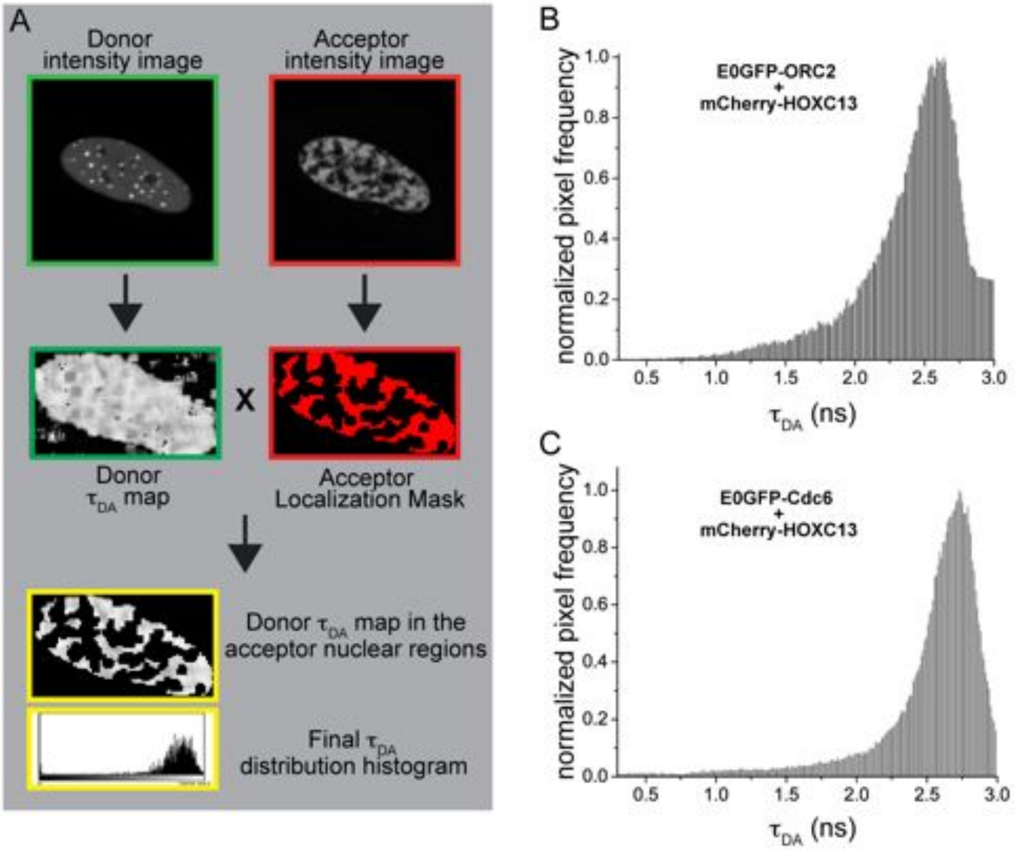


Figure B.3 Image analysis for the calculation of *E* for the ORC2-HOXC13 and Cdc6-HOXC13 interaction. A) Summary of the image analysis method. The two top images represent a typical analyzed cell in which the donor (fused to ORC2 or Cdc6) was expressed with the acceptor (fused with HOXC13). The donor image was fitted with Eq. B6 in all nuclear pixels and we derived a lifetime map of the donor interacting with the acceptor (τ_{DA} map, left middle image). The acceptor image was converted into a mask of acceptor localization (right middle image) and superimposed onto the τ_{DA} map. In this way, it was possible to select only the τ_{DA}-map pixels in which both donor and acceptor were present. The final τ_{DA} distributions derived from all analyzed cells were summed and a cumulative τ_{DA} distribution was obtained for the study of ORC2-HOXC13 (B) and Cdc6-HOXC13 (C) interactions. The weighted mean values (±SEM) of these distributions are reported in Table B.1.

Table B.1 *E* and distance calculation by FLIM. τ_D and τ_{DA} components (mean±SEM) derived from FLIM image analysis were used to calculate *E* using Eq. B7, and further converted into distance of interaction using Eq. B4, with R_0= 5.1nm [237]. τ_D values are the same as those reported in the first row of Table 2.2. For the case of ORC2 protein, only a selection of cells (those displaying the closest ORC2-HOXC13 interaction, as explained in Figure 2.32) were considered in the τ_{DA} calculations (n refers to the number of analyzed cells).

Table B.1	**τ_D (ns)**	**τ_{DA} (ns)**	***E* (%)**	**Distance (nm)**
E^0GFP-ORC2 + mCherry-HOXC13	3.03±0.04 n=51	2.41±0.01 n=48	20.46±0.36	6.40±0.04
E^0GFP-Cdc6 + mCherry-HOXC13	3.00±0.04 n=49	2.52±0.05 n=106	16.00±0.53	6.72±0.07

Acknowledgments

There are many persons I want to thank, who helped me during this Ph.D. program. My first thought goes to Fabio Beltram. He welcomed me like a true member of his lab and, thanks to this, I was put in the ideal working conditions for the project to be developed. I also want to thank Daniele Arosio, for his critical, punctual and often decisive suggestions about how to get and present data. My thanks go to Stefano Luin whom I involved in the last part of the project, which turned out to me to be the most interesting one. And really, without him, I just could not have done it. Many thanks also to Giovanni Signore, for always being helpful for whatever reason at anytime of the day, especially during the writing of this thesis. Thanks to Gulnara Abdurashidova for stimulating discussions and Michela Serresi for optimal suggestions in cell culture procedures. Thanks also to Paolo Faraci, for excellent technical support. Many thanks go to Laura Comelli and Barbara D'Innocenzo, who worked in parallel with me in the HOX project and gave me important information and reagents. And last, but just to remember them better, my thanks are for my closest lab-mates: Nanda Ricci, Andrea Callegari and Valerio Voliani, with whom I shared lab-life and had the best and most useful discussions.

References

1. Jacob, F. and S. Brenner, *[On the regulation of DNA synthesis in bacteria: the hypothesis of the replicon.].* C R Hebd Seances Acad Sci, 1963. **256**: p. 298-300.
2. Kornberg, A., *Control of initiation of the Escherichia coli chromosome.* Cold Spring Harb Symp Quant Biol, 1991. **56**: p. 275-8.
3. DePamphilis, M.L., *Eukaryotic DNA replication: anatomy of an origin.* Annu Rev Biochem, 1993. **62**: p. 29-63.
4. Huberman, J.A., *Cell cycle. A licence to replicate.* Nature, 1995. **375**(6530): p. 360-1.
5. Stillman, B., *Cell cycle control of DNA replication.* Science, 1996. **274**(5293): p. 1659-64.
6. Sclafani, R.A. and T.M. Holzen, *Cell cycle regulation of DNA replication.* Annu Rev Genet, 2007. **41**: p. 237-80.
7. Bell, S.P. and A. Dutta, *DNA replication in eukaryotic cells.* Annu Rev Biochem, 2002. **71**: p. 333-74.
8. Kunkel, T.A. and P.M. Burgers, *Dividing the workload at a eukaryotic replication fork.* Trends Cell Biol, 2008. **18**(11): p. 521-7.
9. Conti, C., J.A. Seiler, and Y. Pommier, *The mammalian DNA replication elongation checkpoint: implication of Chk1 and relationship with origin firing as determined by single DNA molecule and single cell analyses.* Cell Cycle, 2007. **6**(22): p. 2760-7.
10. Duggin, I.G., et al., *The replication fork trap and termination of chromosome replication.* Mol Microbiol, 2008. **70**(6): p. 1323-33.
11. Bussiere, D.E. and D. Bastia, *Termination of DNA replication of bacterial and plasmid chromosomes.* Mol Microbiol, 1999. **31**(6): p. 1611-8.
12. DePamphilis, M.L., *Replication origins in metazoan chromosomes: fact or fiction?* Bioessays, 1999. **21**(1): p. 5-16.
13. Gilbert, D.M., *Making sense of eukaryotic DNA replication origins.* Science, 2001. **294**(5540): p. 96-100.
14. Cadoret, J.C., et al., *Genome-wide studies highlight indirect links between human replication origins and gene regulation.* Proc Natl Acad Sci U S A, 2008. **105**(41): p. 15837-42.

15. Karnani, N., et al., *Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection*. Mol Biol Cell, 2010. **21**(3): p. 393-404.
16. Gerbi, S.A. and A.K. Bielinsky, *DNA replication and chromatin*. Curr Opin Genet Dev, 2002. **12**(2): p. 243-8.
17. Gilbert, D.M., *In search of the holy replicator*. Nat Rev Mol Cell Biol, 2004. **5**(10): p. 848-55.
18. McNairn, A.J. and D.M. Gilbert, *Epigenomic replication: linking epigenetics to DNA replication*. Bioessays, 2003. **25**(7): p. 647-56.
19. Mechali, M., *DNA replication origins: from sequence specificity to epigenetics*. Nat Rev Genet, 2001. **2**(8): p. 640-5.
20. Vas, A. and J. Leatherwood, *Where does DNA replication start in archaea?* Genome Biol, 2000. **1**(3): p. REVIEWS1020.
21. Mendez, J. and B. Stillman, *Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis*. Mol Cell Biol, 2000. **20**(22): p. 8602-12.
22. Diffley, J.F., *Regulation of early events in chromosome replication*. Curr Biol, 2004. **14**(18): p. R778-86.
23. Tatsumi, Y., et al., *The ORC1 cycle in human cells: I. cell cycle-regulated oscillation of human ORC1*. J Biol Chem, 2003. **278**(42): p. 41528-34.
24. Wohlschlegel, J.A., et al., *Inhibition of eukaryotic DNA replication by geminin binding to Cdt1*. Science, 2000. **290**(5500): p. 2309-12.
25. Diffley, J.F., *DNA replication: building the perfect switch*. Curr Biol, 2001. **11**(9): p. R367-70.
26. Noton, E. and J.F. Diffley, *CDK inactivation is the only essential function of the APC/C and the mitotic exit network proteins for origin resetting during mitosis*. Mol Cell, 2000. **5**(1): p. 85-95.
27. Tanaka, S. and J.F. Diffley, *Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation*. Genes Dev, 2002. **16**(20): p. 2639-49.
28. Sporbert, A., et al., *DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters*. Mol Cell, 2002. **10**(6): p. 1355-65.
29. Thomson, A.M., P.J. Gillespie, and J.J. Blow, *Replication factory activation can be decoupled from the replication timing program by modulating Cdk levels*. J Cell Biol, 2010. **188**(2): p. 209-21.
30. Zou, L. and B. Stillman, *Formation of a preinitiation complex by S-phase cyclin CDK-dependent loading of Cdc45p onto chromatin*. Science, 1998. **280**(5363): p. 593-6.

31. Schepers, A. and P. Papior, *Why are we where we are? Understanding replication origins and initiation sites in eukaryotes using ChIP-approaches*. Chromosome Res, 2010. **18**(1): p. 63-77.

32. Norio, P., et al., *Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development*. Mol Cell, 2005. **20**(4): p. 575-87.

33. Hiratani, I., et al., *Global reorganization of replication domains during embryonic stem cell differentiation*. PLoS Biol, 2008. **6**(10): p. e245.

34. MacAlpine, H.K., et al., *Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading*. Genome Res, 2010. **20**(2): p. 201-11.

35. Bartek, J., C. Lukas, and J. Lukas, *Checking on DNA damage in S phase*. Nat Rev Mol Cell Biol, 2004. **5**(10): p. 792-804.

36. Nyberg, K.A., et al., *Toward maintaining the genome: DNA damage and replication checkpoints*. Annu Rev Genet, 2002. **36**: p. 617-56.

37. Zhou, B.B. and S.J. Elledge, *The DNA damage response: putting checkpoints in perspective*. Nature, 2000. **408**(6811): p. 433-9.

38. Santocanale, C. and J.F. Diffley, *A Mec1- and Rad53-dependent checkpoint controls late-firing origins of DNA replication*. Nature, 1998. **395**(6702): p. 615-8.

39. Shirahige, K., et al., *Regulation of DNA-replication origins during cell-cycle progression*. Nature, 1998. **395**(6702): p. 618-21.

40. Mimura, S., et al., *Phosphorylation-dependent binding of mitotic cyclins to Cdc6 contributes to DNA replication control*. Nature, 2004. **431**(7012): p. 1118-23.

41. Lau, E., et al., *The functional role of Cdc6 in S-G2/M in mammalian cells*. EMBO Rep, 2006. **7**(4): p. 425-30.

42. Bell, S.P. and B. Stillman, *ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex*. Nature, 1992. **357**(6374): p. 128-34.

43. Vashee, S., et al., *Sequence-independent DNA binding and replication initiation by the human origin recognition complex*. Genes Dev, 2003. **17**(15): p. 1894-908.

44. Duncker, B.P., I.N. Chesnokov, and B.J. McConkey, *The origin recognition complex protein family*. Genome Biol, 2009. **10**(3): p. 214.

45. Aggarwal, B.D. and B.R. Calvi, *Chromatin regulates origin activity in Drosophila follicle cells*. Nature, 2004. **430**(6997): p. 372-6.

46. Danis, E., et al., *Specification of a DNA replication origin by a transcription complex*. Nat Cell Biol, 2004. **6**(8): p. 721-30.

47. Speck, C., et al., *ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA*. Nat Struct Mol Biol, 2005. **12**(11): p. 965-71.

48. DePamphilis, M.L., *The 'ORC cycle': a novel pathway for regulating eukaryotic DNA replication*. Gene, 2003. **310**: p. 1-15.

49. Ohta, S., et al., *The ORC1 cycle in human cells: II. Dynamic changes in the human ORC complex during the cell cycle*. J Biol Chem, 2003. **278**(42): p. 41535-40.

50. Li, C.J., A. Vassilev, and M.L. DePamphilis, *Role for Cdk1 (Cdc2)/cyclin A in preventing the mammalian origin recognition complex's largest subunit (Orc1) from binding to chromatin during mitosis*. Mol Cell Biol, 2004. **24**(13): p. 5875-86.

51. Zhang, Z., et al., *Structure and function of the BAH-containing domain of Orc1p in epigenetic silencing*. Embo J, 2002. **21**(17): p. 4600-11.

52. Pak, D.T., et al., *Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes*. Cell, 1997. **91**(3): p. 311-23.

53. Prasanth, S.G., et al., *Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance*. Embo J, 2004. **23**(13): p. 2651-63.

54. Chesnokov, I.N., *Multiple functions of the origin recognition complex*. Int Rev Cytol, 2007. **256**: p. 69-109.

55. Prasanth, S.G., K.V. Prasanth, and B. Stillman, *Orc6 involved in DNA replication, chromosome segregation, and cytokinesis*. Science, 2002. **297**(5583): p. 1026-31.

56. Prasanth, S.G., et al., *Dynamics of pre-replication complex proteins during the cell division cycle*. Philos Trans R Soc Lond B Biol Sci, 2004. **359**(1441): p. 7-16.

57. Hemerly, A.S., et al., *Orc1 controls centriole and centrosome copy number in human cells*. Science, 2009. **323**(5915): p. 789-93.

58. Blow, J.J. and B. Hodgson, *Replication licensing--defining the proliferative state?* Trends Cell Biol, 2002. **12**(2): p. 72-8.

59. Randell, J.C., et al., *Sequential ATP hydrolysis by Cdc6 and ORC directs loading of the Mcm2-7 helicase*. Mol Cell, 2006. **21**(1): p. 29-39.

60. Robinson, N.P., et al., *Identification of two origins of replication in the single chromosome of the archaeon Sulfolobus solfataricus*. Cell, 2004. **116**(1): p. 25-38.

61. Remus, D., et al., *Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing*. Cell, 2009. **139**(4): p. 719-30.

62. Nguyen, V.Q., C. Co, and J.J. Li, *Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms*. Nature, 2001. **411**(6841): p. 1068-73.

63. Drury, L.S., G. Perkins, and J.F. Diffley, *The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle*. Curr Biol, 2000. **10**(5): p. 231-40.

64. Delmolino, L.M., P. Saha, and A. Dutta, *Multiple mechanisms regulate subcellular localization of human CDC6*. J Biol Chem, 2001. **276**(29): p. 26947-54.

65. McGarry, T.J. and M.W. Kirschner, *Geminin, an inhibitor of DNA replication, is degraded during mitosis*. Cell, 1998. **93**(6): p. 1043-53.

66. Luo, L., et al., *The cell-cycle regulator geminin inhibits Hox function through direct and polycomb-mediated interactions*. Nature, 2004. **427**(6976): p. 749-53.

67. Drury, L.S. and J.F. Diffley, *Factors affecting the diversity of DNA replication licensing control in eukaryotes*. Curr Biol, 2009. **19**(6): p. 530-5.

68. Adachi, Y., J. Usukura, and M. Yanagida, *A globular complex formation by Nda1 and the other five members of the MCM protein family in fission yeast*. Genes Cells, 1997. **2**(7): p. 467-79.

69. Wyrick, J.J., et al., *Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins*. Science, 2001. **294**(5550): p. 2357-60.

70. Morgan, D.O., *The Cell Cycle*. 2007, London: New Science.

71. Sheu, Y.J. and B. Stillman, *Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression*. Mol Cell, 2006. **24**(1): p. 101-13.

72. Masai, H., et al., *Phosphorylation of MCM4 by Cdc7 kinase facilitates its interaction with Cdc45 on the chromatin*. J Biol Chem, 2006. **281**(51): p. 39249-61.

73. Yabuuchi, H., et al., *Ordered assembly of Sld3, GINS and Cdc45 is distinctly regulated by DDK and CDK for activation of replication origins*. Embo J, 2006. **25**(19): p. 4663-74.

74. Aparicio, T., et al., *The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication*. Nucleic Acids Res, 2009. **37**(7): p. 2087-95.

75. Aparicio, O.M., A.M. Stout, and S.P. Bell, *Differential assembly of Cdc45p and DNA polymerases at early and late origins of DNA replication*. Proc Natl Acad Sci U S A, 1999. **96**(16): p. 9130-5.

76. Zegerman, P. and J.F. Diffley, *Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast*. Nature, 2007. **445**(7125): p. 281-5.

77. Kubota, Y., et al., *A novel ring-like complex of Xenopus proteins essential for the initiation of DNA replication*. Genes Dev, 2003. **17**(9): p. 1141-52.

78. Boskovic, J., et al., *Molecular architecture of the human GINS complex*. EMBO Rep, 2007. **8**(7): p. 678-84.

79. Forsburg, S.L., *Eukaryotic MCM proteins: beyond replication initiation*. Microbiol Mol Biol Rev, 2004. **68**(1): p. 109-31.

80. Izumi, M., et al., *The human homolog of Saccharomyces cerevisiae Mcm10 interacts with replication factors and dissociates from nuclease-resistant nuclear structures in G(2) phase*. Nucleic Acids Res, 2000. **28**(23): p. 4769-77.

81. Moyer, S.E., P.W. Lewis, and M.R. Botchan, *Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase*. Proc Natl Acad Sci U S A, 2006. **103**(27): p. 10236-41.

82. Marahrens, Y. and B. Stillman, *A yeast chromosomal origin of DNA replication defined by multiple functional elements*. Science, 1992. **255**(5046): p. 817-23.

83. Hyrien, O., C. Maric, and M. Mechali, *Transition in specification of embryonic metazoan DNA replication origins*. Science, 1995. **270**(5238): p. 994-7.

84. Abdurashidova, G., et al., *Start sites of bidirectional DNA synthesis at the human lamin B2 origin*. Science, 2000. **287**(5460): p. 2023-6.

85. Keller, C., et al., *The origin recognition complex marks a replication origin in the human TOP1 gene promoter*. J Biol Chem, 2002. **277**(35): p. 31430-40.

86. Ladenburger, E.M., C. Keller, and R. Knippers, *Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication*. Mol Cell Biol, 2002. **22**(4): p. 1036-48.

87. Cvetic, C. and J.C. Walter, *Eukaryotic origins of DNA replication: could you please be more specific?* Semin Cell Dev Biol, 2005. **16**(3): p. 343-53.

88. Remus, D., E.L. Beall, and M.R. Botchan, *DNA topology, not DNA sequence, is a critical determinant for Drosophila ORC-DNA binding*. Embo J, 2004. **23**(4): p. 897-907.

89. Antequera, F., *Genomic specification and epigenetic regulation of eukaryotic DNA replication origins*. Embo J, 2004. **23**(22): p. 4365-70.

90. Bielinsky, A.K., *Replication origins: why do we need so many?* Cell Cycle, 2003. **2**(4): p. 307-9.

91. Li, F., et al., *Spatial distribution and specification of mammalian replication origins during G1 phase*. J Cell Biol, 2003. **161**(2): p. 257-66.

92. Abdurashidova, G., et al., *Functional interactions of DNA topoisomerases with a human replication origin*. Embo J, 2007. **26**(4): p. 998-1009.

93. DePamphilis, M.L., et al., *Regulating the licensing of DNA replication origins in metazoa*. Curr Opin Cell Biol, 2006. **18**(3): p. 231-9.

94. Minami, H., et al., *Binding of AlF-C, an Orc1-binding transcriptional regulator, enhances replicator activity of the rat aldolase B origin*. Mol Cell Biol, 2006. **26**(23): p. 8770-80.

95. Atanasiu, C., et al., *ORC binding to TRF2 stimulates OriP replication*. EMBO Rep, 2006. **7**(7): p. 716-21.

96. Sibani, S., G.B. Price, and M. Zannis-Hadjopoulos, *Ku80 binds to human replication origins prior to the assembly of the ORC complex*. Biochemistry, 2005. **44**(21): p. 7885-96.

97. Rampakakis, E., D. Di Paola, and M. Zannis-Hadjopoulos, *Ku is involved in cell growth, DNA replication and G1-S transition*. J Cell Sci, 2008. **121**(Pt 5): p. 590-600.

98. Schepers, A., et al., *Human origin recognition complex binds to the region of the latent origin of DNA replication of Epstein-Barr virus*. Embo J, 2001. **20**(16): p. 4588-602.

99. Thomae, A.W., et al., *Interaction between HMGA1a and the origin recognition complex creates site-specific replication origins*. Proc Natl Acad Sci U S A, 2008. **105**(5): p. 1692-7.

100. Dominguez-Sola, D., et al., *Non-transcriptional control of DNA replication by c-Myc*. Nature, 2007. **448**(7152): p. 445-51.

101. de Stanchina, E., et al., *Selection of homeotic proteins for binding to a human DNA replication origin*. J Mol Biol, 2000. **299**(3): p. 667-80.

102. Gabellini, D., et al., *Early mitotic degradation of the homeoprotein HOXC10 is potentially linked to cell cycle progression*. Embo J, 2003. **22**(14): p. 3715-24.

103. Salsi, V., et al., *HOXD13 binds DNA replication origins to promote origin licensing and is inhibited by geminin*. Mol Cell Biol, 2009. **29**(21): p. 5775-88.

104. Falaschi, A., G. Abdurashidova, and G. Biamonti, *DNA replication, development and cancer: a homeotic connection?* Crit Rev Biochem Mol Biol, 2010. **45**: p. 14-22.

105. Lewis, E.B., *A gene complex controlling segmentation in Drosophila*. Nature, 1978. **276**(5688): p. 565-70.

106. Gehring, W.J., et al., *Homeodomain-DNA recognition*. Cell, 1994. **78**(2): p. 211-23.

107. Gehring, W.J., M. Affolter, and T. Burglin, *Homeodomain proteins*. Annu Rev Biochem, 1994. **63**: p. 487-526.

108. Gehring, W.J. and Y. Hiromi, *Homeotic genes and the homeobox*. Annu Rev Genet, 1986. **20**: p. 147-73.

109. Levine, M. and T. Hoey, *Homeobox proteins as sequence-specific transcription factors*. Cell, 1988. **55**(4): p. 537-40.

110. McGinnis, W. and R. Krumlauf, *Homeobox genes and axial patterning*. Cell, 1992. **68**(2): p. 283-302.

111. Scott, M.P. and A.J. Weiner, *Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila*. Proc Natl Acad Sci U S A, 1984. **81**(13): p. 4115-9.

112. Stein, S., et al., *Checklist: vertebrate homeobox genes*. Mech Dev, 1996. **55**(1): p. 91-108.

113. Care, A., et al., *HOXB7: a key factor for tumor-associated angiogenic switch*. Cancer Res, 2001. **61**(17): p. 6532-9.

114. Myers, C., A. Charboneau, and N. Boudreau, *Homeobox B3 promotes capillary morphogenesis and angiogenesis*. J Cell Biol, 2000. **148**(2): p. 343-51.

115. Ford, H.L., et al., *Abrogation of the G2 cell cycle checkpoint associated with overexpression of HSIX1: a possible mechanism of breast carcinogenesis*. Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12608-13.

116. Magli, M.C., et al., *Coordinate regulation of HOX genes in human hematopoietic cells*. Proc Natl Acad Sci U S A, 1991. **88**(14): p. 6348-52.

117. Cillo, C., et al., *Homeobox genes and cancer*. Exp Cell Res, 1999. **248**(1): p. 1-9.

118. Troy, P.J., et al., *Transcriptional repression of peri-implantation EMX2 expression in mammalian reproduction by HOXA10*. Mol Cell Biol, 2003. **23**(1): p. 1-13.

119. Shi, X., et al., *Hoxa-9 represses transforming growth factor-beta-induced osteopontin gene transcription*. J Biol Chem, 2001. **276**(1): p. 850-5.

120. Veraksa, A., M. Del Campo, and W. McGinnis, *Developmental patterning genes and their conserved functions: from model organisms to humans*. Mol Genet Metab, 2000. **69**(2): p. 85-100.

121. Scott, M.P., *Vertebrate homeobox gene nomenclature*. Cell, 1992. **71**(4): p. 551-3.

122. Krumlauf, R., *Hox genes in vertebrate development*. Cell, 1994. **78**(2): p. 191-201.

123. Meyer, A., *The evolution of body plans: HOM/Hox cluster evolution, model systems, and the importance of phylogeny*. New uses for new phylogenies, ed. P.H.B. Harvey, A.J.L.; Smith, J.M.; Nees, S. 1996, Oxford: Oxford University Press.

124. Greer, J.M., et al., *Maintenance of functional equivalence during paralogous Hox gene evolution*. Nature, 2000. **403**(6770): p. 661-5.

125. Takahashi, Y., et al., *Expression profiles of 39 HOX genes in normal human adult organs and anaplastic thyroid cancer cell lines by quantitative real-time RT-PCR system*. Exp Cell Res, 2004. **293**(1): p. 144-53.

126. McGinnis, W., et al., *A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans*. Cell, 1984. **37**(2): p. 403-8.

127. Qian, Y.Q., et al., *The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors*. Cell, 1989. **59**(3): p. 573-80.

128. Billeter, M., et al., *Determination of the three-dimensional structure of the Antennapedia homeodomain from Drosophila in solution by 1H nuclear magnetic resonance spectroscopy*. J Mol Biol, 1990. **214**(1): p. 183-97.

129. Fraenkel, E., et al., *Engrailed homeodomain-DNA complex at 2.2 A resolution: a detailed view of the interface and comparison with other engrailed structures*. J Mol Biol, 1998. **284**(2): p. 351-61.

130. Kissinger, C.R., et al., *Crystal structure of an engrailed homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions*. Cell, 1990. **63**(3): p. 579-90.

131. Hirsch, J.A. and A.K. Aggarwal, *Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity*. Embo J, 1995. **14**(24): p. 6280-91.

132. Klemm, J.D., et al., *Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules*. Cell, 1994. **77**(1): p. 21-32.

133. Li, T., et al., *Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA*. Science, 1995. **270**(5234): p. 262-9.

134. Wilson, D.S., et al., *High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA*. Cell, 1995. **82**(5): p. 709-19.

135. Wolberger, C., et al., *Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions*. Cell, 1991. **67**(3): p. 517-28.

136. Billeter, M., *Homeodomain-type DNA recognition*. Prog Biophys Mol Biol, 1996. **66**(3): p. 211-25.

137. Connolly, J.P., J.G. Augustine, and C. Francklyn, *Mutational analysis of the engrailed homeodomain recognition helix by phage display*. Nucleic Acids Res, 1999. **27**(4): p. 1182-9.

138. Castronovo, V., et al., *Homeobox genes: potential candidates for the transcriptional control of the transformed and invasive phenotype*. Biochem Pharmacol, 1994. **47**(1): p. 137-43.

139. Ades, S.E. and R.T. Sauer, *Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex*. Biochemistry, 1995. **34**(44): p. 14601-8.

140. Piper, D.E., et al., *Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation*. Cell, 1999. **96**(4): p. 587-97.

141. Mann, R.S., K.M. Lelli, and R. Joshi, *Hox specificity unique roles for cofactors and collaborators*. Curr Top Dev Biol, 2009. **88**: p. 63-101.

142. Ades, S.E. and R.T. Sauer, *Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50*. Biochemistry, 1994. **33**(31): p. 9187-94.

143. Berry, M. and W. Gehring, *Phosphorylation status of the SCR homeodomain determines its functional activity: essential role for protein phosphatase 2A,B'*. Embo J, 2000. **19**(12): p. 2946-57.

144. Yaron, Y., et al., *Identification of novel functional regions important for the activity of HOXB7 in mammalian cells*. J Immunol, 2001. **166**(8): p. 5058-67.

145. Iyaguchi, D., et al., *DNA recognition mechanism of the ONECUT homeodomain of transcription factor HNF-6*. Structure, 2007. **15**(1): p. 75-83.

146. Joshi, R., et al., *Functional specificity of a Hox protein mediated by the recognition of minor groove structure*. Cell, 2007. **131**(3): p. 530-43.

147. Mann, R.S. and S.K. Chan, *Extra specificity from extradenticle: the partnership between HOX and PBX/EXD homeodomain proteins*. Trends Genet, 1996. **12**(7): p. 258-62.

148. Moens, C.B. and L. Selleri, *Hox cofactors in vertebrate development*. Dev Biol, 2006. **291**(2): p. 193-206.

149. Shen, W.F., et al., *AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins*. Mol Cell Biol, 1997. **17**(11): p. 6448-58.

150. Shen, W.F., et al., *The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target*. J Biol Chem, 1997. **272**(13): p. 8198-206.

151. LaRonde-LeBlanc, N.A. and C. Wolberger, *Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior*. Genes Dev, 2003. **17**(16): p. 2060-72.

152. Rohs, R., et al., *The role of DNA shape in protein-DNA recognition*. Nature, 2009. **461**(7268): p. 1248-53.

153. Lawrence, H.J. and C. Largman, *Homeobox genes in normal hematopoiesis and leukemia*. Blood, 1992. **80**(10): p. 2445-53.

154. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer*. Cell, 2000. **100**(1): p. 57-70.

155. Cillo, C., et al., *Homeobox genes in normal and malignant cells*. J Cell Physiol, 2001. **188**(2): p. 161-9.

156. Calvo, R., et al., *Altered HOX and WNT7A expression in human lung cancer*. Proc Natl Acad Sci U S A, 2000. **97**(23): p. 12776-81.

157. Li, H., C.J. Huang, and K.B. Choo, *Expression of homeobox genes in cervical cancer*. Gynecol Oncol, 2002. **84**(2): p. 216-21.

158. La Starza, R., et al., *Human homeobox gene HOXC13 is the partner of NUP98 in adult acute myeloid leukemia with t(11;12)(p15;q13)*. Genes Chromosomes Cancer, 2003. **36**(4): p. 420-3.

159. Lam, D.H. and P.D. Aplan, *NUP98 gene fusions in hematologic malignancies*. Leukemia, 2001. **15**(11): p. 1689-95.

160. Gurevich, R.M., P.D. Aplan, and R.K. Humphries, *NUP98-topoisomerase I acute myeloid leukemia-associated fusion gene has potent leukemogenic activities independent of an engineered catalytic site mutation*. Blood, 2004. **104**(4): p. 1127-36.

161. Nebral, K., et al., *NUP98 is fused to topoisomerase (DNA) IIbeta 180 kDa (TOP2B) in a patient with acute myeloid leukemia with a new t(3;11)(p24;p15)*. Clin Cancer Res, 2005. **11**(18): p. 6489-94.

162. Boudreau, N., et al., *Induction of the angiogenic phenotype by Hox D3*. J Cell Biol, 1997. **139**(1): p. 257-64.

163. Kim, A.L., et al., *Conformational and molecular basis for induction of apoptosis by a p53 C-terminal peptide in human cancer cells*. J Biol Chem, 1999. **274**(49): p. 34924-31.

164. Raman, V., et al., *Compromised HOXA5 function can limit p53 expression in human breast tumours*. Nature, 2000. **405**(6789): p. 974-8.

165. Abdurashidova, G., et al., *Cell cycle modulation of protein-DNA interactions at a human replication origin*. Embo J, 1998. **17**(10): p. 2961-9.

166. Luo, L. and M. Kessel, *Geminin coordinates cell cycle and developmental control*. Cell Cycle, 2004. **3**(6): p. 711-4.

167. Shaner, N.C., P.A. Steinbach, and R.Y. Tsien, *A guide to choosing fluorescent proteins*. Nat Methods, 2005. **2**(12): p. 905-9.

168. Baird, G.S., D.A. Zacharias, and R.Y. Tsien, *Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral*. Proc Natl Acad Sci U S A, 2000. **97**(22): p. 11984-9.

169. Piatkevich, K.D. and V.V. Verkhusha, *Advances in engineering of fluorescent proteins and photoactivatable proteins with red emission*. Curr Opin Chem Biol, 2009. **14**(1): p. 23-9.

170. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein*. Nat Biotechnol, 2004. **22**(12): p. 1567-72.

171. Shaner, N.C., et al., *Improving the photostability of bright monomeric orange and red fluorescent proteins*. Nat Methods, 2008. **5**(6): p. 545-51.

172. Shu, X., et al., *Novel chromophores and buried charges control color in mFruits*. Biochemistry, 2006. **45**(32): p. 9639-47.

173. Tsukamoto, T., et al., *Visualization of gene activity in living cells*. Nat Cell Biol, 2000. **2**(12): p. 871-8.

174. Platani, M., et al., *Cajal body dynamics and association with chromatin are ATP-dependent*. Nat Cell Biol, 2002. **4**(7): p. 502-8.

175. Gerlich, D., et al., *Global chromosome positions are transmitted through mitosis in mammalian cells*. Cell, 2003. **112**(6): p. 751-64.

176. Voss, T.C., I.A. Demarco, and R.N. Day, *Quantitative imaging of protein interactions in the cell nucleus*. Biotechniques, 2005. **38**(3): p. 413-24.

177. Misteli, T. and E. Soutoglou, *The emerging role of nuclear architecture in DNA repair and genome maintenance*. Nat Rev Mol Cell Biol, 2009. **10**(4): p. 243-54.

178. Schneider, R. and R. Grosschedl, *Dynamics and interplay of nuclear architecture, genome organization, and gene expression*. Genes Dev, 2007. **21**(23): p. 3027-43.

179. Rippe, K., *Dynamic organization of the cell nucleus*. Curr Opin Genet Dev, 2007. **17**(5): p. 373-80.

180. Misteli, T., *Beyond the sequence: cellular organization of genome function*. Cell, 2007. **128**(4): p. 787-800.

181. Groth, A., et al., *Chromatin challenges during DNA replication and repair*. Cell, 2007. **128**(4): p. 721-33.

182. Mueller, F., P. Wach, and J.G. McNally, *Evidence for a common mode of transcription factor interaction with chromatin as revealed by improved quantitative fluorescence recovery after photobleaching*. Biophys J, 2008. **94**(8): p. 3323-39.

183. Schmiedeberg, L., et al., *High- and low-mobility populations of HP1 in heterochromatin of mammalian cells*. Mol Biol Cell, 2004. **15**(6): p. 2819-33.

184. Souza, P.P., et al., *The histone methyltransferase SUV420H2 and Heterochromatin Proteins HP1 interact but show different dynamic behaviours*. BMC Cell Biol, 2009. **10**: p. 41.

185. Sprague, B.L., et al., *Analysis of binding at a single spatially localized cluster of binding sites by fluorescence recovery after photobleaching*. Biophys J, 2006. **91**(4): p. 1169-91.

186. Sprague, B.L., et al., *Analysis of binding reactions by fluorescence recovery after photobleaching*. Biophys J, 2004. **86**(6): p. 3473-95.

187. Phair, R.D. and T. Misteli, *High mobility of proteins in the mammalian cell nucleus*. Nature, 2000. **404**(6778): p. 604-9.

188. Phair, R.D., et al., *Global nature of dynamic protein-chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins*. Mol Cell Biol, 2004. **24**(14): p. 6393-402.

189. Festenstein, R., et al., *Modulation of heterochromatin protein 1 dynamics in primary Mammalian cells*. Science, 2003. **299**(5607): p. 719-21.

190. Kimura, H. and P.R. Cook, *Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B*. J Cell Biol, 2001. **153**(7): p. 1341-53.

191. Gerlich, D., et al., *Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication*. Curr Biol, 2006. **16**(15): p. 1571-8.

192. Hemmerich, P., et al., *Dynamics of inner kinetochore assembly and maintenance in living cells*. J Cell Biol, 2008. **180**(6): p. 1101-14.

193. Wallrabe, H. and A. Periasamy, *Imaging protein molecules using FRET and FLIM microscopy*. Curr Opin Biotechnol, 2005. **16**(1): p. 19-27.

194. O'Keefe, R.T., S.C. Henderson, and D.L. Spector, *Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences*. J Cell Biol, 1992. **116**(5): p. 1095-110.

195. Leonhardt, H., et al., *Dynamics of DNA replication factories in living cells.* J Cell Biol, 2000. **149**(2): p. 271-80.

196. Schermelleh, L., et al., *Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation.* Nucleic Acids Res, 2007. **35**(13): p. 4301-12.

197. Easwaran, H.P., H. Leonhardt, and M.C. Cardoso, *Cell cycle markers for live cell analyses.* Cell Cycle, 2005. **4**(3): p. 453-5.

198. Meister, P., A. Taddei, and S.M. Gasser, *In and out of the replication factory.* Cell, 2006. **125**(7): p. 1233-5.

199. Sporbert, A., et al., *PCNA acts as a stationary loading platform for transiently interacting Okazaki fragment maturation proteins.* Nucleic Acids Res, 2005. **33**(11): p. 3521-8.

200. Dimitrova, D.S. and D.M. Gilbert, *The spatial position and replication timing of chromosomal domains are both established in early G1 phase.* Mol Cell, 1999. **4**(6): p. 983-93.

201. Jackson, D.A. and A. Pombo, *Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells.* J Cell Biol, 1998. **140**(6): p. 1285-95.

202. Ma, H., et al., *Spatial and temporal dynamics of DNA replication sites in mammalian cells.* J Cell Biol, 1998. **143**(6): p. 1415-25.

203. McNairn, A.J., et al., *Chinese hamster ORC subunits dynamically associate with chromatin throughout the cell-cycle.* Exp Cell Res, 2005. **308**(2): p. 345-56.

204. Phair, R.D. and T. Misteli, *Kinetic modelling approaches to in vivo imaging.* Nat Rev Mol Cell Biol, 2001. **2**(12): p. 898-907.

205. Kreitz, S., et al., *The human origin recognition complex protein 1 dissociates from chromatin during S phase in HeLa cells.* J Biol Chem, 2001. **276**(9): p. 6337-42.

206. Martini, E., et al., *Recruitment of phosphorylated chromatin assembly factor 1 to chromatin after UV irradiation of human cells.* J Cell Biol, 1998. **143**(3): p. 563-75.

207. Nielsen, A.L., et al., *Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family.* Embo J, 1999. **18**(22): p. 6385-95.

208. Langer, P.R., A.A. Waldrop, and D.C. Ward, *Enzymatic synthesis of biotin-labeled polynucleotides: novel nucleic acid affinity probes.* Proc Natl Acad Sci U S A, 1981. **78**(11): p. 6633-7.

209. Nakayasu, H. and R. Berezney, *Mapping replicational sites in the eucaryotic cell nucleus.* J Cell Biol, 1989. **108**(1): p. 1-11.

210. Dimitrova, D.S. and R. Berezney, *The spatio-temporal organization of DNA replication sites is identical in primary, immortalized and transformed mammalian cells*. J Cell Sci, 2002. **115**(Pt 21): p. 4037-51.

211. Ferrari, A., et al., *ROCK-mediated contractility, tight junctions and channels contribute to the conversion of a preapical patch into apical surface during isochoric lumen initiation*. J Cell Sci, 2008. **121**(Pt 21): p. 3649-63.

212. Le Clainche, C., et al., *IQGAP1 stimulates actin assembly through the N-WASP-Arp2/3 pathway*. J Biol Chem, 2007. **282**(1): p. 426-35.

213. Tannock, I.F., *Population kinetics of carcinoma cells, capillary endothelial cells, and fibroblasts in a transplanted mouse mammary tumor*. Cancer Res, 1970. **30**(10): p. 2470-6.

214. Beaudouin, J., et al., *Dissecting the contribution of diffusion and interactions to the mobility of nuclear proteins*. Biophys J, 2006. **90**(6): p. 1878-94.

215. Essers, J., et al., *Nuclear dynamics of PCNA in DNA replication and repair*. Mol Cell Biol, 2005. **25**(21): p. 9350-9.

216. Cheutin, T., et al., *Maintenance of stable heterochromatin domains by dynamic HP1 binding*. Science, 2003. **299**(5607): p. 721-5.

217. Muller, K.P., et al., *Multiscale analysis of dynamics and interactions of heterochromatin protein 1 by fluorescence fluctuation microscopy*. Biophys J, 2009. **97**(11): p. 2876-85.

218. Howell, J.L. and R. Truant, *Live-cell nucleocytoplasmic protein shuttle assay utilizing laser confocal microscopy and FRAP*. Biotechniques, 2002. **32**(1): p. 80-2, 84, 86-7.

219. Tagawa, A., et al., *Assembly and trafficking of caveolar domains in the cell: caveolae as stable, cargo-triggered, vesicular transporters*. J Cell Biol, 2005. **170**(5): p. 769-79.

220. Picard, D., E. Suslova, and P.A. Briand, *2-color photobleaching experiments reveal distinct intracellular dynamics of two components of the Hsp90 complex*. Exp Cell Res, 2006. **312**(19): p. 3949-58.

221. Xouri, G., et al., *Cdt1 associates dynamically with chromatin throughout G1 and recruits Geminin onto chromatin*. Embo J, 2007. **26**(5): p. 1303-14.

222. Carrero, G., et al., *Characterizing fluorescence recovery curves for nuclear proteins undergoing binding events*. Bull Math Biol, 2004. **66**(6): p. 1515-45.

223. Jiang, W., N.J. Wells, and T. Hunter, *Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6*. Proc Natl Acad Sci U S A, 1999. **96**(11): p. 6193-8.

224. Kim, J. and E.T. Kipreos, *Control of the Cdc6 replication licensing factor in metazoa: the role of nuclear export and the CUL4 ubiquitin ligase*. Cell Cycle, 2008. **7**(2): p. 146-50.

225. Paolinelli, R., et al., *Acetylation by GCN5 regulates CDC6 phosphorylation in the S phase of the cell cycle*. Nat Struct Mol Biol, 2009. **16**(4): p. 412-20.

226. Tsakraklides, V. and S.P. Bell, *Dynamics of Pre-replicative Complex Assembly*. J Biol Chem, 2010. **285**(13): p. 9437-43.

227. Hayakawa, T., et al., *Cell cycle behavior of human HP1 subtypes: distinct molecular domains of HP1 are required for their centromeric localization during interphase and metaphase*. J Cell Sci, 2003. **116**(Pt 16): p. 3327-38.

228. Auth, T., E. Kunkel, and F. Grummt, *Interaction between HP1alpha and replication proteins in mammalian cells*. Exp Cell Res, 2006. **312**(17): p. 3349-59.

229. Craig, J.M., et al., *Analysis of mammalian proteins involved in chromatin modification reveals new metaphase centromeric proteins and distinct chromosomal distribution patterns*. Hum Mol Genet, 2003. **12**(23): p. 3109-21.

230. Radichev, I., et al., *Genetic analysis of human Orc2 reveals specific domains that are required in vivo for assembly and nuclear localization of the origin recognition complex*. J Biol Chem, 2006. **281**(32): p. 23264-73.

231. Kim, J., H. Feng, and E.T. Kipreos, *C. elegans CUL-4 prevents rereplication by promoting the nuclear export of CDC-6 via a CKI-1-dependent pathway*. Curr Biol, 2007. **17**(11): p. 966-72.

232. Fujita, M., et al., *Nuclear organization of DNA replication initiation proteins in mammalian cells*. J Biol Chem, 2002. **277**(12): p. 10354-61.

233. Alexandrow, M.G. and J.L. Hamlin, *Cdc6 chromatin affinity is unaffected by serine-54 phosphorylation, S-phase progression, and overexpression of cyclin A*. Mol Cell Biol, 2004. **24**(4): p. 1614-27.

234. Lidonnici, M.R., et al., *Subnuclear distribution of the largest subunit of the human origin recognition complex during the cell cycle*. J Cell Sci, 2004. **117**(Pt 22): p. 5221-31.

235. McNairn, A.J. and D.M. Gilbert, *Overexpression of ORC subunits and increased ORC-chromatin association in transformed mammalian cells*. J Cell Biochem, 2005. **96**(5): p. 879-87.

236. Young, M.R. and B.K. Tye, *Mcm2 and Mcm3 are constitutive nuclear proteins that exhibit distinct isoforms and bind chromatin during specific cell cycle stages of Saccharomyces cerevisiae*. Mol Biol Cell, 1997. **8**(8): p. 1587-601.

237. Albertazzi, L., et al., *Quantitative FRET Analysis With the EGFP-mCherry Fluorescent Protein Pair*. Photochem Photobiol, 2009. **85**: p. 287-97.

238. Arosio, D., et al., *Spectroscopic and structural study of proton and halide ion cooperative binding to gfp*. Biophys J, 2007. **93**(1): p. 232-44.

239. Byrd, K. and V.G. Corces, *Visualization of chromatin domains created by the gypsy insulator of Drosophila*. J Cell Biol, 2003. **162**(4): p. 565-74.

240. Hakes, D.J. and R. Berezney, *DNA binding properties of the nuclear matrix and individual nuclear matrix proteins. Evidence for salt-resistant DNA binding sites*. J Biol Chem, 1991. **266**(17): p. 11131-40.

241. Mancini, M.G., et al., *Subnuclear partitioning and functional regulation of the Pit-1 transcription factor*. J Cell Biochem, 1999. **72**(3): p. 322-38.

242. He, D.C., J.A. Nickerson, and S. Penman, *Core filaments of the nuclear matrix*. J Cell Biol, 1990. **110**(3): p. 569-80.

243. Gilbert, D.M., *Replication timing and transcriptional control: beyond cause and effect*. Curr Opin Cell Biol, 2002. **14**(3): p. 377-83.

244. Audit, B., et al., *Open chromatin encoded in DNA sequence is the signature of 'master' replication origins in human cells*. Nucleic Acids Res, 2009. **37**(18): p. 6064-75.

245. Sequeira-Mendes, J., et al., *Transcription initiation activity sets replication origin efficiency in mammalian cells*. PLoS Genet, 2009. **5**(4): p. e1000446.

246. Malyavantham, K.S., et al., *Spatio-temporal dynamics of replication and transcription sites in the mammalian cell nucleus*. Chromosoma, 2008. **117**(6): p. 553-67.

247. Bondos, S.E., X.X. Tan, and K.S. Matthews, *Physical and genetic interactions link hox function with diverse transcription factors and cell signaling proteins*. Mol Cell Proteomics, 2006. **5**(5): p. 824-34.

248. Shen, W.F., et al., *HOXA9 forms triple complexes with PBX2 and MEIS1 in myeloid cells*. Mol Cell Biol, 1999. **19**(4): p. 3051-61.

249. Williams, T.M., M.E. Williams, and J.W. Innis, *Range of HOX/TALE superclass associations and protein domain requirements for HOXA13:MEIS interaction*. Dev Biol, 2005. **277**(2): p. 457-71.

250. Kong, D. and M.L. DePamphilis, *Site-specific DNA binding of the Schizosaccharomyces pombe origin recognition complex is determined by the Orc4 subunit*. Mol Cell Biol, 2001. **21**(23): p. 8095-103.

251. Sabo, A., et al., *Acetylation of conserved lysines in the catalytic core of cyclin-dependent kinase 9 inhibits kinase activity and regulates transcription*. Mol Cell Biol, 2008. **28**(7): p. 2201-12.

252. Bancaud, A., et al., *Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin*. Embo J, 2009. **28**(24): p. 3785-98.

253. Bulinski, J.C., et al., *Rapid dynamics of the microtubule binding of ensconsin in vivo*. J Cell Sci, 2001. **114**(Pt 21): p. 3885-97.

254. Dundr, M., et al., *A kinetic framework for a mammalian RNA polymerase in vivo*. Science, 2002. **298**(5598): p. 1623-6.

255. Rabut, G., V. Doye, and J. Ellenberg, *Mapping the dynamic organization of the nuclear pore complex inside single living cells*. Nat Cell Biol, 2004. **6**(11): p. 1114-21.

256. Chubb, J.R., et al., *Chromatin motion is constrained by association with nuclear compartments in human cells*. Curr Biol, 2002. **12**(6): p. 439-45.

257. Abney, J.R., et al., *Chromatin dynamics in interphase nuclei and its implications for nuclear structure*. J Cell Biol, 1997. **137**(7): p. 1459-68.

258. Gerlich, D., J. Mattes, and R. Eils, *Quantitative motion analysis and visualization of cellular structures*. Methods, 2003. **29**(1): p. 3-13.

259. Lakowicz, J., *Principles of Fluorescence Spectroscopy*. Third Edition ed. 2006, New York, NY 10013, USA: Springer Science+Business Media, LLC.

260. Sekar, R.B. and A. Periasamy, *Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations*. J Cell Biol, 2003. **160**(5): p. 629-33.

261. Day, R.N., A. Periasamy, and F. Schaufele, *Fluorescence resonance energy transfer microscopy of localized protein interactions in the living cell nucleus*. Methods, 2001. **25**(1): p. 4-18.

262. Elangovan, M., R.N. Day, and A. Periasamy, *Nanosecond fluorescence resonance energy transfer-fluorescence lifetime imaging microscopy to localize the protein interactions in a single living cell*. J Microsc, 2002. **205**(Pt 1): p. 3-14.

263. Heim, R. and R.Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer*. Curr Biol, 1996. **6**(2): p. 178-82.

264. Pollok, B.A. and R. Heim, *Using GFP in FRET-based applications*. Trends Cell Biol, 1999. **9**(2): p. 57-60.

265. Förster, T., *Energiewanderung und Fluoreszenz*, in *Naturwissenschaften*, 1946. **33**: p. 166-175.

266. Förster, T., *Comparative Effects of Radiation*. Radiat. Res. Supp., 1960. **2**.

267. Tramier, M., et al., *Sensitivity of CFP/YFP and GFP/mCherry pairs to donor photobleaching on FRET determination by fluorescence lifetime imaging microscopy in living cells*. Microsc Res Tech, 2006. **69**(11): p. 933-9.

268. Millington, M., et al., *High-precision FLIM-FRET in fixed and living cells reveals heterogeneity in a simple CFP-YFP fusion protein*. Biophys Chem, 2007. **127**(3): p. 155-64.