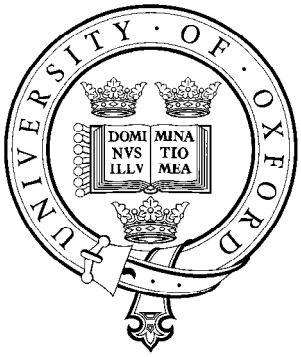
Specialised Transcription Factories



Meng Xu

Lincoln College

The Sir William Dunn School of Pathology

Medical Sciences Division

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas 2007

Abstract

Specialised Transcription Factories

Meng Xu

Lincoln College

Doctor of Philosophy in Medical Sciences

Michaelmas 2007

The intimate relationship between the higher-order chromatin organisation and the regulation of gene expression is increasingly attracting attention in the scientific community. Thanks to high-resolution microscopy, genome-wide molecular biology tools (3C, ChIP-on-chip), and bioinformatics, detailed structures of chromatin loops, territories, and nuclear domains are gradually emerging. However, to fully reveal a comprehensive map of nuclear organisation, some fundamental questions remain to be answered in order to fit all the pieces of the jigsaw together. The underlying mechanisms, precisely organising the interaction of the different parts of chromatin need to be understood.

Previous work in our lab hypothesised and verified the "transcription factory" model for the organisation of mammalian genomes. It is widely assumed that active polymerases track along their templates as they make RNA. However, after allowing engaged polymerases to extend their transcripts in tagged precursors (e.g., Br-U or Br-UTP), and immunolabelling the now-tagged nascent RNA, active transcription units are found to be clustered in nuclei, in small and numerous sites we call "transcription factories". Previous work suggested the transcription machinery acts

both as an enzyme as well as a molecular tie that maintains chromatin loops, and the different classes of polymerases are concentrated in their own dedicated factories. This thesis aims to further characterise transcription factories.

Different genes are transcribed by different classes of RNA polymerase (i.e., I, II, or III), and the resulting transcripts are processed differently (e.g., some are capped, others spliced). Do factories specialise in transcribing particular subsets of genes?

This thesis developed a method using replicating minichromosomes as probes to examine whether transcription occurs in factories, and whether factories specialise in transcribing particular sets of genes. Plasmids encoding the SV40 origin of replication are transfected into COS-7 cells, where they are assembled into minichromosomes. Using RNA fluorescence *in situ* hybridisation (FISH), sites where minichromosomes are transcribed are visualised as discrete foci, which specialise in transcribing different groups of genes. Polymerases I, II, and III units have their own dedicated factories, and different polymerase II promoters and the presence of an intron determine the nuclear location of transcription. Using chromosome conformation capture (3C), minichromosomes with similar promoters are found in close proximity. They are also found close to similar endogenous promoters and so are likely to share factories with them.

In the second part of this thesis, I used RNA FISH to confirm results obtained by tiling microarrays. Addition of tumour necrosis factor alpha (TNF α) to human umbilical vein endothelial cells induces an inflammatory response and the transcription of a selected sub-set of genes. My collaborators used tiling arrays to demonstrate a wave of transcription that swept along selected long genes on stimulation. RNA FISH confirmed these results, and that long introns are co-

transcriptionally spliced. Results are consistent with one polymerase being engaged on an allele at any time, and with a major checkpoint that regulates polymerase escape from the first few thousand nucleotides into the long gene.

Acknowledgements

It is my great pleasure to thank Professor Peter R. Cook for giving me the opportunity to study in the Sir William Dunn School of Pathology. Without his inspiration and constructive criticism, none of my work could be possible. I thank all members of the Cook group, especially Dr. S. Melnik, Dr. C. Eskiw, Dr. A. Rapp, and Dr. M. Shamsher for stimulating discussions and advice at all stages, and Dr. D. Carter, Dr. M. Formaget, Mr. J. Bartlett and Mr. K. Finan for their help.

I am grateful to Dr. S. Murphy, Dr. A. Akoulitchev, and Professor N. Proudfoot for their valuable suggestions; Dr. A. Ramadass, Dr. M. Dye, Dr. E. Fodor, Dr. T. Kollakowski and Dr. J. Harraway for reagents; and Dr. N. White and Mr. P. Yajnik for help on microscopy.

I thank Professor Tatsuhiko Kodama and his group at the University of Tokyo for an excellent collaboration; and Dr. V. Buckle, Dr. J. Brown and Ms. J. Green at the Weatherall Institute of Molecular Medicine for their help on FISH experiments.

I wish to thank Dr. T. Thesen, Miss K. Tanzhaus, Miss L. Harriss, Mr. E. Harvat, Dr. T. Deng, and Miss S. Olego-Fernandez for their friendship, and my family for their complete support without reservation.

The Clarendon Bursary, the UK Government Overseas Research Student Award Scheme, and the E.P. Abraham Research Fund awarded me scholarships for three years to study at the University of Oxford.

Contents

List of Tables

List of Figures

Glossary

3C	chromosome conformation capture
BAC	Bacterial Artificial Chromosomes
CAB box	Cajal body box
CBP	cap binding protein
CF	cleavage factor
CBC	cap-binding complex
CPSF	cleavage and polyadenylation specificity factor
CT	chromosome territories
CTCF	CCCTC-binding factor
CTD	carboxy-terminal domain
Ch	chapter
ChIP	chromatin immunoprecipitation
Chr	chromosome
CstF	cleavage stimulatory factor
DAPI	4',6-diamidino-2-phenylindole dihydrochloride
DEPC	diethyl pyrocarbonate
DFC	dense fibrillar component
DMSO	Dimethyl Sulfoxide
DRB	5,6-dichlorobenzimidazole 1-β-D-ribofuranoside
DSE	distal sequence element
DTT	Dithiothreitol
EDTA	ethylenediamine-tetraacetic acid
EJC	exon junction complex
EM	electron microscopy
FACT	facilitates chromatin transcription
FACS	Fluorescence-activated cell sorting
FC	fibrillar centre
FISH	fluorescence in situ hybridisation
FITC	fluorescein isothiocyanate

GC	granular component
HAT	histone-acetyltransferase
HDE	histone downstream element
HEXIM1	hexamethylene bisacetamide-induced protein 1
HP	heterochromatin protein
HUVEC	human vascular endothelial cell
ICD	interchromatin domain
ICR	imprinting control region
IGCs	interchromatin granule clusters
LCR	locus control region
LINE	long interspersed nucleotide elements
MAR	matrix-attachment region
miRNA	microRNA
MHC	major histocompatibility
mt	mitochondrial
NMD	nonsense-mediated decay
NOR	nucleolar organiser regions
NPC	nuclear pore complex
NTS	nascent transcript and splicing
n-TEF	negative transcription elongation factor
OPT	Oct1/PTF/transcription
OR	olfactory receptor
ORF	open reading frame
PAP	poly(A) polymerase
PFA	paraformaldehyde
PHAX	phosphorylated adaptor for RNA export
PIC	preinitiation complex
PML	promyelocytic leukaemia
PTC	premature termination codon
PNC	perinucleolar compartment
pol	polymerase
poly(A)	polyadenine
p-TEF	positive transcription elongation factor

PTF	PSE-binding transcription factor
PSE	proximal sequence element
qPCR	quantitative PCR
qRT-PCR	quantitative reverse transcription PCR
RISC	RNA-induced silencing complexes
rRNA	ribosomal RNA
SAR	scaffold-attachment region
scaRNA	small Cajal body RNA
SINE	short interspersed element
siRNA	small interfering RNA
SLBP	stem-loop binding protein
SMN	survival of motor neurons
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoproteins
SRB	suppressor of RNA polymerase B
TBP	TATA-box binding protein
TF	transcription factor
TMG	trimethylguanosine
TNF α	tumour necrosis factor α
tRNA	transfer RNA
UBF	upstream binding factor
UCE	upstream control element
UTR	untranslated region
Xi	inactive X chromosome
Xic	X-inactivation centre
Xa	active X chromosome

1 Introduction

Ever since the completion of the human genome project in 2003, we have entered the post-genomic era. The enormous information revealed by genome-wide transcriptomic, proteomic, and epigenetic research combined with progressively powerful bioinformatics is pushing genetics up to a new peak. In contrast to the information explosion, we still know little about the mechanisms underlying genome organisation and how the organisation influences vital biological functions like replication and transcription.

Recently, some details of the relationship between nuclear structure and function have begun to emerge, and the subject is becoming increasingly popular. On one hand, gene replication and expression are heavily regulated by chromatin structure and the associated epigenetic modifications; on the other hand, the transcription and replication machineries seem to drive the self organisation of the genome within the nucleus.

Here I first review the different machineries involved in transcription, and the relation between transcription and nuclear structure. Next, I introduce a model of nuclear organisation in which transcription is concentrated in discrete 'factories' that contain the appropriate machinery. This leads to my project – investigating whether factories specialise in transcribing different gene subsets.

1.1 Different machineries involved in transcription

Genetic information flows from DNA into RNA, and then into the proteins that perform the biological functions. The process by which a DNA sequence is copied by an RNA polymerase to produce a complementary RNA is called transcription. In eukaryotes, transcription mostly occurs in the nucleus, where the DNA is primarily located.

The core of the enormous polymerising machine containing many proteins can be purified as an active complex known as RNA polymerase (pol). Three different pols, namely pol I, II, and III have been isolated from eukaryotes. They are structurally similar to each other, although involved in transcribing different types of genes. Pol II transcribes all protein-coding genes, histone genes, some small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and some other non-coding RNAs; pol I transcribes ribosomal RNA (rRNA); pol III transcribes transfer RNA (tRNA), some snRNAs and various other small non-coding regulatory RNAs.

Transcription involves three basic steps: initiation (when a DNA template and pol come together), elongation (as RNA is polymerised using ribonucleotide triphosphates as substrates), and termination (when the pol and template separate). The traditional model for transcription involves a pol that initiates by attaching to a template before it elongates by tracking along the DNA; on termination, it dissociates from the template. An alternative model consistent with all my results involves the template initiating by attaching to a pol immobilised in a factory, followed by elongation as the pol reels in the template and extrudes nascent RNA; on termination, the template dissociates from the pol and the factory (Cook, 2001).

During or after transcription, transcripts are processed differently (some are capped, spliced, and/or 3' edited). Protein-coding transcripts are processed into mature messenger RNAs (mRNAs), and are exported to the cytoplasm to be translated into protein. Non-coding RNAs are further edited to participate in transcription regulation or other functional events.

A dedicated pol is involved in the transcription that occurs in mitochondria in the cytoplasm. Recently, a plant specific nuclear pol – pol IV – was discovered. It is required for the production of small interfering RNAs (siRNAs) that direct RNA-mediated DNA methylation of repeated genomic sequences; this blocks transcription of homologous promoters (Kanno et al., 2005). Pol IV function is required to maintain chromatin organisation, as mutation of the catalytic subunits causes heterochromatin dispersal (Pontes et al., 2006).

Here I review the genes transcribed by different classes of polymerases and their different processing mechanisms.

1.1.1 RNA polymerase II and the genes it transcribes

Pol II is the predominant nuclear polymerase. It is responsible for transcribing protein-coding genes and some small regulatory RNAs. It is often present as a large complex known as the "holoenzyme" (Koleske and Young, 1994; Struhl, 1996; Kadonaga, 1998). The holoenzyme is a huge protein conglomerate containing the core pol II, a subset of general transcription factors (TFs), nine suppressor of RNA polymerase B (SRB) proteins, and other specific transcription factors (such as GATA-1, Sp1) (Bjorklund and Kim, 1996; Kadonaga, 1998). The crystal structure of the yeast core pol II structure, in the acts of initiation and

elongation, has been well characterised at a resolution of 2.8 - 3.3 Å (Cramer et al., 2001; Gnatt et al., 2001; Armache et al., 2003). It contains 12 different subunits, namely Rpb1 - Rpb12, and a total mass of 513.6 kD. The human pol II has a similar composition.

1.1.1.1 Initiation by polymerase II is complex

In contrast to prokaryotes, initiation in eukaryotes is a complex event that requires cooperative action by transcription factors (Roeder, 1996b). At the transcription start site – or promoter – specific DNA sequences – *cis*-control elements – determine accurate initiation; these include the TATA box, CCAAT box, GC box, heat-shock element, etc. (Pelham, 1982; Kadonaga, 2004). Different subsets of genes have their own distinctive promoters, and biochemical studies have given us a picture of how the initiation complex is assembled at the promoter. [It was thought that most promoters of protein-coding genes contained a TATA box located 25 nucleotides (nt) 5' from the starting site, and the assembly pathway was analysed using TATA-containing templates. However, it now turns out that only a minority of human genes contain such TATA boxes (e.g., Jin *et al*., 2006).] The TATA box is recognised by TFIID, which contains the TATA-box binding protein (TBP). TFIIB binds to the TFIID-promoter complex, before the TFIIF/core pol II is recruited. TFIIE and TFIIH then join to complete the preinitiation complex (PIC). TFIIH contains a DNA helicase that can unwind the DNA duplex. The CDK7 kinase of TFIIH phosphorylates the carboxy-terminal domain (CTD) of the largest subunit of the core pol II (Pabo and Sauer, 1992; Orphanides et al., 1996; Roeder, 1996b; Roeder, 1996a). The CTD forms a mobile tail consisting of 52 repeats of a consensus

sequence Y1-S2-P3-T4-S5-P6-S7, which can be phosphorylated at Ser2 and Ser5 (Cramer et al., 2001).

Unlike prokaryotes, eukaryotic transcription also needs chromatin-remodelling factors to mobilise nucleosomes and deal with higher-order chromatin structures, as well as a range of enzymes that modify histones and associated proteins (Kadonaga, 2004).

1.1.1.2 Elongation is tightly coupled with processing

Soon after initiation, pol II associates with negative transcription elongation factors (N-TEFs) to terminate prematurely and generate only a short transcript; this process can be repeated many times to yield many short transcripts (Marshall and Price, 1992). Positive transcription elongation factor (P-TEFb) is required for escape from this abortive phase into productive elongation (Marshall and Price, 1995).

During elongation, pol II associates with various elongation factors, including TFIIF, elongin, ELL, ELL2, DSIF (DRB sensitivity-inducing factor), SII and FACT (facilitates chromatin transcription) (Conaway et al., 2000). [In the alternative model for transcription (above), pol II elongates by reeling in the template and unwinding the DNA to expose a new base to an incoming triphosphate at the polymerisation site.] The two strands are unwound and rewound during elongation, and this creates superhelical tension in DNA which is rapidly removed by DNA topoisomerase (Alberts, 2002). Optical tweezers have been used to analyse elongation by *Escherichia coli* pol; it occurs in discrete steps averaging 3.7 ± 0.6 Å, equivalent to one base at a time (Abbondanzieri et al., 2005). However, we await a similar visualisation using the eukaryotic polymerase.

The speed of elongation by pol II has been determined (using fluorescence *in situ* hybridisation; FISH) as 1.1-1.4 kb/min on the *Drosophila UBX* and mammalian β-actin genes (Darzacq et al., 2005). A recent study using photobleaching and photoactivation of fluorescent MS2 proteins to label nascent RNA in living cells reported an elongation speed of 4.3 kb/min (Darzacq et al., 2007).

Primary transcripts produced by pol II are processed in a number of ways (Nevins, 1983; Kiss, 2004); processing is tightly coupled with transcription, and the CTD of pol II plays a critical role in binding the needed factors (Thompson et al., 1993; Steinmetz, 1997).

1.1.1.3 5’ capping of mRNA

The 5’ end of protein-coding transcripts is capped shortly after initiation, when the nascent chain is 20 - 40 nt long (Hirose and Manley, 2000). Capping occurs in three steps: 1) RNA 5’ triphosphatase removes the γ-phosphate of the 5’ (first) nucleotide; 2) guanylyltransferase transfers a GMP to the resulting diphosphate end; 3) the N7 position of the cap guanine is methylated, forming the m7G(5')ppp(5') cap (Shuman, 1995).

During capping, the capping enzyme directly and selectively binds to the phosphorylated form of the pol II CTD, at either Ser2 or Ser5 (Yue et al., 1997). This prevents degradation of the transcript (Hirose and Manley, 2000).

A cap-binding complex (CBC) recognises the cap on the transcript, and will eventually direct the transcript to the correct export pathway to the cytoplasm. The CBP is then replaced by cytoplasmic translation initiation factor eIF-4E (Shatkin and Manley, 2000). The 5’ cap – which is introduced co-transcriptionally – is then a

distinctive signal of pol II specific RNAs. Pol I and III transcripts either lack caps, or capping is post transcriptional (Table 1.1).

1.1.1.4 3’ polyadenylation of mRNA and termination

In higher eukaryotes, most protein-coding genes contain a poly(A) tail (the replication-dependent histone genes are an exception). Around 200 adenosine residues are added to the 3’–OH by the poly(A) machinery (Zhao et al., 1999). The 3’–OH is generated by a specific cleavage at a CA dinucleotide, and is marked with an AAUAAA sequence 10-30 nt 5’ to the cleavage site, a U-rich or GU-rich downstream element 10-30 nt 3’ to the cleavage site, and in some cases a U-rich upstream enhancer (Colgan and Manley, 1997; Zhao et al., 1999; Proudfoot et al., 2002).

The mammalian poly(A) machinery consists of the cleavage and polyadenylation specificity factor (CPSF) which binds to AAUAAA; a cleavage stimulatory factor (CstF) which binds to the downstream element; cleavage factor I and II (CF I and CF II) which direct cleavage; a poly(A) polymerase (PAP) which promotes cleavage and adds the A residues one by one; and a poly(A) binding protein PABP II which binds to the new poly(A) tail and enhances the activity of PAP and so controls of length of the tail (Wahle and Ruegsegger, 1999; Zhao et al., 1999; Proudfoot, 2004).

Evidence suggests cleavage is cotranscriptional, and the pol II CTD plays an important part (McCracken et al., 1997b; Hirose and Manley, 2000). The CPSF and CstF copurify with the CTD (McCracken et al., 1997b), and processing of full-length synthetic pre-mRNAs injected into oocytes is less efficient than co-transcriptional

processing *in vivo* (Bird et al., 2004). Inhibition of CTD phosphorylation by a kinase inhibitor also blocks cleavage (Ahn et al., 2004; Bird et al., 2004).

3'end processing may be a trigger of transcription termination (Greger and Proudfoot, 1998). The loss of the poly(A) tail results in failure to terminate transcription, which causes read-through and the silencing of the downstream genes through transcriptional interference (Greger et al., 2000). A study using microinjected plasmid DNA encoding the TFIIIA gene into *Xenopus* oocyte nuclei indicated the poly(A) signal is required for termination, but not for cotranscriptional cleavage (Osheim et al., 1999). An *in vitro* study also showed the termination site is further downstream from the poly(A) site (Tran et al., 2001). A transcription pausing site downstream of the poly(A) site has also been identified; this seems to enhance recognition of the poly(A) site (Birse et al., 1998; Yonaha and Proudfoot, 2000). The currently prevalent model suggests nascent RNA is looped out between the 3' end and the associated processing machinery (which in turn associates with the CTD of pol II) and the pol II RNA exit channel (Cramer et al., 2001; Proudfoot, 2004).

1.1.1.5 Splicing of mRNA

Most protein-coding genes contain non-coding introns which have to be removed in order to generate a functional message. Pre-mRNAs contain consensus elements marking the introns: the 5' splice site is marked by AG|GURAGU ('|': exon-intron junction), the 3' by YAG|RNNN (Proudfoot et al., 2002).

The complex process of splicing is performed by spliceosomes containing small nuclear ribonucleoproteins (snRNPs) and associated proteins with five snRNAs (U1, U2, U4, U5, and U6) (Kramer, 1996). First, the 2'OH of the branchpoint adenosine attacks the 5' exon-intron border, forming a lariat (the intron

and the 3' exon); the 5' exon becomes free. The two exon-intron junctions are brought together by snRNPs. Second, the 3'OH of the free 5' exon attacks the intron-3'exon border. The two exons fuse together when the lariat is released (Moore and Sharp, 1993; Sharp, 1994; Proudfoot et al., 2002). Evidence suggests snRNAs play a leading catalytic role in the spliceosome, and form intermolecular base pairs with the transcript (Newman, 1997; Valadkhan and Manley, 2000); the protein complex interact with the catalytic core, and is responsible for the assembly and stabilisation of the active spliceosome (Collins and Guthrie, 2000).

Electron microscopy indicates splicing can occur cotranscriptionally (Beyer and Osheim, 1988; Wetterberg et al., 2001). The CTD again plays an important part, as certain mutations specifically affect splicing, but not transcription or other processing steps (Hirose and Manley, 2000; Maniatis and Reed, 2002). An *in vitro* assay indicated pre-mRNA synthesised by pol II is immediately directed into the spliceosome assembly pathway, while transcripts synthesised by T7 polymerase are not (Das et al., 2007).

Genome-wide analyses indicate that 40-60% of human genes have alternative splice forms (Modrek and Lee, 2002). A functional unit may be added or removed in the resulting alternative protein product (Modrek et al., 2001). The rate of elongation can influence the choice between alternative splicing sites (de la Mata et al., 2003).

1.1.1.6 Interaction of capping, splicing, and 3' end processing

5' capping, splicing, and 3' end processing are clearly interconnected with each other (Maniatis and Reed, 2002; Neugebauer, 2002; Proudfoot, 2004). The CBC enhances the interaction of U1 snRNA with the cap-proximal 5' splice site

(Lewis and Izaurralde, 1997). Splicing factors at the 3' intron promote transcript 3' end cleavage (McCracken et al., 2002), and the poly(A) machinery promotes recognition of the 3' terminal intron (Berget, 1995).

The pol II CTD regulates the coupling of transcription and processing, through changes in pattern of phosphorylation. Capping, splicing, and polyadenylation are all inhibited by truncation of the CTD (Cho et al., 1997; McCracken et al., 1997a; McCracken et al., 1997b). Ser5 is first phosphorylated during initiation, which recruits the capping machinery, and then the capping enzymes are released after Ser5 is dephosphorylated (Ho and Shuman, 1999; Komarnitsky et al., 2000). Ser2 is subsequently phosphorylated, recruiting the splicing and 3' end processing factors (Komarnitsky et al., 2000; Cho et al., 2001).

One model suggests pol II carries all the processing machinery along with it as it transcribes (McCracken et al., 1997b), but this seems unlikely (Wetterberg et al., 2001). Another model involves an immobilised 'factory' that contains all the needed factors to couple transcription and processing (Iborra et al., 1996; Cook, 1999) (Fig.1.4). Another study using electron tomography of the active BR3 gene described a NTS complex (nascent transcript and splicing complex) containing 20–25 nascent transcripts, pol II, and splicing factors. As the gene is transcribed, the size and shape of this complex changed greatly, indicating a dynamic exchange with the soluble pool (Wetterberg et al., 2001).

1.1.1.7 Replication-dependent histone RNAs

Some histone proteins (i.e. H3.3, H2A.Z, H1, H3-cid and MacroH2A) are encoded by poly(A) mRNAs and metabolised like other mRNAs (Marzluff, 2005). However, some histone mRNAs – the majority of the replication-dependent ones –

lack poly(A) tails (Marzluff et al., 2002) or introns. Intron insertion interferes with proper 3' end formation (Pandey et al., 1990). There are five classes of replication-dependent histones. The four core histones, H2A, H2B, H3, and H4 make up the nucleosome, and histone H1 binds to the linker DNA between nucleosomes (Marzluff et al., 2002).

The 3' end of the mammalian histone mRNA possesses a highly conserved 26-nt sequence, 24-70 nt 3' of the stop codon (Marzluff et al., 2002). The untranslated region (UTR) of histone RNA forms a 16 nt stem-loop, interacting with the stem-loop binding protein (SLBP) (Dominski and Marzluff, 1999). The mature 3' end is obtained by a single cleavage event downstream of the stem-loop. 9 to 12 nt 3' of the cleavage site is the histone downstream element (HDE), the core (AAGAG) of which is complementary to the 5' end of U7 snRNA. The U7 snRNP complex is essential for histone pre-mRNA 3' processing (Marzluff, 2005). A unique Sm binding site follows the HDE site, essential for U7 snRNP function (Grimm et al., 1993).

The histone mRNAs are highly cell-cycle regulated, expressed primarily in S-phase of somatic cells (Marzluff et al., 2002). Pre-mRNA processing plays an important role in this regulation. The major regulator SLBP is synthesized just before S-phase and degraded rapidly at the end of S-phase (Zheng et al., 2003).

1.1.1.8 U snRNAs transcribed by polymerase II

snRNAs, normally 60-450 nt long, are assembled into snRNPs with a set of RNP proteins (Kiss, 2004). The major U1, U2, U4, and U5 snRNAs are required for splicing (Lamond and Spector, 2003). Although also transcribed by pol II, they have different transcription initiation and processing mechanisms.

The promoters of snRNAs do not contain a TATA box; instead, they are composed of a basal proximal sequence element (PSE), and an upstream enhancer-like distal sequence element (DSE) (Hernandez, 2001).

During transcription, the 5' end of the snRNA precursor is capped with a 7mGppp (Mattaj, 1986; Kiss, 2004). snRNAs do not have any introns or poly(A) tails, and the 3' end signal is a conserved sequence called the "3' box" (Uguen and Murphy, 2003). Transcription continues for at least 250 nt beyond the 3' box in U2, and the 3' box only functions as a termination signal when pol II initiates at the PSE (Cuello et al., 1999). Efficient processing of snRNA requires the pol II CTD, as truncation or treatment with CTD kinase inhibitors causes a dramatic reduction in proper 3' end formation of U2 transcripts (Medlin et al., 2003).

After export to the cytoplasm via a large export complex, the pre-snRNA is assembled with a ring of seven Sm proteins into the snRNP core complex (Paushkin et al., 2002). The 7mG cap of pre-snRNA is then hypermethylated. The whole Sm snRNP is reimported into the nucleus via importin β and other factors recognising Sm proteins and the cap (Palacios et al., 1997).

After reimport into the nucleus, snRNPs accumulate in Cajal bodies, where they are site-specifically pseudouridylated and 2'-O-methylated to create mature spliceosomal snRNPs. This modification is mediated by small Cajal body RNAs (scaRNAs) (Gall, 2000; Jady et al., 2003; Nesic et al., 2004). Finally mature spliceosomal snRNPs are stored in splicing speckles (Lamond and Spector, 2003).

1.1.1.9 scaRNA and snoRNA

Most small nucleolar RNAs (snoRNAs) direct the site-specific 2'-O-methylation and pseudouridylation of rRNAs and U6 snRNAs, while small Cajal

body RNAs (scaRNAs) direct similar modifications of pol II transcribed U snRNAs (Darzacq et al., 2002; Kiss, 2002). Most snoRNAs and scaRNAs are processed from spliced introns by exonucleolytic trimming. A common feature of the host gene transcripts is a 5' terminal oligopyrimidine sequence (Hirose et al., 2003), with the only exceptions being U3, U8 and U13 box C/D snoRNAs which are transcribed from independent genes (Samarsky et al., 1998). 3' end trimming and synthesis of the trimethylguanosine (TMG) cap of the U3 box C/D snoRNA may occur in Cajal body (Hirose et al., 2003).

snoRNAs and scaRNA are targeted to the nucleolus by their conserved C/D and H/ACA box elements (Terns and Terns, 2002). The scaRNAs have a CAB box which targets them to the final destination – the Cajal Body (Richard et al., 2003).

1.1.1.10 LINE elements

LINE-1 (long interspersed nucleotide elements–1) elements in mammals are active retrotransposable elements. They transpose in the genome by amplifying themselves via retrotransposition. There are 900,000 copies in the human genome – equivalent to 21% (Deininger and Batzer, 2002). LINEs are 4-7 kb long, composed of two open reading frames (ORFs): ORF1 encodes an RNA-binding protein, and ORF2 an endonuclease and a reverse transcriptase (Kazazian, 2004). The 5' UTR contains an internal promoter sequence, and the 3' UTR contains a poly(A) tail (Deininger and Batzer, 2002). An intron is post-transcriptionally removed, which otherwise inhibits retrotransposition (Dewannieux et al., 2003).

1.1.1.11 Xist and Tsix

Xist and *Tsix* RNA are non-coding RNAs that are exclusively expressed during X chromosome inactivation (Borsani et al., 1991). The inactive X (Xi)

expresses *Xist*, and the active X (Xa) expresses *Tsix* on the antisense strand in a mutually exclusive manner (Zhang et al., 2007). *Xist* RNA is 17-kb long, spliced, polyadenylated, and stably associates with the entire Xi (Xu et al., 2006).

1.1.1.12 MicroRNA

MicroRNAs (miRNAs) are 17-25 base-pair long non-coding RNAs (Lai, 2002; Zeng et al., 2002). They are imperfectly complementary to the 3' UTR of specific mRNAs, to selectively block their translation (Lee et al., 1993). Similar to siRNAs, miRNAs are processed by RNase III, and associated with RNA-induced silencing complexes (RISCs), for post-transcriptional repression and mRNA degradation (He and Hannon, 2004; Yekta et al., 2004).

Processing of miRNA involves two steps. First, in the nucleus, nascent miRNAs (pri-miRNA) are processed by Drosha, an RNase III enzyme, into ~70 nt precursors (pre-miRNA) containing imperfect stem-loop structures (Lee et al., 2003). Next, pre-miRNAs are exported into the cytoplasm, where they are cleaved by another RNase III enzyme – Dicer – to give a 17-25 bp imperfect dsRNA duplex. This duplex contains both mature miRNA and its complementary strand (Ketting et al., 2001).

Evidence suggests human pri-miRNAs are capped and polyadenylated (Cai et al., 2004). For example, chromatin immunoprecipitation (ChIP) indicates pri-miRNA associates with pol II (Lee et al., 2004). However, analysis of the genes in the human chromosome (Chr) 19 miRNA cluster shows they are interspersed among Alu repeats which are transcribed by pol III (Borchert et al., 2006).

1.1.2 Polymerase I and the ribosomal genes it transcribes

In most eukaryotes, the sole function of RNA pol I is the transcription of the genes encoding the large ribosomal RNAs within the nucleolus. Approximately 400 copies of the 43-kb human ribosomal gene tandem repeat are distributed among the short arms of the human acrocentric Chrs 13, 14, 15, 21 and 22 (Henderson et al., 1972). A single polycistronic rDNA transcription unit (the 45S rRNA gene in man) encodes the precursor to 28S, 18S, and 5.8S rRNAs. It also contains functional elements that regulate pre-rRNA transcription (i.e., the rDNA promoter, enhancers, an origin of replication, transcription terminators, and a replication fork barrier) (Grummt, 2003).

1.1.2.1 Transcription of ribosomal RNA

Efficient initiation of rRNA transcription requires interactions between the DNA-binding protein, upstream binding factor (UBF1), promoter selectivity factor (SL1 in humans, TIP-IB in mouse) for the targeting of pol I, and assembly factors (TIF-IA and TIF-IC in mouse) to form a stable pol I PIC at rRNA promoters (Bell et al., 1988; Grummt, 2003). The rDNA promoter consists of a start site proximal core promoter and an upstream control element (UCE).

UBF is highly abundant in mammalian cells (~5 x 10^5 copies per cell), and has an additional role as an architectural transcription factor that maintains the "open" chromatin state of secondary constrictions (Bazett-Jones et al., 1994; Mais et al., 2005). UBF binds selectively to rDNA, and exclusively localises to nucleoli during interphase and remains associated with rDNA repeats during mitosis in active nucleolar organiser regions (NORs) (Roussel et al., 1996). A study using fluorescent recovery after photobleaching (FRAP) indicated that *in vivo*, transcription initiation

at a ribosomal promoter occurs every ~1.4 sec, and elongation occurs at 95 nt/sec on a human rDNA gene of 13.3 kb (Dundr et al., 2002).

Transcription of rRNA was first directly visualised using electron microscopy imaging of "Miller spreads" (Miller, 1981). "Miller spreads" are prepared by swelling isolated nuclei in distilled water and a household detergent to disperse the chromatin, before fixation for electron microscopy. A gradient of lateral nascent RNP fibrils are seen in the classical "Christmas tree" structure, with each fibril attached to the central chromatin axis by a polymerase (Figure.1.2A).

1.1.2.2 Processing of ribosomal RNA

Extensive chemical modifications occur in the 13.3 kb precursor rRNA before individual rRNAs are cleaved out and ribosomes assembled. In eukaryotes, the two major modifications, 2'-O-ribose methylation and pseudouridylation, are directed by C/D box and H/ACA box snoRNAs. The snoRNA sequence complementary to target rRNA modification sites forms a canonical duplex (Kiss, 2002; Boisvert et al., 2007). In contrast to the transcription machinery, the rRNA-processing machinery does not remain associated with NORs during mitosis (Boisvert et al., 2007).

Transcription of the rRNA genes by pol I is essential for the regulation of the cell cycle. Transcription is maximal in S and G2 phases, shut down during mitosis, and slowly recovers in G1 (Kuhn et al., 1998).

1.1.3 Polymerase III and the genes it transcribes

Yeast pol III has 17 subunits, 10 of which are unique to pol III (C subunits), two are common with pol I (AC subunits), and five are common to all three pols (ABC subunits) (Schramm and Hernandez, 2002). The human pol III has 16 subunits.

Three of the pol III specific subunits form a subcomplex that is involved in promoter-dependent transcription initiation (Wang and Roeder, 1997).

Pol III transcribes three classes of essential small non-coding RNAs. Class 1 contains the 5S rRNA; class 2 encompasses tRNAs, 7SL RNA, Alu RNAs, and some small viral RNAs; and class 3 includes U6 snRNA and 7SK (Hernandez, 1993).

Pol III transcription generally terminates at a run of four or more thymidine (T) residues, but some of the poly(T)s are not recognized as termination signals. Most 3' ends of pol III genes are generated by termination (Gunnery et al., 1999).

1.1.3.1 5S rRNA

In contrast to other rRNAs, 5S rRNA genes are transcribed by pol III (Haeusler and Engelke, 2006). In *S. cerevisiae*, 5S RNA genes are located in the nucleolus, but in man they are in the nucleoplasm (Dechampesme et al., 1999). A major cluster of 5S genes on human Chr1 contains 100-150 repeats (Sorensen and Frederiksen, 1991; Matera and Ward, 1992). 200-300 dispersed 5S genes scattered across the genome (Sorensen and Frederiksen, 1991). FISH indicates the major cluster often localises to the nucleolar periphery (Matera et al., 1995). It is easy to see this may facilitate the coregulation of transcription of the 5S genes with those encoding the other ribosomal RNAs.

The promoter of 5S RNA genes contains box A and box C elements. The 5S specific activator TFIIIA is required to recognise these elements before TFIIIC, TFIIIB and pol III recruitment – as with tRNA (Roeder, 2003).

5S rRNA processing is simpler than that of the other rRNAs. The 5′ end is unmodified, while the 3′ end is extended by 7 to 13 nucleotides (poly U) before

termination. The additional sequences are then removed sequentially from the 3' terminus by a slow-acting exonuclease (Piper et al., 1983; Kressler et al., 1999).

1.1.3.2 tRNA

Transfer RNAs (tRNAs) are short RNAs that function as essential adaptors to translate mRNA into a polypeptide. Each tRNA only recognises a particular amino acid and matches it to specific codons in the mRNA (White, 2005).

The internal promoter of tRNA genes contains box A and box B elements. They are recognised by the transcription initiation factor TFIIIC that recruits TFIIIB before assembly with pol III into a functional PIC (Roeder, 2003). In yeast, tRNA transcription occurs in the nucleolus. In human it is nucleoplasmic, and its products then move to nucleoli for processing and maturation (Carmo-Fonseca et al., 2000).

tRNA processing involves several steps. First, the 5' leader is removed by RNase P (Xiao et al., 2002), and the 3' trailer by ELAC2 (Takaku et al., 2003); second, a trinucleotide CCA is added to the 3' end; third, as many as 80 modifications occur before tRNA becomes mature (Hopper and Phizicky, 2003).

In human, 6% of tRNA genes contain introns. These are 14-60 nt long and located one nt 3' to the anticodon (Ogden et al., 1984). tRNA splicing starts with an endonuclease cleavage at the 3' splice sites, before this intermediate is processed by a 2' phosphotransferase to produce a mature tRNA (Trotta et al., 1997).

1.1.3.3 U6 snRNA

Unlike other snRNA, U6 snRNA is transcribed by pol III and confined to the nucleus. The promoter of the human U6 snRNA gene consists of not only a DSE (-215 to -240) and a PSE (-65 to -48) like other snRNAs, but also a TATA box-like element (-32 to -25) (Schramm and Hernandez, 2002).

In contrast to other pol III genes, the U6 snRNA is capped at the 5' end with a 7mGppp like the pol II transcribed snRNAs (Singh and Reddy, 1989); however, the timing of this process remains a mystery. U6 snRNA has a short poly(U) tail at the 3' end as a termination signal. It is associated with Sm like proteins (Spiller et al., 2007). The pseudouridylation and 2'-O-methylation of U6 snRNA is directed by snoRNAs instead of scaRNAs (Kiss, 2004). After transiently accumulating at the Cajal body, mature U6 snRNA is found mainly in splicing speckles (Huang and Spector, 1996).

1.1.3.4 7SK

The human 7SK RNA is an snRNA consisting of 331 nucleotides (Murphy et al., 1986). The phosphorylation of pol II CTD is a signal for many transcription and processing check points (above). Phosphorylation of Ser2 is mediated by P-TEFb containing CDK9 and cyclin T1 (Peng et al., 1998). The inhibition of P-TEFb by HEXIM1 (hexamethylene bisacetamide-induced protein 1) is specifically regulated by 7SK (Nguyen et al., 2001; Yang et al., 2001; Yik et al., 2003).

The 7SK gene has a similar upstream promoter to that of U6 snRNA, with a DSE, a PSE, and a TATA box-like element (Murphy et al., 1986). Its RNA also possesses an mpppG cap at the 5' end which is added post-transcriptionally (Shumyatsky et al., 1990). The RNA has a short poly(U) tail at the 3' end, which is further uridylated or adenylated post-transcriptionally (Chen et al., 2000).

1.1.3.5 7SL, *Alu*, B1, and B2 RNAs

7SL RNA is a class 2 pol III RNA; the gene contains an internal promoter. The RNA forms part of the signal recognition particle that inserts proteins into membranes (White, 2004). 7SL RNA generates processed pseudogenes, some of

which have the S domain deleted and an A-rich sequence added. After amplification, two such monomers fuse to generate *Alu* elements: the left gains a functional B box, and the right loses its promoter (Dewannieux et al., 2003; Kazazian, 2004).

Alu RNAs belongs to the short interspersed element (SINE) family. SINEs relocate in the genome via retrotransposition (Shedlock and Okada, 2000). *Alu*s share a ~300 nt consensus sequence, and are present in 1.1 million copies (or 11%) in the human genome (Kazazian, 2004). Pol III transcribed *Alu*s can be mobilised by LINEs (Dewannieux et al., 2003). *Alu*s can cause insertional mutation, recombination, gene conversion and alterations in gene expression (Batzer and Deininger, 2002).

In mouse, B1 and B2 RNAs transcribed from SINEs have very similar structures to *Alu* RNAs (Allen et al., 2004). Under heat-shock or translation inhibition, B1 and B2 transcription is upregulated (Liu et al., 1995). B2 RNA can bind to pol II and inhibits its initiation, but not elongation (Allen et al., 2004; Espinoza et al., 2004).

1.1.4 Export pathways for different transcripts

RNA export from the nucleus to the cytoplasm occurs via nuclear pore complexes (NPCs) (Ryan and Wente, 2000). It is dependent on the assembly of the RNA into the appropriate ribonucleoprotein complex. Non-coding RNAs are exported using Ran-dependent transport factors, while protein-coding RNAs are exported by Ran-independent pathways (Cullen, 2003).

1.1.4.1 Ran-dependent export pathways of non-coding RNAs

Both U snRNAs and rRNAs export is mediated by the karyopherin, Crm1 (Cullen, 2003). Pre-snRNA is assembled into a large export complex, which includes the CBC, the phosphorylated adaptor for RNA export (PHAX), and the Crm1/RanGTP complex. After export through the NPC, PHAX is dephosphorylated, and the snRNA export complex disassembled through GTP hydrolysis of Ran (Ohno et al., 2000; Kiss, 2004).

In yeast, after extensive processing, the 28S, 5.8S and 5S rRNA are first assembled with 50 ribosomal proteins into the 60S pre-ribosomal subunit. The 18S rRNA assembles with 33 ribosomal proteins into the 40S preribosomal subunit (Venema and Tollervey, 1999). They are exported separately, although both pathways involve Crm1 (Hurt et al., 1999; Moy and Silver, 2002). The 60S subunit, but not 40S, requires an adaptor Nmd3p to access the pathway (Hedges et al., 2006).

tRNA export requires Exp-t, also a member of the karyopherin family, which actively interacts with Ran-GTP. The presence of an unremoved intron does not affect export (Arts et al., 1998), in contrast to the case of mRNA (below).

pre-miRNAs are exported into the cytoplasm by exportin 5, a Ran-GTP dependent transporter (Lund et al., 2004).

1.1.4.2 Ran-independent export pathways of mRNAs

After processing, UAP56 is recruited to mRNA molecules. Next, Aly is recruited, which binds the TAP-Nxt nuclear RNA export factor. Another unidentified factor may substitute for the role of Aly (Cullen, 2003).

Both processing of the 3’ end and splicing are critical for mRNA export as pre-mRNA mutants defective in either splicing or 3’ end formation fail to be

exported (Custodio et al., 1999). *In vitro* evidence suggested Aly is recruited during mRNA splicing (Zhou et al., 2000). Furthermore, 3' processing, but not polyadenylation, facilitates mRNA export (Cullen, 2003). The export of histone mRNAs, which do not contain poly(A) tails, is also mediated by TAP (Huang et al., 2003). During stress, Rpb4p, the fourth largest subunit of pol II mediates mRNA export (Farago et al., 2003).

RNA length plays a role in choice of transport pathway. When shortened to <120 nt, mRNAs are exported by the U snRNA pathway, while inserting 300 nt drives U1 snRNA to the mRNA export pathway (Ohno et al., 2002).

1.1.5 Mitochondrial DNA transcription

Mitochondria (mt) are found in the cytoplasm, and they are essential for cellular ATP production. In mammals, the 16,600 bp circular double-stranded mtDNA encodes 13 proteins, 2 rRNAs and 22 tRNAs, (Falkenberg et al., 2002).

Mt transcription produces polycistronic precursors from both strands which are then processed into mRNAs, rRNAs, and tRNAs (Montoya et al., 1981; Shadel and Clayton, 1997). Processing is mainly achieved through cleavage by the tRNA processing enzymes (Rossmanith et al., 1995). The mammalian mtDNA genome does not contain introns (Shadel and Clayton, 1997). The 5' end of mt-mRNAs carries a monophosphate and is uncapped; the 3' end is polyadenylated. The mt-rRNAs have 3' poly(A) tails, and mt-tRNAs a CCA addition like the cellular tRNAs (Gagliardi et al., 2004). Mitochondria have their own transcription machinery, which is different from that in the nucleus. The catalytic subunit and replicative helicase of

mtDNA directed RNA pol (POLRMT) are similar to those in bacteriophages T7 and T3 pols (Shutt and Gray, 2006).

1.1.6 Summary

Table 1.1. Genes are transcribed and processed differently

Gene	Function	Transcription		Processing		
		pol	location	capping	splicing	3' editing
45S rRNA	protein synthesis	I	no	-	-	pseudouridylation / 2'-O-methylation (snoRNA)
mRNA	encoding protein	II	n	co	co	poly(A)
spliceosomal snRNA	mRNA processing	II	n	co	-	as rRNA (scaRNA) 3' box
histone	nucleosome formation	II	n	co	-	stem loop (U7 snRNA)
snoRNA/ scaRNA	snRNA and rRNA processing	II	n	co	from intron	trimming
Xist / Tsix	X inactivation	II	n	co	-	poly(A)
LINEs	Retrotransposon	II	n	co	post	poly(A)
tRNA	protein synthesis	III	n/no	-	post	CCA
5S rRNA	protein synthesis	III	n/no	-	-	poly (U)
7SK	CTD regulation	III	n	post	-	poly (U)
U6 snRNA	mRNA processing	III	n	post	-	poly (U)/(snoRNA)
7SL	protein transport	III	n	-	-	A rich sequence
Alu	retrotransposon	III	n	-	-	A rich sequence
miRNA	gene silencing	II /III?	n	co	-	poly(A)/Drosha/Dicer
mitochondrial	protein coding/ tRNA/rRNA	POLRMT	c	-	-	poly(A)/CCA/ A rich sequence

[Abbreviations]: n: nucleus; no: nucleolus; c: cytoplasm; co: co-transcriptional processing; post: post-transcriptional processing.

1.2 The structure of the nucleus

The mammalian nucleus is separated from the cytoplasm by a double-membrane nuclear envelope. It contains the essential genetic information and machineries for gene expression. Nuclear pores penetrate the envelope and allow selective transport between nucleus and cytoplasm (Stoffler et al., 1999; Cook, 2001). The peripheral nuclear lamina is considered responsible for anchoring chromatin to the nuclear periphery and maintaining envelope structure. It contains lamins A/C and B located inside the envelope (Moir et al., 2000).

Most of the nucleus is filled with chromatin during interphase. Individual chromosomes can only be resolved during mitosis. The nucleoli contain high concentrations of RNA, and are responsible for rRNA synthesis and ribosome assembly (Boisvert et al., 2007). Increasing numbers of specialised interchromatin nuclear domains (Spector, 2001) have been identified. The structures of these nuclear domains are highly dynamic, with components rapidly exchanging with the soluble pool (Misteli, 2001).

1.2.1 Chromatin packaging during interphase

A typical human chromosome contains ~100 Mbp DNA with a contour length of ~34 mm, sophisticatedly packaged into a ~10 μm nucleus (Cook, 2001). Nucleosomes are the fundamental structural units of eukaryotic chromatin. Each nucleosome contains 147 bp DNA wrapped in 1.75 turns around an octamer of core histones (Luger et al., 1997). Nucleosomes may change their positions along the DNA, and the histone tails undergo numerous modifications (Fuks, 2005). A string of nucleosomes are packaged into high-order structures which are either condensed

as heterochromatin or open as euchromatin. Heterochromatin is often observed at the nuclear periphery or nucleolus, but can also aggregate into internal chromocentres. Finally these chromosomes are organised into territories (Cavalli, 2002; Cook, 2002) (reviewed in Ch.1.3.2).

1.2.2 Nuclear domains

1.2.2.1 The nucleolus

Nucleoli contain hundreds of tandem rRNA genes and numerous factors necessary for transcription of rDNA, processing of pre-rRNA, and ribosome assembly (Ch.1.1.2.). It may also be a site of covalent RNA modification and protein assembly for RNP complexes, such as snRNPs, telomerase and several other pol III transcribed small RNAs, such as 5S rRNA, some tRNAs, RNase P RNA, 7SL RNA and miRNAs (Gerbi et al., 2003; Boisvert et al., 2007).

Electron microscopic imaging of nucleoli reveal three distinct subcompartments: the fibrillar centre (FC), dense fibrillar component (DFC), and granular component (GC) (Huang, 2002). Transcription takes place on the surface of the FC (Hozak et al., 1994). Processing of nascent transcripts occurs in the GC (Fatica and Tollervey, 2002).

1.2.2.2 Splicing speckles

25-50 nuclear speckles containing pre-mRNA splicing factors can be localised against a diffuse pool using immuno-fluorescence. Aggregates of speckles, interchromatin granule clusters (IGCs) are suggested to be involved in the assembly and modification of pre-mRNA splicing factors (Spector, 1993). They are not sites of

transcription, as they contain inactive pol II phosphorylated on Ser2 of the CTD (Pombo and Cook, 1996; Xie et al., 2006), nor are they sites of splicing (Misteli, 2000); rather, they seem to be stores of processing factors (Spector, 2001). Active genes are often found at the periphery of speckles (Xing et al., 1995).

1.2.2.3 Cajal bodies

1-10 Cajal bodies are present in the nucleus with a diameter of 0.2-1.0 μm (Gall, 2000); they contain factors involved in snRNP and snoRNP biogenesis (Ch.1.1.1.8 and 1.1.1.9). These factors move through the Cajal body to nuclear speckles or nucleoli for snRNP or snoRNP maturation (Sleeman and Lamond, 1999). Gemini of Cajal bodies are often coincident with or adjacent to Cajal bodies. However, they contain survival of motor neurons gene product (SMN) and Gemin2 which are absent from Cajal bodies (Matera, 1999). Gemini may play an additional role in snRNP maturation (Spector, 2001).

1.2.2.4 OPT domain

OPT (Oct1/PTF/transcription) domains are 1.0-1.5 μm, and contain the PSE-binding transcription factor (PTF) and Oct-1 – two transcriptional activators (Pombo et al., 1998). They assemble in G1 phase and disappear in S phase. A 30 Mbp region on Chr6 (band 6p21) often associate with this domain (Pombo et al., 1998).

1.2.2.5 Promyelocytic leukaemia (PML) bodies

A typical human nucleus contains 10-30 PML bodies, of 0.3-1.0 μm. These bodies contain PML protein, SUMO1, HAUSP and CBP. In patients suffering from acute promyelocytic leukaemia, PML bodies break into smaller domains. This is caused by a translocation in which the PML gene is fused to the α-retinoic acid

receptor. PML bodies are suggested to be involved in transcriptional regulation, apoptosis, DNA repair, and cell growth (Wang et al., 1998; Eskiw et al., 2004).

1.2.2.6 Other nuclear domains

Paraspeckles are often found adjacent to speckles in both primary and transformed human cells. They contain proteins (PSP1, PSP2, and p54/nrb) that can cycle to and from nucleoli (Fox et al., 2002). Paraspeckles may be involved in regulating nuclear retention of stable RNAs (Prasanth et al., 2005).

Perinucleolar compartment (PNC) and SAM68 nuclear bodies are associated with the periphery of nucleoli. They contain pol III transcribed RNAs and RNA-binding proteins, and are mainly found in cancer cells (Dundr and Misteli, 2001).

The association with the nucleolus of both paraspeckles and PNCs is transcription-dependent (Dundr and Misteli, 2001; Fox et al., 2002).

1.3 Intimate relation between gene regulation and nuclear structure

In eukaryotes, the organisation of chromatin into high-order structures is tightly connected with the control of gene expression and other chromosomal processes (e.g., chromosome stabilisation and segregation, imprinting, dosage compensation, and recombination) (Hall et al., 2002).

Transcriptional states are also determined by high-order chromatin structure – for example, “euchromatin” (on) and “heterochromatin” (off) (Turner, 2000). It is proposed these structures are dependent on the local concentration and combination of different histone modifications (Huang, 2002; Margueron et al., 2005a; Heintzman et al., 2007). Using ChIP-on-chip (Kim and Ren, 2006; Barski et al.,

2007), a high resolution map of these modifications in human nuclei is emerging. A comprehensive analysis of functional elements the in human genome by the ENCODE project is also taking place (The ENCODE consortium, 2007). These studies are revealing a tight coupling between epigenetic regulation and the processes of replication and transcription.

1.3.1 Histone modifications and DNA methylation

In different cell types, combinations of histone modifications determine gene expression at a specified time and place (Turner, 2007). Active promoters are marked by an overall increase in acetylation of core histones and trimethylation of Lys4 of histone H3 (H3K4) (Jenuwein and Allis, 2001; Santos-Rosa et al., 2002; Margueron et al., 2005b), and enhancers by monomethylation of H3K4 (Heintzman et al., 2007). The monomethylations of H3K27, H3K9, H4K20, H3K79, and H2BK5 are all linked to gene activation, while trimethylations of H3K27, H3K9, and H3K79 are linked to repression. H2A.Z also associates with functional regulatory elements (Splinter et al., 2006; Barski et al., 2007).

Phosphorylation of the pol II CTD is also related to histone modification. H3K4 methylation by Set1p methylase is a marker of early elongation, correlating with Ser5 phosphorylation of the CTD (Ng et al., 2003). By contrast, H3K36 methylated by Set2p marks late elongation and termination, correlating with Ser2 phosphorylation (Bannister et al., 2005).

Histone modifications appear at early stages of differentiation (Okamoto et al., 2004). In preimplantation embryos, Xi demonstrated H3K4 hypomethylation and

H3K9 hypoacetylation at the 8-cell stage; next, H3K27 is methylated. MacroH2A is enriched in the morula; H3K9 is methylated in the blastocyst (Lucchesi et al., 2005).

The different modifications provide distinctive binding sites for effectors (Jenuwein and Allis, 2001). At the simple level, the bromodomain in histone-acetyltransferase (HAT)-associated transcriptional co-activators selectively interacts with acetylated lysine (Dhalluin et al., 1999). The chromodomain of heterochromatin protein HP1 binds methylated H3K9, but little with H3K4 (Bannister et al., 2001). The chromodomain of polycomb protein (PC) binds to trimethylated H3K27 (Czermin et al., 2002). At a higher level, a number of studies implicated that the plant homeodomain finger found in a variety of proteins is a highly specialised methyl-lysine binding domain to discriminate different methylation states (Mellor, 2006).

In addition, DNA methylation plays an important role in transcription silencing/activation. Methylation results in loss of DNase-I hypersensitive sites and transcriptional repression (Keshet et al., 1986). In vertebrates, 60-90% of all CpG dinucleotides are methylated (Ng and Bird, 1999). The methylation-free CpGs are found in high densities in GC-rich regions (CpG islands) at the 5' promoter region of many human genes (Bird, 2002). Evidence suggest DNA methylation does not silence active promoters, but affects genes that are already silent (Bird, 1999). Inhibition of histone deacetylases can reactivate genes previously silenced by DNA methylation (Ng and Bird, 1999). On the other hand, DNA methylation may also feedback on lysine methylation, suggesting an interaction between DNA methylation and histone modification (Fuks, 2005).

1.3.2 Chromosome territories

Compared with the extensive detail we know about histone modifications and DNA methylation, high-order of chromatin organisation is less understood. Current models suggest that chromosomes in the nucleus are organised into chromosome territories (CTs) (Cremer and Cremer, 2001). CTs are built up from a hierarchy of chromatin fibres with increasing thickness: the 10 and 30 nm fibres and the 60–130 nm fibres (Belmont and Bruce, 1994). Chromatin loops build up rosettes, which are further organised into CTs (Munkel et al., 1999).

CTs were originally believed to be separated by an interchromatin domain (ICD), containing the transcription and processing machineries (Zirbel et al., 1993). Active genes are located at the periphery of CTs and loop out to contact with the ICD (Kosak and Groudine, 2004), interacting with the different nuclear bodies (Cremer and Cremer, 2001). In fact, CTs are highly dynamic. High resolution FISH showed CTs intermingle significantly, correlating with the frequency of chromosome translocations (Branco and Pombo, 2006). Recent research on chromatin loops demonstrates that spatial interactions of genes are not at all restricted within the chromosome (Osborne et al., 2004; Ling et al., 2006; Splinter et al., 2006) .

1.3.3 Chromatin loops

The chromatin structure at the mouse β-globin locus has been studied comprehensively using FISH and chromosome conformation capture (3C) (Fraser and Bickmore, 2007). The *Hbb-b1* gene lies 50 kb away on Chr7 from its locus control region (LCR). A globin stabilising protein *Eraf* is located 24.4 Mb away from *Hbb*. *Hbb-b1* contacts the LCR and *Eraf* only when they are actively

transcribed in erythroid nuclei, but not in brain when they are inactive (Carter et al., 2002; Tolhuis et al., 2002; Osborne et al., 2004).

Recently, genome-based 3C derivatives enable a wider view of chromatin loops (Ling et al., 2006; Lomvardas et al., 2006; Wurtele and Chartrand, 2006; Zhao et al., 2006). Using associated chromosome trap, Ling *et al.* showed insulin-like growth factor 2 (Ifg2)/H19 imprinting control region (ICR) on Chr7 colocalises with one allele of Wsb1/Nf1 on Chr11; omission of CCCTC-binding factor (CTCF) or deletion of the maternal ICR abolishes the association (Ling et al., 2006). Using circular 3C, Zhao *et al.* identified 114 unique sequences from all autosomes, several of which interact primarily with the maternal H19 ICR (Zhao et al., 2006). Combining 3C with microarrays (3C-on-chip) for larger scale analysis, de Laat and colleagues confirmed the active β-globin locus in fetal liver preferentially contacts transcribed loci on Chr7, while in fetal brain it contacts different silent loci. Furthermore, a housekeeping gene in a gene-dense region on Chr8 forms long-range contacts with other active gene clusters, both in *cis* and in *trans* (Simonis et al., 2006). These studies revealed an extensive network of intra- and inter-chromosomal interactions.

What is the underlying mechanism organising specific interaction of chromatin loops? What is the driving force and how are the loops regulated?

1.3.4 Models of nuclear organisation

An early and widely accepted model of nuclear organisation was helical hierarchies, in which nucleosomes strings are coiled into solenoids, and solenoids into higher-order structures (Sedat and Manuelidis, 1978). Another opinion is

random packing of nucleosomal strings (Sachs et al., 1995). However, it is very difficult for these models to justify the specific interactions of chromatin loops.

A third model suggests loops attach to a peripheral lamina or to internal structures like skeletons/matrices/scaffolds (Saitoh and Laemmli, 1993; Spann et al., 2002). A number of studies suggest nuclear scaffold/matrix-attachment regions (S/MARs) can attach newly-made nucleic acids to some nuclear skeleton (Davie, 1995; Heng et al., 2004). S/MARs may function as the mediators of loop attachment. However, these anchors alone are not sufficient for chromatin loops to form (Heng et al., 2004). Furthermore, the successive hypo- and hyper-tonic extractions for preparation of nuclear matrix are questionable, since various macromolecular rearrangements could occur (Pederson, 2000). The basic structure of S/MARs and its relation to factories remain undefined.

Other evidence suggests the transcription machinery organises the loops (Cook, 1999; de Laat and Grosveld, 2003; Osborne et al., 2004; Zhou et al., 2006). In this model, the chromatin fibre is attached through polymerases or transcription factors to a “transcription factory” (Cook, 1999) or “active chromatin hub” (de Laat and Grosveld, 2003), both refer to a spatial clustering of *cis* regulatory elements and active genes (Chakalova et al., 2005). Attachments of chromatin loops to a factory or hub are highly dynamic as polymerases initiate and terminate, transcription factors associate and disassociate (Cook, 2002). Frequent exchange of protein components occur between the factories or hubs and the soluble pool (Kimura, 1999).

1.4 Transcription factories

The term "factory" is used to define 'a building or range of buildings with plant for the manufacture of goods'. The first visualised "factories" were sites of viral replication seen after infection. Using [^{3}H]thymidine, a cytoplasmic focus of viral DNA synthesis was visualised by autoradiography (Cairns, 1960). The term "factory" (Joklik, 1968) was then applied to active sites of replication and transcription in uninfected cells (Hozak et al., 1993; Jackson et al., 1993).

1.4.1 Nucleolar transcription factories

The nucleolus is the best characterised transcription factory (Ch.1.2.2). The FC contains high concentrations of pol I and UBF. Transcription occurs on the surface of FCs (Hozak et al., 1994; Cmarko et al., 2000; Koberna et al., 2002). In HeLa cells during interphase, a nucleolar factory contains ~30 FCs, each containing ~500 pols associated with 4 active genes (Jackson et al., 1998) (Figure.1.2B). Such templates are stripped off FCs during the preparation of Miller spreads to give the "Christmas trees" with their ~125 "branches" (Figure. 1.2A).

In yeast, the number of active transcription units and FCs increases with pol I initiation rate, however the number of active pol I per nucleolus remains constant (French et al., 2003). This indicates the factory number and diameter is related to the rate of rRNA synthesis (Cook, 2001).

1.4.2 Immobile polymerases

It is widely assumed that active polymerases track along their templates as they make RNA. However, evidence suggests active pols are immobilised. 1) If pols track along a helical template, the resulting transcript will become entwined about

the DNA. An additional precise mechanism is required to solve this entanglement problem. However, if pols are fixed, they extrude un-entwined transcripts as they reel in and rotate and template. 2) Nascent RNA, transcribed DNA, and polymerising activity all remain after most chromatin is removed after ribonuclease treatment (Dickinson et al., 1990; Jackson and Cook, 1993; Jackson et al., 1993). If pols track along the DNA template, transcriptional activity should be lost (Fig 1.1). 3) Nascent RNA is concentrated in a few sites (below).

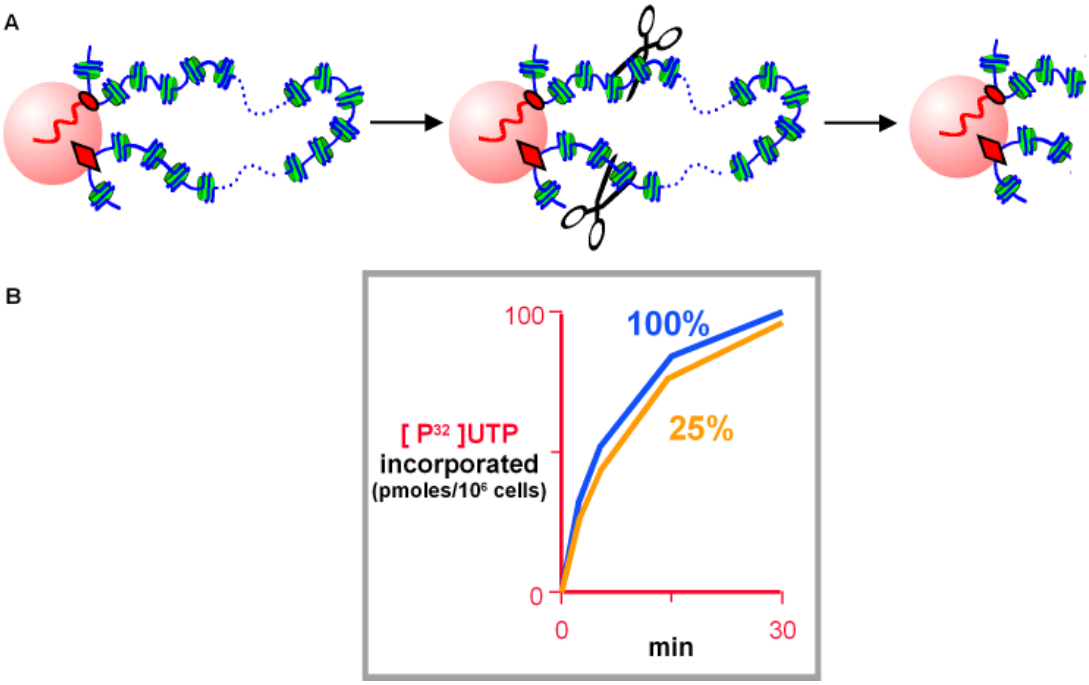


Fig. 1.1. Nuclease digestion indicates pols are immobilised

A. Model. If active pols (oval) tie chromatin to factories (sphere), polymerising activity should remain after nuclease digestion and removing the detached chromatin.

B. Results of a nuclease digestion experiment (Dickinson et al., 1990). [^{3}H]labelled cells were encapsulated, lysed and washed, incubated with or without *Eco*RI and detached fragments electroeluted. The time-course of incorporation of [^{32}P]UTP into acid-insoluble material by beads treated in the two ways is shown. Blue curve: control, digestion with *Eco*RI but without electroelution to remove chromatin; orange curve: digested and electroeluted (25% chromatin remained). Although most chromatin was removed, most polymerising activity remains.

1.4.3 Localising active transcription sites

Active sites of transcription can be visualised by either growing cells in Br-U (Iborra et al., 1998), or permeabilising cells before allowing engaged pols to extend their transcripts in Br-UTP, and immunolabelling the resulting Br-RNA (Jackson et al., 1993). Tagged nascent transcripts are found in a limited number of discrete sites (Jackson et al., 1993; Wansink et al., 1993; Iborra et al., 1996; Fay et al., 1997) (Fig 1.2A). Active polymerases and transcription factors colocalise with such discrete foci, termed "transcription factories" (Iborra et al., 1996; Grande et al., 1997). Moreover, RNA-FISH coupled to immunolabelling confirmed active transcription units on different chromosomes colocalise in such factories rich in pol II (Osborne et al., 2004), indicating factories are the organisers of chromatin loops.

Other evidence supports the idea that active polymerases are the major molecular ties attaching chromatin loops to the factories. For example, the lampbrush loops seen after swelling oocyte nuclei in hypo-tonic buffers are only found when the relevant regions are transcriptionally active (Gall and Murphy, 1998). Similarly, in living cells transcription inhibitors like DRB or α-amanitin increase the mobility of loops (Chubb et al., 2002), presumably by breaking the attachment of loops to the underlying structure. In addition, TFs act as additional and more transient ties as sequences containing transcription-factor binding sites also resist nucleolytic detachment (Jackson and Cook, 1993). These ties are likely to persist for only a few seconds (Misteli, 2001; Chen et al., 2002).

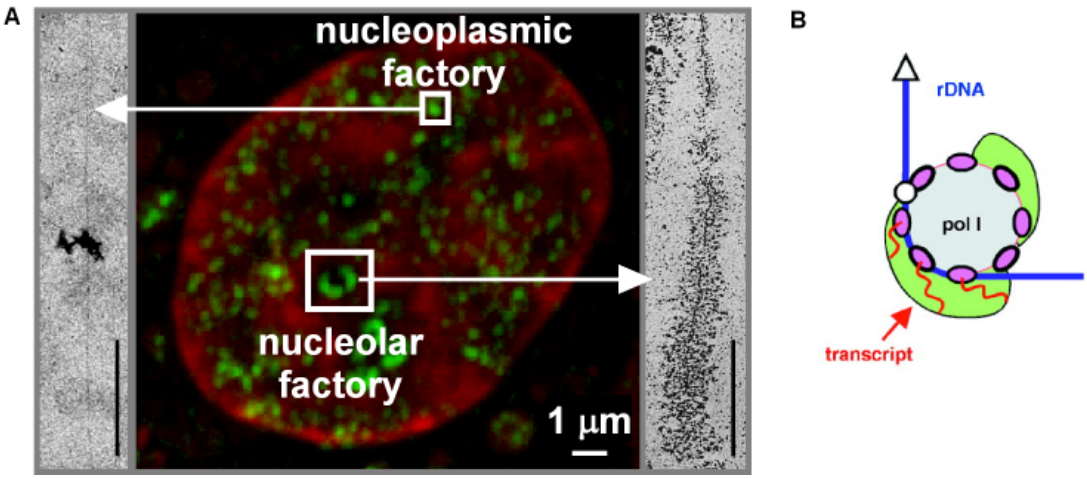


Fig. 1.2. Transcription factories in HeLa cells

Adapted from (Cook, 1999).

A. Localising nascent RNA.

Middle. Cells were permeabilised, nascent RNA extended in Br-UTP, cryosectioned (100 nm), Br-RNA immunolabelled with FITC (green), nucleic acids counterstained with TOTO-3 (red), and an image collected on a confocal microscope. Nascent RNA (green) is concentrated in factories in the cytoplasm (made by mt pols), nucleoplasm (by pol II and III), and nucleoli (by pol I). Bar: 1 µm. Image provided by A. Pombo.

Left. One transcript (probably made by pol II) is associated with DNA (which runs from top to bottom). Breaking a nucleoplasmic transcription factory like the one in the middle releases ~8 active units like this (Jackson et al., 1998). Bar: 1 µm (~2.9 kb DNA).

Right. One active rRNA transcription unit with ~125 transcripts, formed by stripping away one of the two crescents in a nucleolar transcription factory shown in the middle (Miller and Bakken, 1972). Bar: 1 µm

B. Model for a nucleolar factory. The nucleolar factory in (A) contains two crescent-shaped foci (DFC), each with nascent transcripts from one transcription unit. The transcripts and template in only one of the two crescents are shown in (B). rDNA slides (open arrowhead) through pols (ovals) on the surface of a FC containing pol I, as nascent transcripts are extruded to form the DFC.

1.4.4 The relation between factories and other nuclear structures

Many studies suggest lamins play an important part in nuclear organisation and transcription regulation (Spann et al., 2002; Akhtar and Matera, 2006; Vlcek and Foisner, 2007). Lamins seem to be critical factors for anchoring chromatin to the nuclear periphery (Goldman et al., 2002). Alteration of lamin organisation by over expression of a dominant negative lamin mutant inhibits pol II transcription (Kumaran et al., 2002; Spann et al., 2002). Factories are probably attached to a lamin-containing skeleton.

Other skeletons containing nuclear actin and myosin probably also interact with factories, as these proteins copurify with pol II and micro-injecting antibodies against them inhibit transcription (Pestic-Dragovich et al., 2000).

1.4.5 Size and number of nucleoplasmic transcription factories

Nucleoplasmic factories are very small and numerous. In a typical HeLa cell, there are on average ~8,000 pol II factories, each containing ~8 active molecules of pol II; each factory is ~50 nm in diameter, and the distance between factories is ~300 nm (Jackson et al., 1998; Kimura et al., 1999). Therefore, they can only be resolved one from another using the highest resolution microscopy [i.e. electron microscopy (EM) or confocal imaging of cryosections (Iborra and Cook, 1998; Jackson et al., 1998)].

The average contour length of loops in HeLa cells was estimated to be ~86 kb. This was calculated from the fraction of chromatin resisting electroelution after nuclease digestion (Jackson et al., 1990). Coincidently, on average, one S/MAR is present at every 60-90 kb of genomic DNA (Boulikas, 1995). Two populations of

loops seem to be present, ~20% are around ~7.5 kbp, and the rest evenly distributed from 50-175 kbp (Jackson et al., 1990). We could only speculate the shorter represent the active loops, and longer inactive.

Table 1.2 lists the number of transcription factories in a HeLa cell from various early studies. The number of factories seems to vary among different cell types. A quantitative study on Br-labelled nascent RNA foci suggested the number of factories falls during differentiation of mouse embryonic stem cells, as nuclei become smaller, but site density and diameter remain roughly constant. Similar site densities and diameters are found in salamander (amphibian) cells with 11-fold larger genomes, and in aneuploid HeLa cells (Faro-Trindade and Cook, 2006). Another study on erythroid cells reported an average of 100−300 pol II factories per nucleus, counted from immuno-labelled pol II foci (Osborne et al., 2004). Regardless the method used to determine the number of factories, it is necessary to ensure all the foci are seen by establishing the detection threshold (Faro-Trindade and Cook, 2006).

1.4.6 Transcription factory model for genome organisation

Our current model suggests active pols on different transcription units cluster into a factory surrounded by a “cloud” of loops. The cluster is rich in pols and other factors and the high concentration supports efficient interaction with DNA. Pols and TFs act as molecular ties, transiently tethering transcription units to the factory. Such dynamic attaching and detaching of the chromatin loops to the transcriptional machinery is the driving force behind transcriptional regulation (Cook, 1999; Cook, 2002) (Fig.1.3).

Table 1.2. Numbers of pols, nascent transcripts, and factories in a HeLa cell

Marker analyzed (approach used)	Number (diameter, nm)					
	pol I	pol II + III	pol II			pol III
			$II_A + II_O$	II_A	II_O	
1 Total pols (westerns)			320,000	150,000	170,000	
2 Active pols (sarkosyl, westerns)			86,000	22,000	64,000	
3 Nascent transcripts ([^{32}P]UTP)	15,000		65,000			10,000
4 Nascent transcripts (Br-UTP; LM)	15,000	55,000				
5 Bio-RNA sites (bio-CTP; EM)		7,800				
6 Br-RNA sites (Br-UTP; LM/EM)		2,400	8,000			1,800
7 Br-RNA sites (Br-U; LM/EM)	30	8,600				
8 Pol II/II_O sites (EM)			8,000		10,800	
9 Diameter bio-/Br-RNA sites (EM)		50				43
10 Diameter pol sites (EM)			50		46	
11 Pols/transcription unit (calculated)	125		1			1
12 Active pols/site (calculated)	500		8		7	5

Data from (Iborra et al., 1996; Iborra and Cook, 1998; Jackson et al., 1998; Kimura et al., 1999; Pombo et al., 1999). pol II_A / IIo are pol II with hypo- / hyperphosphorylated II CTD (Bregman et al., 1995).

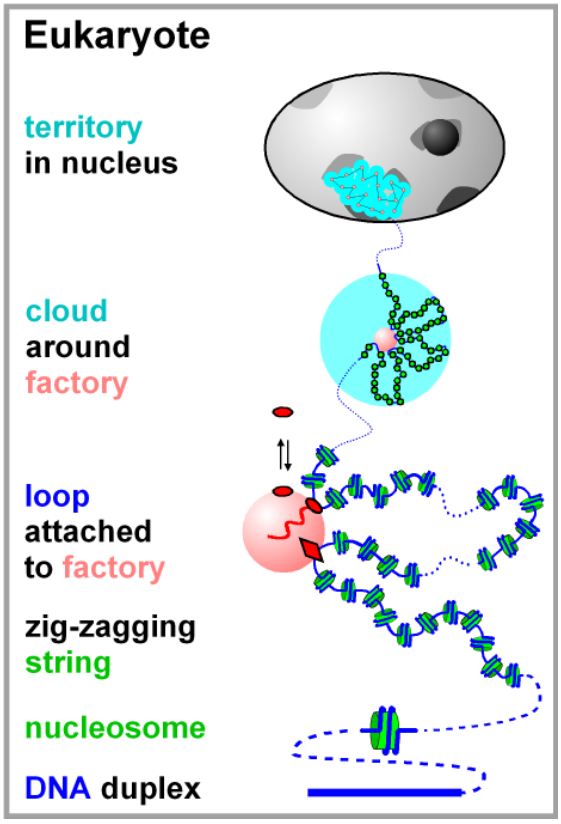


Fig. 1.3. A model for eukaryotic genome organisation

Adapted from (Cook, 1999; Cook, 2001)

DNA winds around the core histones, and forms a zig-zagging string of nucleosomes. This string is organised into loops by attachment to TFs (diamonds) and engaged pols (ovals). 10-20 such loops (only 3 are shown) form a cloud around the factory. Transcription units on different chromosomes will also bind to the same factory (not shown).

Active pols are bound to a transcription factory and act both as motors that reel in their templates and as one of the critical structural ties that maintain the loops. Loops appear and disappear as pols initiate and terminate, and the factors bind and dissociate. Nucleosomes in long loops are static and acquire a (heterochromatic) histone code. They also aggregate onto the lamina, nucleoli, and chromocentres.

Individual components in the factory exchange continually with others in the soluble pool. 50-200 successive clouds strung along the chromosome form a territory (the general path of DNA between clouds is shown).

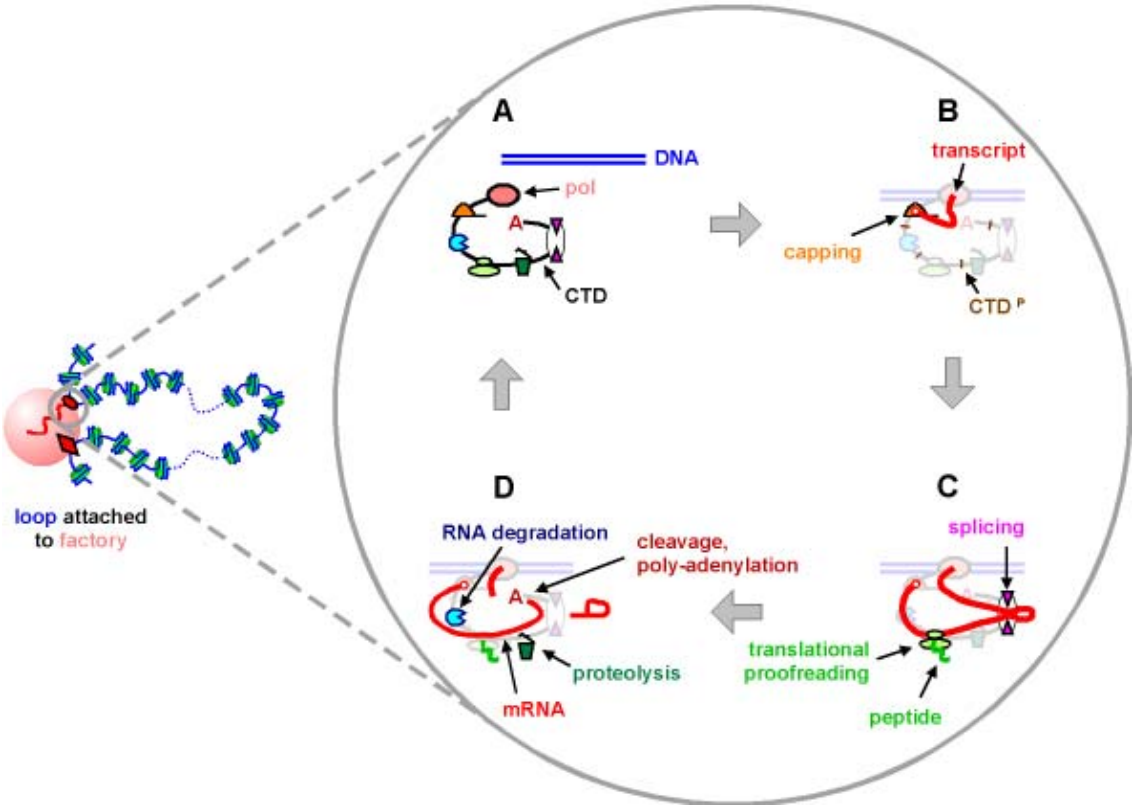


Fig. 1.4. Model of a polymerising complex in a transcription factory

Adapted from (Cook, 2001; Iborra et al., 2004a)

One chromatin loop attached to a transcription factory containing pol II is shown on the left. Only one of ~8 polymerising complexes of a pol II factory is shown on the right. This complex contains all the machinery necessary to create an mRNA.

A. The CTD has the potential to associate with sites involved in capping, transcript degradation, translational proofreading (NMD), proteolysis, splicing, and polyadenylation.

B. Transcription begins as the template bound to the polymerising complex and is reeled in as the transcript is extruded; the CTD is now hyper-phosphorylated, and a cap has been added.

C. The transcript continues to be extruded through a splicing site as the ribosome/NMD machinery begins proofreading the nascent message.

D. Once introns are removed (lariat), the transcript is cleaved and polyadenylated, and is ready to leave for the cytoplasm; but if errors are detected, the faulty transcript and peptide produced during proofreading are degraded by nucleases and proteasomes.

As reviewed in Ch.1.1, transcription and processing (capping, splicing, 3' end processing) are inter-dependent and often occur simultaneously. The pol II CTD organises these functions. This complex also proofreads nascent transcripts via nonsense-mediated decay (NMD), detecting premature termination codons (PTCs) (Hilleren and Parker, 1999). Nascent peptides were detected in the nucleus, suggesting nuclear translation underlies NMD (Iborra et al., 2004a; Iborra et al., 2004b). The components of NMD machineries and translation colocalise with nucleoplasmic factories, and co-purify with CTD, indicating a molecular cross-talk between the three (Iborra et al., 2004a). A model for a "mega-complex" organising all gene expression events for generating a protein coding RNA is illustrated (Fig.1.4).

1.4.7 Evidence suggests factories specialise

Previous research showed the different classes of polymerases are concentrated in their own dedicated factories (Pombo et al., 1999) (Ch.5.1). The first evidence suggesting factories specialise further was obtained from a study of the OPT domain (Ch.1.2.5). This domain probably contains two or three transcription factories, and is often associated with ~30 Mb on human Chr6. This implies that genes in this chromosomal region that require PTF are transcribed in association with the OPT domain (Pombo et al., 1998).

Furthermore, globin-related genes seem to be transcribed in particular factories. In fetal liver, active *Eraf* and *Hbb-b1* — 24.4 Mb apart on the genetic map — nevertheless are transcribed in the same pol II factory (Ch.1.3.3), while inactive genes in brain are not (Osborne et al., 2004). Active α- and β-globin genes

frequently juxtapose, correlating with transcription activity (Brown et al., 2006). Besides, active α1 and α2, but not the inactive ζ gene, colocalise with neighbouring housekeeping genes C16orf33, C16orf8, MPG and C16orf35 (Zhou et al., 2006). Furthermore, during immediate-early gene induction in mouse B cells, *Myc* proto-oncogene on Chr15 is preferentially recruited to the same transcription factory as the highly transcribed *Igh* gene located on Chr12, correlating with their frequent translocation (Osborne et al., 2007).

In addition, many genes requiring special processing machinery are found to be associated with the nuclear domain rich in those particular processing factors. Artificial tandem arrays of human U1 and U2 snRNA genes colocalise in the same factories on Cajal bodies, and the association is abolished by promoter mutations that prevent transcription (Frey et al., 1999). In *Saccharomyces cerevisiae*, five different tRNAs colocalise with 5S ribosomal DNA and U14 small nucleolar RNA factories on the surface of the nucleolus, and a point mutation in the promoter eliminates this association (Thompson et al., 2003). On heat shock, satellite III repeats on human 9q12 are transcribed and associate with a factory containing heat-shock factor 1 (Jolly et al., 2004). Gene-dense regions of various human chromosomes, in particular the major histocompatibility (MHC) locus, are transcribed at the periphery of PML bodies (Shiels et al., 2001; Wang et al., 2004). In adenovirus-infected HeLa cells, exon-junction complex (EJC) proteins (REF/Aly, Y14, SRm160, UAP56, RNPS1, and Magoh) and U snRNPs colocalise with factories transcribing a virally-encoded human beta-globin gene. Mutant pre-mRNAs defective in splicing and 3'end processing do not colocalise with SRm160, REF, UAP56, or Sm proteins, suggesting they are transcribed in different factories from the wild type genes (Custodio et al., 2004).

Specialised factories can play an important role in regulating the selective transcription of one of many similar genes. African trypanosomes express one of many variant surface glycoproteins, by incorporating the relevant allele into a special pol I factory in the nucleoplasm (Navarro and Gull, 2001). In each olfactory neuron, the olfactory receptor (OR) genes on different chromosomes compete to associate with an enhancer on Chr14, ensuring only a single OR gene is activated (Lomvardas et al., 2006). Furthermore, there seems to be a cell-type-specific dynamic interaction between interacting chromatin partners. The *Ifng* gene on mouse Chr10 interacts with the *Il5* cytokine promoter, *Rad50* promoter and RHS6 DNase I hypersensitive site of the T_H2 LCR, all located in the T_H2 cytokine gene locus on mouse Chr11. The interactions are extremely strong in naive $CD4^+$ T cells but are greatly reduced after the differentiation of naive T cells to effector T_H1 or T_H2 cells (Spilianakis et al., 2005). A similar mechanism may also contribute to imprinting control. Ifg2/H19 ICR on Chr7 colocalises with one allele of Wsb1/Nf1 on Chr11; deletion of the maternal ICR abolishes the association (Ling et al., 2006).

In this thesis, I aim to analyse factories involved in transcription of specified groups of genes. I developed a probe for active transcription factories based on transcriptionally-active minichromosomes. Previous work demonstrated that plasmids bearing a β-globin transcription unit colocalise with endogenous β-globin loci (Ashe et al., 1997). Maul reviewed that dsDNA viruses are replicated and transcribed in nuclei in active domains containing various components involved in transcription and processing (Maul, 1998). Since high inputs yield many transcriptionally-inert plasmids (Jackson and Cook, 1993) that provide an obscuring background, I used a low-input replicating system; then, replication increases copy

number well above the input so that we can be confident that most of the plasmids analysed represent minichromosomes and not the inactive input.

2 Materials and Methods

2.1 Introduction

The work described in this thesis includes two parts: 1) detecting specific transcription sites using replicating minichromosomes in COS-7 cells; 2) visualising a transcription wave sweeping along a long gene in HUVEC cells.

Part one includes the following tasks: 1) establishing a replicating minichromosome based method to detect factories transcribing different groups of genes; 2) quantifying and visualising plasmid DNA after transfection; 3) labelling and visualising plasmid transcription sites using RNA-FISH; 4) detecting proximity of different plasmids and endogenous genes using 3C;

Part two includes the following tasks: 1) labelling and evaluating oligo-nucleotide probes; 2) visualising the transcription wave in HUVEC cells.

The materials and methods used in both parts are described below. In addition, I started a study on mouse *globin* in different tissues. The methods are included here, but the incomplete results will not be shown.

2.2 Chemicals and kits

The chemicals used in the study are listed below. Unlisted ionic salts, acids, bases and organic solvents were from Sigma or VWR International.

Chemical	Manufacturer and catalogue number
Agarose	Bio-Rad 161-3102
Alexa Fluor 546 protein labelling kit	Invitrogen A10237
Alexa fluor 488 reactive dye	Invitrogen A32750
ATP	Roche 11093070910
Biotin-16-dUTP	Roche 11093070910
Biotin-16-ddUTP	Roche 1427598
Biotin Nick translation mix	Roche 11745824910
Blocking reagent	Roche 1112589
BSA Sigma fraction V	Sigma A3294
Br-UTP	Sigma B-7166
COT Human DNA	Roche 1581 074
Cy3 mono reactive dye pack	Amersham Pa 230001
DAPI (4',6-diamidino-2-phenylindole)	Sigma D9542
Deionization resin	Bio-Rad AG 501-x8(D)
Deoxynucleoside Triphosphate Set	Roche 1969064
DEPC (Diethyl pyrocarbonate)	Sigma D5758
DIG-11-dUTP	Roche 11573152910
DIG-11-UTP	Roche 03359247910
DIG DNA labelling and detection kit	Roche 11093657910
DIG nick-translation mix	Roche 11745816910
DNA marker	Fermentas SM0333
DNase 1	Roche 10776785001
Dimethyl sulfoxide	Sigma D8418
DMEM (Dulbecco's Modified Eagle's Medium)	GIBCO BRL 41966
Dithiothreitol (DTT)	Sigma D-5545
Expand high fidelity PCR system	Roche 03300242001
Ficoll 70	Sigma F2878
Fetal Bovine Serum	PAA A15-101
Formaldehyde	MERCK cat no.3999
Formamide	GIBCO BRL 5515UA
FuGene 6	Roche 11814443001
FISH tag RNA multicolour kit	Invitrogen F32956
GoTaq Green master mix	Promega M7113
Gelatin from Porcine skin	Sigma G2500
Kanamycin	Sigma K1377
Lipofectamine	Invitrogen 18292-011
L-glutamate	GIBCO 25030-024
Glycerol	Sigma G5516

Glycine	**Sigma G7126**
Magnesium Chloride	**VWR International 10494V**
Micro-spin G-50 columns	**Amersham Biosciences 27533001**
Microcon	**Millipore**
Hybond-N+ membrane	**Amersham RPN-303B**
NP-40 alternative	**CaliBioschem 492016**
HEPES (4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid)	**GIBCO 15630-122**
Maleic acid	**Sigma M0375**
Paraformaldehyde (PFA)	**Electron Microscopy Sciences 15710-S**
PCR purification kit	**QIAGEN 28006**
Phosphate Buffered Saline	**Sigma P4417**
Phenol:chloroform:isoamyl alcohol 25:24:1	**Sigma P2069**
Penicillin-Streptomycin-Glutamine	**GIBCO 10378-016**
Poly-L-lysine	**Sigma P2636**
Pepsin	**Sigma P7000**
Platinum SYBR Green qPCR supermix-UDG	**Invitrogen 11733-038**
Polyvinylpyrolidone	**Sigma PVP-40T**
Protease K	**Sigma P7626**
Protease inhibitor cocktail	**Sigma P2714**
Protein G Sepharose beads	**GE healthcare 17-0618-01**
QIAprep Spin Miniprep Kit	**QIAGEN 27106**
QIAfilter Plasmid Maxi Kit	**QIAGEN 12263**
RNase A	**Roche 10109134001**
RNase ZAP	**Sigma R2020**
RNase inhibitor	**Roche 03335399011**
Salmon-sperm DNA	**Sigma D1626**
Saponin	**Sigma S4521**
Sephadex G-50 columns	**GE healthcare 27-5330-01**
SDS (20% solution)	**Bio-Rad 161-0418**
SP6 polymerase	**Roche 10810274001**
Sucrose	**Sigma S-7903**
Superscript II platinum SYBR green one step qPCR system	**Invitrogen 11736-051**
T4 DNA ligase	**New England Biolabs (NEB)**
T7 RNA polymerase	**Roche 10881767001**
Terminal transferase	**Roche 03333566001**
Taq polymerase	**Roche 1146165**
Trizol	**Invitrogen 15596-018**

Tris-HCl buffer (1.5M; pH 8.8)	**Bio-Rad 161-0798**
Tris-HCl buffer (0.5M; pH 6.8)	**Bio-Rad 161-0799**
Triton X-100	**Sigma T-8787**
Trypan blue	**Sigma T8154**
Tween-20	**Sigma 63518**
Trypsin-EDTA	**GIBCO 25200-056**
ULS nucleoacid labelling kit	**Invitrogen U21650**
VectaShield	**Vector laboratories H-1000**
H2O	**Millipore**

2.3 Antibodies

The antibodies used in the study are listed below.

Mouse monoclonal antibody directed against:
biotin (Jackson ImmunoResearch 200 002 096)
Br-U (Caltag MD5100)
UBF (F-9) (Santa-Cruz SC-13125)
SC-35 (Sigma S-4045)
Sheep antibody directed against:
DIG (Roche 1 333 089)
anti-DIG-alkaline phosphatase (Roche 11093274910)
Rabbit antibody directed against:
H3 tri-methyl-K4 (Abcam ab8580)
pol II (Santa Cruz sc-899X)
Secondary antibodies directed against species-specific immunoglobulins:
Cy3 donkey anti-sheep IgG, H+L (Jackson ImmunoResearch 713 165 147)
Cy3 donkey anti-mouse IgG, H+L (Jackson ImmunoResearch 715 165 150)
Cy5 donkey anti-mouse IgG, H+L (Jackson ImmunoResearch 715 175 150)
Alexa 350 donkey anti-sheep IgG, H+L (Invitrogen A21097)
Alexa 488 donkey anti-mouse IgG, H+L (Invitrogen A21202)
Streptavidin A488 (Invitrogen S11223)
mouse normal IgG (Santa-Cruz SC-2035)
Sheep normal IgG (Santa-Cruz SC-2717)

Unconjugated donkey anti-mouse IgG, H+L (Jackson ImmunoResearch 715 005 151) was labelled using Alexa 647 protein labelling kit (Invitrogen A20173) according to the manufacturer's instructions.

2.4 Cell cultures

COS-7 cells (Mellon et al., 1981) were obtained from the Cell Bank of The Sir William Dunn School of Pathology. Cells (kept in 1:9 DMSO:serum, -80ºC) were recovered to 37°C, washed in fresh DMEM + 10% foetal calf serum, and grown on Petri dishes (Nunc), flasks (Corning), 6 well plates (Costar) or coverslips (VWR international) for individual experiments. For subculture, cells were detached with 0.25% trypsin-EDTA. Each batch of cells was kept for no more than 20 generations. To inhibit pol II transcription, cells were incubated with α-amanitin (50 μg/ml) overnight at 37°C.

HUVEC cells fixed on coverslips were obtained directly from Professor Tatsuhiko Kodama at Tokyo University.

18.5 day fetal liver and brain cells from mice (bred by The Sir William Dunn School of Pathology animal facility) were harvested.

2.5 Buffers and solutions

All buffers and solutions were prepared in MilliQ water and autoclaved unless otherwise stated. pH was adjusted with 5M NaOH or 40% HCl.

2.5.1 Buffers for general molecular and cell biology

2.5.1.1 Buffers for transformation

Transformated cells were recovered in **SOC medium**, which contains 2.0 g Bacto-tryptone, 0.5 g Bacto-yeast extract in 100 ml 0.01 M NaCl, 2.5 mM KCl. After autoclave, 1 ml 2 M Mg^{2+} stock (1 M $MgCl_2$, 1 M $MgSO_4$, filtered) and 1 ml 2 M glucose was added to the medium, pH adjusted to 7.0, and the solution filtered with 0.2 μm membrane (Millipore).

Cells transformed with plasmids were selected using **LB plates containing kanamycin**. LB was made by dissolving 5 g Bacto-Tryptone, 2.5 g Bacto-Yeast Extract, 2.5 g NaCl, 7.5 g Bacto-Agar in 500 ml H2O at pH 7.5. After autoclaving, the medium was cooled to 60ºC before adding kanamycin to 50 μg/ml.

Nucleic acids were generally stored in **TE buffer**, which contains 10 mM Tris•Cl (made in sterile H_2O; this cannot be autoclaved) and 1 mM EDTA at pH 7.5.

2.5.1.2 Buffers for gel electrophoresis

DNA and RNA were resolved in 0.5-2% agarose gels in either 1 x TAE or 0.5 x TBE buffer. **10 x TAE** contains Tris-base 48.4 g, glacial acetic acid 10.9 g, EDTA 2.92 g in 1 L H_2O. **10 x TBE** contains Tris-base 54.0 g, boric acid 27.5 g, EDTA 2.92 g in 1 L H2O.

2.5.1.3 Buffers for general cell biology

Phosphate buffered saline (PBS) was used for washing cells. It was made from PBS tablets containing 137 mM NaCl, 2.7 mM KCl and 0.01 M phosphate buffer (8 mM Na_2HPO_4 and 1.5 mM KH_2PO_4).

2.5.2 Buffers for 'Southern' blot

Depurination buffer was made from 11 ml HCl in 989 ml H_2O.

After agarose gel electrophoresis, DNA was denatured in **Denaturation buffer** containing 0.5 M NaOH and 1.5 M NaCl.

After denaturation, agarose gel was incubated in **Neutralisation buffer** containing 0.5 M Tris·Cl and 1.5 M NaCl at pH 7.5.

SSC buffer was used during hybridisation and subsequent washes. **20 x SSC** stock contains 0.3 M tri-sodium citrate and 3 M NaCl at pH 7.0.

'Southern' hybridisation was carried out in **Church buffer** containing 0.5 M sodium phosphate buffer (pH 7.2), 1 mM EDTA (pH 8.0), 1% BSA and 7% SDS.

Maleic buffer contains 1 M maleic acid, 1.5 M NaCl at pH 7.5.

10 x Blocking solution was made by dissolving 10 g block powder in 100 ml maleic buffer. Blocking solution was freshly prepared for every experiment by diluting the stock solution with maleic buffer.

Antibody solution: Anti-DIG-alkaline phosphatase was centrifuged (10,000 rpm; 4ºC, 5 min), and supernatant diluted $1:10^4$ with blocking solution.

Nucleic acid dilution buffer: Labelled DNA probes were diluted in a buffer containing 50 μg/ml salmon sperm DNA in TE before applied to nylon membrane.

After transfer, membrane was washed in **Washing buffer** containing 1 M maleic acid, 1.5 M NaCl and 3% Tween 20 at pH 7.5.

Membrane was washed and incubated in **Detection buffer** with colour substrate for detection. The buffer contains 1 M Tris•Cl , 1 M NaCl at pH 9.5.

2.5.3 Buffers for DNA and RNA FISH

All buffers for RNA FISH were treated with DEPC (20ºC, 16 h) and autoclaved unless otherwise stated.

1 mM phosphate buffer was made by mixing 39 ml 1 M KH_2PO_4 and 61 ml 1 M $KHPO_4 \cdot 2H_2O$, before adjusting the pH to 7.0.

50x Denhardt's contains 1% Ficoll 400, 1% polyvinylpyrolidone and 1% bovine serum albumin in DEPC-treated H_2O, sterilized by filtering, stored at -20ºC.

Hybridization mixture contains 50 ng tagged probe, 25% deionised formamide, 2 x SSC, 200 ng/µl sheared salmon sperm DNA, 5x Denhardt's, 50 mM phosphate buffer and 1 mM EDTA.

Tris/saline (TS) buffer was made from 0.15 M NaCl, 0.1 M Tris•Cl at pH 7.5 in DEPC-treated H_2O.

After hybridization, **Tris/ Saline/ Tween (TST) buffer** was used to wash away unbounded probes. TST buffer contains 0.15 M NaCl, 0.1 M Tris•Cl and 0.05% Tween 20 at pH 7.5 in DEPC-treated H_2O.

Blocking reagent 20x stock was made by dissolving 27 g blocking power in l00 ml in H_2O (27% solution), and 1 ml aliquots stored at -20ºC.

2.5.4 Buffers for 3C

2.5.4.1 Buffers for conventional 3C

Cells were lysed in **Lysis buffer**, which contains 10 mM Tris·Cl (pH 8.0), 10 mM NaCl, and 0.2% NP-40. Protease inhibitors (0.1 mM PMSF, 1:500 protease inhibitor cocktail) were added.

Digested chromatin was diluted and ligated in **Ligation buffer**, freshly made with 30 mM Tris·Cl (pH 8), 10 mM $MgCl_2$, 10 mM DTT and 1 mM ATP.

2.5.4.2 Buffers for ChIP-3C

Cells were lysed in **Hypotonic buffer** containing 10 mM Tris·Cl (pH 7.2), 2 mM $MgCl_2$ and 0.5% Triton X-100 2.5 ml.

Restriction wash buffer differs according to the specific restriction enzyme used. For *Bam*HI and *Pst*I, the **high stringency buffer** contains 50 mM Tris·Cl (pH 8.1), 500 mM NaCl, and 10 mM $MgCl_2$; the **low stringency buffer** contains 50 mM Tris·Cl (pH 8.1), 150 mM NaCl, and 10 mM $MgCl_2$.

Before ChIP, cells were permeabilised in **CSK buffer** containing 100 mM NaCl, 300 mM Sucrose, 10 mM PIPES (pH 6.8), 3 mM $MgCl_2$, 10 μM leupeptin, 1 mM EGTA, 1.2 mM PMSF and 0.5% Triton X-100.

2.5.5 Buffers for run-on

Physiological buffer was used to mimic the cellular environment, and label sites of transcription with Br-UTP at an optimal condition. Before run-on, cells were incubated on ice in physiological buffer containing 100 mM potassium acetate, 30 mM KCl, 10 mM Na_2HPO_4, 1 mM $MgCl_2$, 1 mM Na_2ATP, 1 mM DTT, and 50 mg/ml Ficoll, type 70 at pH 7.4 (Jackson et al., 1993).

Before run-on, cells were permeabilised in **Permeabilisation buffer**, a physiological buffer containing an optimised amount of saponin for each cell line. For COS-7 cells, a concentration of 70 μg/ml was used.

The concentration of saponin was determined by applying a series dilution of saponin to cells plated on 24 well plates, and stained with trypan-blue. The highest concentration excluding trypan-blue was selected.

Run-on was performed in a **Transcription buffer** containing 100 μM ATP, CTP, GTP, Br-UTP and 400 μM $MgCl_2$ at pH 7.4.

2.6 Conventional molecular biology

Conventional molecular biology methods were performed according to (Sambrook and Russell, 2001) and the manufacturer's instructions.

2.6.1 Cloning

Restriction enzymes and buffers were obtained from NEB. Sticky and blunt end ligations were carried out using T4 ligase with its accompanying buffer from NEB. Transformation was carried out using JM-109 competent cells (Promega L1001). Plasmids were isolated from bacteria using QIAprep Spin Miniprep kit or QIAFilter Maxiprep kit. The concentration of nucleic acids was measured by spectrometry using a Nanodrop ND-100 spectrometer (Nanodrop Technologies).

2.6.2 Gel electrophoresis

DNA and RNA were resolved in agarose gels in the appropriate concentration in 1 x TAE or 0.5 x TBE buffer. Gels were subsequently stained with 0.5 μg/ml ethidium bromide. In some cases, DNA was analysed by E-Gel Pre-Cast gel system (Invitrogen) using E-gel 48 gels and low molecular weight DNA ladders.

2.6.3 PCR

The content of the standard PCR mix is as follows: 5 μl 10x amplification buffer (containing 1.5 mM $MgCl_2$), 1 μl 20 mM solution of dNTPs, 2.5 μl 20 mM forward primer, 2.5 μl 20 mM reverse primer, 1 U *Taq* polymerase, 1-5 μl DNA template, with H_2O added to a total of 50 μl. In some cases, GoTaq Green master mix was used. Primers were obtained from MWG International or Sigma.

PCR reactions were carried out using a DNA Engine DYAD (MJ research). Standard PCR cycles are: 94ºC 3 min, followed by 30 cycles of 94ºC 45 sec, Ta (appropriate annealing temperature) 45 sec, 72ºC 1 min. Then 94ºC 1 min, Ta 1 min, 72ºC 10 min, and stored at 4ºC. Table 2.1 listed the primers and annealing temperature for each individual reaction.

Table 2.1. PCR primers and annealing temperatures

Product	Primers	Ta (ºC)	Cycles
Cloning			
pol I	**5' GCGCGCATTAATATTACGCCAAGCTTACGG**	**57.2**	**30**
	5' GCGCGCTAGCTTGTTTCTACTAATAACCCG		
HBB	**5' GGGAAGCTTCCTGAGAACTTCAGGGTGAGTC**	**52**	**30**
intron 2	**5' GGGCTGCAGAAAGTGATGGGCCAGCACAC**		
U2G	**5' GCGCAAGCTTTGTCTTTTCGCTGGCTCG**	**55**	**30**
	5' GGGCTGCAGCGCGTCACAGGACTCGTG		
Probes			
EGFP **(nick**	**5' CGTAAACGGCCACAAGTTC**	**57.5**	**30**
translation)	**5' TCTCGTTGGGGTCTTTGCTC**		
DsRed **(nick**	**5' CGAGTTCATGCGCTTCAAGG**	**57.5**	**30**
translation)	**5' TGCTCCACGATGGTGTAGTC**		
Neo^r **(nick**	**5' GGATCGTTTCGCATGATT**	**58.3**	**30**
translation)	**5' GAGTCCCGCTCAGAAGAA**		
U2G **(nick**	**5' ATCGCTTCTCGGCCTTTTG**	**55**	**30**
translation)	**5' GGAGGTACTGCAATACCAGG**		
Read-	**5' CCGGATCAAGAGCTACCAAC**	**56**	**30**
through	**5' GGGATAACGCAGGAAAGAAC**		

2.6.4 Purification of nucleic acids

PCR products were purified using a PCR purification kit (QIAGEN). The length of nucleic acids was limited to between 100 bp and 7 kb, and 10 μg was applied per column.

Phenol extraction. When QIAGEN columns could not be used, nucleic acids were purified from protein using phenol extraction. An equal volume of phenol:chloroform:isoamyl alcohol 25:24:1 (pH 7.8) was added to the solution, before vortexing until an emulsion formed. The mixture were centrifuged (13,000 rpm; 20ºC; 1 min), and aqueous phase transferred to a fresh tube. The same steps were repeated until no protein was visible at the interface. An equal volume of chloroform were added to the aqueous phase, mixed and isolated by centrifugation as described above to remove phenol.

Ethanol precipitation. 1/10 volume of 3 M sodium acetate and 2 ½ volume of ice-cold ethanol were added to solution. Samples were stored at -80ºC for 30 min before centrifuged (14,000 rpm; 0ºC; 30 min). The precipitates were washed with 70% ethanol and centrifuged to remove supernatant (14,000 rpm; 0ºC; 10 min). The nucleic acid precipitates were dried and resuspend in H_2O to a desired concentration.

2.7 Construction of plasmids

Plasmids with a common backbone but encoding transcription units with different promoters, introns, and 3' ends were constructed (Fig 3.4). The SV40 origin of replication (*ori)* in the backbone allows plasmids to replicate in COS-7 cells (which express SV40 T antigen); *ori* also encodes the SV40 early promoter (Mellon et al., 1981). The naming convention is illustrated by reference to plasmid

II^{CMV},*EGFP*$_V$,pA which possess the pol II CMV promoter upstream of *EGFP* followed by the (β-*globin*) intron and a poly(A) signal, while plasmid 0,*EGFP*,pA lacks a promoter and intron. Plasmids were constructed as follows: 0,*EGFP*,pA from pEGFP-N1 (Clontech) by replacing base pairs 8-591 with a promoter linker (sense 5'[AC12]GGGGGGATTAATGGGGGGGCTAGCGGG; antisense 5'[AC12]CCCGCTAGCCCCCCCATTAATCCCCCC; where [AC12]C is 5' amine-modifier C12); I^{45S},*EGFP*,pA by inserting the 250 bp truncated human pol I promoter (Pleschka et al., 1996) between the *Ase*I and *Nhe*I sites of 0,*EGFP*,pA; III^{7SK},*EGFP*,pA by inserting the 234 bp human 7SK promoter (Boyd et al., 1995) between the *Pst*I and *Kpn*I sites of 0,*EGFP*,pA; II^{U2},*U2G*,3' box and II^{U2p-},*U2G*,3'box from pEGPF-C1 (Clontech) by deleting base pairs 8-1367 and inserting the 623 bp *U2G* fragment (Medlin et al., 2003) or the promoter-less *U2G* fragment (Cuello et al., 1999) between the *Ase*I and *Pst*I sites; II^{CMV},*EGFP*$_V$,pA by inserting 906 bp of the β-globin intron 2 (HBB, *Homo sapiens*; chr11: positions 5,203,482 – 5,204,406 in NCBI 36 assembly; prepared by PCR amplification of human DNA) between the *Hind*III and *Pst*I sites of pEGPF-C1. Analogous constructs encoding DsRed were made from pDsRed-N1 (Clontech). Appropriate construction was confirmed by sequencing. Two primers on the common backbone were used for sequencing: Fwd 5' TAACCGTATTACCGCCATGC and Rev 5' TCCGTA TGTTGCATCACCTTC.

2.8 Transfection

Initially, COS-7 cells were transfected with Lipofectamine or FuGene6 according to the manufacturer's instructions. A comparison of the two transfection

reagents and an optimisation of FuGene6 transfection are described in Chapter 3.3.1. Transfection efficiencies were determined by measuring EGFP levels by microscopy and FACS. An optimised FuGene6 transfection protocol was used for all future experiments.

For Lipofectamine, 2 x 10^5 cells per well (for 6-well plates) were plated in DMEM + 10% FCS without antibiotics one day before transfection. For each sample, a complex of the following was prepared and incubated (20ºC; 15 min): 1) 1 µg of plasmid DNA plus 20 µg salmon sperm DNA in 100 µl of DMEM (without serum); 2) 5 µl of Lipofectamine in 100 µl of DMEM (20ºC; 45 min). Cells were then washed with 2 ml DMEM, and the prepared complex plus 0.8 ml DMEM were added. 6 h after transfection, medium was replaced with 2 ml of DMEM + 10% FCS.

For FuGene6, 2 x 10^5 cells per well (for 6-well plates) or 1 x 10^6 (for 90 mm dishes) were plated in DMEM + 10% FCS without antibiotics one day before transfection. For each well, 6 µl FuGene6 were diluted in 100 µl DMEM, vortexed (1 sec), and incubated (20ºC; 5 min) before the addition of 0.2 µg of plasmid DNA (or 1.2 µg per 90 mm dish) plus 0.8 µg salmon sperm DNA (or 4.8 µg per 90 mm dish). The complex was incubated (20ºC; 45 min) before adding to cells.

2.9 Preparation of bacteria artificial chromosomes (BACs)

BACs (listed below) were obtained from BAC PAC resources, Children's Hospital Oakland Research Institute (CHORI). Upon receipt, DH10 *E.coli* culture was streaked to single colonies on a LB agar plate containing 12.5 µg/ml chloramphenicol. A glycerol stock was prepared for each BAC and stored at -80°C. Colonies were then checked by PCR, using primers listed in Table 2.2.

Table 2.2. BAC clones and their PCR primers

RP23-325M1	**5'AATGTGGTGAGGAGGTTGGG**
	5'AAAGGCAGACATAGATGCGG
RP24-170L4	**5'TGTCAGCTAACCTGAATCCC**
	5'TCCAAGAGGATTTCCACCAC
RP24-344M21	**5'CACTCAAAGTTAGTGCCTGC**
	5'CAACCCCAGAAACAGACATC
RP23-222I22	**5'CCTCTTGCTCTAAATGCCTG**
	5'TGCTTTTGATCGTGTGCTTC
RP24-104I11	**5'ATCCCACCACCAAGCCATAC**
	5'ACACCTTTTATCCCAGCACC

2.9.1 DNA isolation of BAC clones

BACs were isolated using a method modified from the conventional method based on alkaline lysis. A single colony was grown (37^{o}C; 16 h) in 2 ml LB containing 12.5 μg/ml chloramphenicol. After centrifugation (3,000 rpm; 20^{o}C; 10 min), the pellet was resuspend in 0.3 ml 50 mM Tris (pH 8), 10 mM EDTA, 100 μg/ml RNase A). 0.3 ml 0.2 N NaOH, 1% SDS was added and mixed (20^{o}C; 5 min). 0.3 ml 3 M potassium acetate (pH 5.5) was added to the solution with gentle shaking, and incubated on ice (5 min). The protein and genomic DNA was removed by centrifugation (10,000 rpm; 4^{o}C; 10 min). The pellet containing the BAC was precipitated using isopropanol, followed by 70% ethanol wash, and resuspend in 40 μl TE.

2.10 Assessing copy number by 'Southern' blot

2.10.1 Hirt extraction

8 h after transfection, cells were trypsinised, washed and replated; trypsinisation and washing minimize the background of input DNA. Cells were then lysed (20ºC; 20 min) by adding 1 ml of 0.6% SDS (pH 7.5) containing 0.01 M EDTA. The viscous lysate was scraped from the Petri dish with a disposable cell lifter (Fisher Scientific) and collected into a 2 ml tube. Then, 0.375 ml 5 M NaCl was added to the lysate to make a final concentration of 1 M, mixed by gently inverting the tube 10 times. Samples were stored at 4ºC for 8 h before centrifugation at 17,000 g (4ºC; 30 min) to remove SDS, protein, and genomic DNA. The supernatant with plasmid DNA and RNA were treated with 50 μg/ml RNase (37ºC; 1 h). Plasmid DNA was phenol extracted, and ethanol precipitated.

2.10.2 Preparation of probes

Pure plasmids were labelled using a DIG-labelling kit by random-priming (Feinberg and Vogelstein, 1983), using an adaptation of the manufacturer's procedure. 300 ng plasmid DNA was linearised (*Eco*RI; 37ºC; 1 h), denatured (95ºC; 10 min) and chilled on ice before incubation (37ºC; 16 h) with hexanucleotides, dNTPs and 'Klenow' enzyme.

The efficiency of labelling the probes were evaluated by dot-blot hybridisation (Kafatos et al., 1979) followed by direct detection using the NBT (Nitro-Blue Tetrazolium Chloride)/BCIP (5-Bromo-4-Chloro-3'-Indolyphosphate p-Toluidine Salt) system (Roche). Labelled probes were serially diluted and dotted on to a nylon membrane (Amersham Biosciences). DNA was crosslinked to the

membrane by baking (120ºC; 30 min). Membrane was then washed in washing buffer (20ºC; 2 min), and incubated sequentially in blocking solution (20ºC; 30 min) and antibody solution (20ºC; 30 min), before washing twice (20ºC; 15 min) and equilibration in detection buffer (20ºC; 3 min). Membrane was then incubated in the dark (20ºC; 16 h) in colour substrate diluted in detection buffer, before direct visualisation.

2.10.3 'Southern' blotting

The Hirt extract was digested with *Eco*RI (37ºC; 1 h), and diluted into 0.2 pg – 2 ng per µl before loading on a 1% agarose gel. *Eco*RI-digested pEGFP-C1 at 200 pg/µl was used as a marker. The gel was then incubated in denaturation buffer (20ºC; 15 min twice), rinsed with H_2O and neutralisation buffer (20ºC; 15 min twice) and equilibrated in 20 x SSC (20ºC; 15 min). DNA was transferred onto nylon membrane using the capillary transfer method (Southern, 1975). Membrane was then washed in 2x SSC, nucleic acid fixed by baking (120ºC; 30 min), and pre-hybridised at the hybridisation temperature (67 ºC; 30 min) in Church buffer. 25 ng random-primed probes were denatured and added to the Church buffer. Hybridisation was carried out in a Micro-4 (Hybaid) hybridisation oven (67ºC; 16 h). Membrane was washed in 2x SSC containing 0.1% SDS (twice 20ºC; 5 min) and incubated in high-stringency buffer containing 0.1 x SSC and 0.1% SDS (68ºC; twice 15 min). Direct detection using NBT/BCIP was then performed as described in 2.10.2.

2.11 Assessing copy number by real-time PCR

2.11.1 Extraction of total DNA and RNA

Total DNA and RNA (genomic and plasmid) were isolated using Trizol according to the manufacturer's instructions. Samples were further purified using phenol extraction and ethanol precipitation.

2.11.2 Standard curve

A standard curve was first generated using known amounts of pure plasmid DNA (Fig. 4.1.A). A standard curve for *GAPDH* was made using a known number of cells. The amount of plasmid was first determined by spectrometry. A series dilution of the standard plasmid DNA was added to cells, before total DNA was extracted, and copy number determined by real-time PCR using *EGFP* primers.

For RNA copy number, a standard curve was generated similarly using known amounts of pure *EGFP* RNA prepared by *in vitro* transcription. The transcript level of different plasmids was evaluated using quantitative reverse transcription-PCR (qRT-PCR), normalised to the *GAPDH* transcript level in each sample; transfection efficiencies were determined by FACS or microscopy using DsRed fluorescence.

2.11.3 DNA copy number determined by real-time PCR

Total DNA from ~10^7 cells was purified, and quantitative PCR (qPCR) performed using a Roter-Gene 3000 (Corbett Life Science; Sydney, Australia); signals were normalised relative to those given by *GAPDH*. Briefly, each 25 µl reaction contained Platinum SYBR Green qPCR SuperMix-UDG w/ROX, and 5

pmol of each primer as in Table 2.3. Cycling conditions were: 50°C for 30 min, 95°C for 10 min, followed by 40 cycles at 94°C for 15 s and 57.5°C for 30 s. Samples were run in triplicate and each experiment was repeated three times.

Absolute copy numbers were obtained by reference to standard curves. Concentration and purity were determined from the absorbance.

Table 2.3. Real-time PCR primers

EGFP	**5' ACGTAAACGGCCACAAGTTC**
	5' AAGTCGTGCTGCTTCATGTG
GAPDH	**5' GAAGGTGAAGGTCGGAGTC**
	5' GAAGATGGTGATGGGATTTC.

2.11.4 RNA copy number determined by qRT-PCR

2.11.4.1 *In vitro* transcription

EGFP RNA standards for real-time PCR were generated by *in vitro* transcription using T7 RNA polymerase. The template (pcDNA-EGFP-N1) was constructed by inserting *EGFP* (from pEGFP-N1) between the *Hind*III and *Bam*HI sites into pcDNA3 (Invitrogen). The integrity and size of RNA were verified by gel electrophoresis, and concentration and purity determined from the absorbance.

2.11.4.2 Quantitative reverse-transcription PCR (qRT-PCR)

Total RNA was purified, cDNA made using the SuperScript III Platinum One-Step System for RT-PCR, and normalized to *GAPDH* mRNA levels in the same sample. qRT-PCR was conducted similarly to qPCR (above). Absolute copy numbers were obtained by reference to a standard curve.

2.12 Fluorescence *in situ* hybridisation (FISH)

2.12.1 DNA FISH

The DNA FISH protocol was adapted from (Brown, 2002).

2.12.1.1 Preparation of probes

DNA FISH probes for plasmids were prepared by random priming using a DIG-labelling kit as described in Ch.2.10.2, purified using Sephadex G-50 columns, ethanol precipitated, and resuspended in water at a concentration of 0.1 μg/μl.

DNA FISH probes for the mouse *globin* region were prepared by nick translation using a DIG- or biotin-nick translation kit according to the manufacturer's instructions. BACs were incubated (15°C; 3 h, 20 min) in the mix, and chilled on ice until the size of products (determined by gel electrophoresis) reached a median of 150 bp. Labelled probes were purified as above.

2.12.1.2 Preparation of coverslips/slides

24 h prior to hybridisation, COS-7 cells (grown on coverslips in 6 well plates) were transfected with the appropriate plasmids.

18.5 day fetal liver and brain was harvested, dissected and transferred to 100 μl cold PBS, disrupted by aspiration and passed through a 70 μm sieve to ensure single cell suspension. Cells were then washed in PBS, incubated in 75 mM KCl at 7 x 10^4/ml on ice (20 min). 500 μl cells were centrifuged onto poly-L-lysine coated slides (VWR International) by using Cytospin 3 (Shandon, Pittsburgh; 2000 rpm; 20°C; 10 min). Slides were incubated immediately (10 min) in KCM buffer (10 mM Tris·Cl, pH 7.7; 120 mM KCl, 20 mM NaCl, 0.1% Triton X-100) and washed twice in PBS.

2.12.1.3 Fixation and permeabilisation

Cells were rinsed briefly in PBS, fixed (17 min; 20°C) in 4% paraformaldehyde (PFA), 0.05% acetic acid in 0.15 M NaCl, washed three times (5 min; 20°C) in PBS, incubated (30 min; 37°C) with 50 µg/ml RNase, permeabilised (5 min; 37°C) in 0.01% pepsin (pH 2.0), rinsed in water, postfixed (5 min; 20°C) in 4% PFA in PBS, and washed (10 min; 20°C) in PBS.

2.12.1.4 Hybridisation

Cellular and plasmid DNA was denatured (2 min; 70°C) in 70% deionised formamide, 2x SSC, and immediately dehydrated (3 min) by passage through 70%, 90%, and then 100% ice-cold ethanol, and coverslips sealed with rubber cement (Marubu, Tamm, Germany) prior to hybridisation (16 h, 37°C) in a moist chamber. 50 ng tagged probe were mixed with hybridisation mix, and denatured (10 min; 94°C) prior to hybridization.

2.12.1.5 Detection

Cells were then washed three times in 2x SSC (15 min; 37°C), twice (5 min; 20°C) in TST, and incubated in 1.4% blocking reagent in TS. Immuno-detection was carried out using sheep anti-DIG (60 min; 20°C; 1:250), followed by donkey anti-sheep Cy3 (30 min; 20°C; 1:250), or using mouse anti-biotin (30 min; 20°C; 1:250) followed by donkey anti-mouse Alexa 488 (30 min; 20°C; 1:200); cells were washed 3x (5 min; 20°C) in PBS after incubation with each antibody, and mounted in Vectashield with .1 µg/ml DAPI.

2.12.2 RNA FISH

2.12.2.1 Preparation of probes

A selection of techniques was used to prepare FISH probes. For each individual target, the choice of labelling methods was optimised. A comparison of different labelling methods is discussed in Ch.4.4.1.1.

2.12.2.1.1 Nick translation

Most RNA FISH probes were prepared by nick translation. Target DNA sequences were amplified by PCR (see Ch.2.6.3), purified by gel extraction, PCR performed again using the purified product as a template, and further purified as described in Ch.2.6.4. Nick translation was carried out using a DIG- or biotin-nick translation kit. In most cases, 1 kb template was incubated (15°C; 90 min) in the mix, and chilled on ice while the size of products was determined by gel electrophoresis. The reaction was stopped if a median of 150 bp was achieved. Labelled probes were purified using Sephadex G-50 spin columns (GE Biosciences), ethanol precipitated, and resuspended in DEPC-treated water at a concentration of 0.1 μg/μl.

2.12.2.1.2 Asymmetrical PCR

Three ~200 bp fragments for *EGFP* or *DsRed* probes were synthesized using touchdown PCR (Don et al., 1991), followed by asymmetrical PCR (Finckh et al., 1991). The PCR conditions were as follows: 1) Touchdown PCR: 94ºC 3 min, followed by one cycle of 94ºC 45 sec, 61ºC (annealing temperature, 3ºC higher than the calculated Ta) 45 sec, 72ºC 1 min 30 sec. For the subsequent 14 cycles, decrease the annealing temperature by 1ºC per cycle. Then, another 30 cycles with: 94ºC 45 sec, 47ºC 45 sec, 72ºC 1 min 30 sec. The final product was stored at 4ºC. 2)

Asymmetrical PCR: Use 5 ng of the product from touchdown PCR as the template for asymmetrical PCR. The content of the PCR mix is as follows: 5 µl 10 x amplification buffer (containing 1.5 mM $MgCl_2$), 11.3 µl DIG or biotin labelling mix (20 mM dATP, 20 mM dGTP, 20 mM dCTP, 13 mM dTTP, 7 mM DIG-dUTP or biotin-dUTP), 2.5 µl 20 mM reverse primer only, 1 U *Taq* polymerase (Roche), 1-5 µl DNA template, add H_2O to a total of 50 µl. Table 2.4 lists the primers for *EGFP* and *DsRed* 200 bp probes.

Table 2.4. Touchdown PCR primers

EGFP **(asymmetrical PCR)**	**Pair 1**	**5' CGTAAACGGCCACAAGTTC**
		5' TTGAAGAAGTCGTGCTGC
	Pair 2	**5' GCACCATCTTCTTCAAGGAC**
		5' CGGCCATGATATAGACGTTG
	Pair 3	**5' AGCAGAAGAACGGCATCAAG**
		5' TCTCGTTGGGGTCTTTGCTC
DsRed **(asymmetrical PCR)**	**Pair 1**	**5' CGTCATCACCGAGTTCATGC**
		5' TTGGAGCCGTACTGGAACTG
	Pair 2	**5' ACAAGAAGCTGTCCTTCCCCC**
		5' TGGTCTTCTTCTGCATCACG
	Pair 3	**5' ACAAGGCCCTGAAGCTGAAG**
		5' TGCTCCACGATGGTGTAGTC

2.12.2.1.3 Chemical synthesis

Pre-U2 probe was chemically synthesised. The sequence is 5' [bio]CCGGAGGGGGTGCACCGTTCCT[bio]. [bio] is biotin tetra-ethyleneglycol.

2.12.2.1.4 Chemical labelling

Five 50-mer oligonucleotide probes were synthesised and labelled as described (Femino et al., 2003). The sequences of the oligonucleotides were selected

using conventional algorithms. A thymine was substituted by an amino-modifier C6-dT for every 10 bases (Gene Design, Tokyo, Japan), and labelled with Cy3 or Alexa 488 using Cy3 Mono Reactive Dye Pack (GE healthcare) or Alexa Fluor 488 reactive dye decapack for microarrays (Invitrogen).

12 μg oligonucleotides were purified by phenol extraction, ethanol precipitated, before resuspension in 6 μl 0.5 M $NaHCO_3$ (pH 9.3). One vial of dye (Alexa 488 pre-dissolved in 2 μl DMSO) was added into the solution, thoroughly mixed, incubated (20ºC; 60 min), mixed at 15 min intervals, and incubated for another 60 min (20ºC). Another vial of dye was added, mixed and incubated as described. A third vial was added with an additional 6 μl 0.5 M $NaHCO_3$ (pH 9.3), mixed and incubated with the same procedures.

Probes were then purified using G-50 columns, ethanol precipitated, purified and concentrated using a Microcon-30 column. The labelling efficiency was determined by spectrometry.

2.12.2.1.5 *In vitro* transcription

EGFP RNA probes were prepared by *in vitro* transcription as described in Ch.2.11.3.1, except the components of the nucleotide mix were 0.35 mM aminoallyl UTP (Invitrogen) or DIG-11-UTP, 1 mM each ATP, CTP, GTP, and 0.65 mM UTP. RNA labelled with aminoallyl UTP was then further chemically labelled with Alexa 488 or Cy3 as described in 2.12.2.1.4.

2.12.2.1.6 ULS labelling

Unmodified DNA oligonucleotides (30-mer) were directly labelled using ULS labelling kit (Invitrogen) according to the manufacturer's instructions.

2.12.2.1.7 Terminal labelling by terminal transferase

DNA oligonucleotides were 3' end-labelled with terminal deoxynucleotidyl transferase (Tdt) as described (Binnie et al., 2006).

2.12.2.2 Preparation of coverslips and slides

Pre-treatment of coverslips and slides prior to FISH were adapted from published procedures (Celis, 1998). Coverslips were incubated in 2N NaOH (20°C; 2h), rinsed extensively in H_2O, sonicated 5 min in H_2O and ethanol, and submerged in 70% ethanol. Slides were incubated in RNaseZap (20°C; 10 min), rinsed twice in DPEC-treated MilliQ H_2O, and submerged in 70% ethanol.

24 h prior to hybridisation, COS-7 cells (grown on coverslips in 6 well plates) were transfected with the appropriate plasmids using FuGene 6. HUVECs were treated with TNFα, and fixed at the appropriate time.

2.12.2.3 Fixation and permeabilisation

RNA FISH protocol was adapted from published procedures (Celis, 1998; Levsky et al., 2002; Chakalova et al., 2004). A comparison/optimisation of different fixation and permeabilisation procedures is described in Ch.4.4.1.2. The following is the optimal protocol. Cells were rinsed briefly in PBS, fixed (17 min; 20°C) in 4% PFA, 0.05% acetic acid in 0.15 M NaCl, washed three times (5 min; 20°C) in PBS, permeabilised (5 min; 37°C) in 0.01% pepsin (pH 2.0), rinsed in water, postfixed (5 min; 20°C) in 4% PFA in PBS, and washed (10 min; 20°C) in PBS.

2.12.2.4 Hybridisation

Hybridisation (16 h, 37°C) was carried out in a moist chamber. 50 ng tagged probe was added to hybridisation mix, and denatured (10 min; 94°C) prior to hybridisation.

2.12.2.5 Detection

Cells were then washed three times in 2x SSC (15 min; 37°C), twice (5 min; 20°C) in TST. Cells hybridised with DIG- or biotin-labelled probes were incubated in 1.4% blocking reagent in TS. Immuno-detection was carried out using sheep anti-DIG (60 min; 20°C; 1:250), followed by donkey anti-sheep Cy3 (30 min; 20°C; 1:250) or mouse anti-biotin (30 min; 20°C; 1:250), followed by donkey anti-mouse Alexa 488 (30 min; 20°C; 1:200) or donkey anti-mouse Cy5 (30 min; 20°C; 1:150). Cells were washed three times (5 min; 20°C) in PBS after incubation with each antibody, and mounted in Vectashield containing 1. μg/ml DAPI and 2.5 μm Orange/Green Intensity Calibration Beads (0.1% intensity; Molecular Probes) at a concentration of 6 x 10^4/ml. Cells hybridised with probes chemically labelled with fluors were mounted in Vectashield containing 1. μg/ml DAPI and 2.5 μm Orange/Green Intensity Calibration Beads (0.02% intensity) at a concentration of 6 x 10^4/ml.

2.12.3 Immuno-FISH

Cells were fixed (17 min; 20°C) in 4% PFA, 250 mM HEPES (pH 8.0), permeabilised (5 min; 37°C) in 0.5% Triton and 0.5% saponin, postfixed (as in Ch.2.12.2.3), hybridised (as in Ch.2.12.2.4), UBF or SC-35 detected using mouse monoclonal anti-UBF (2 h; 20°C; 1:150) or anti-SC-35 (2 h; 20°C; 1:200) followed

by donkey anti-mouse Alexa 647 (2 h; 20°C; 1:200). DIG-labelled DNA probes were immuno-detected using sheep anti-DIG (60 min; 20°C; 1:250), followed by donkey anti-sheep Cy3 (30 min; 20°C; 1:250)

2.13 Labelling transcription and replication sites

2.13.1 Labelling transcription sites using Br-U

Cells were grown in 10 mM Br-U (37°C; 15 min), rinsed in PBS (20°C; 5 min), fixed (20 min; 20°C) in 4% PFA, 250 mM HEPES (pH 8.0), permeabilised (5 min; 37°C) in 0.5% Triton X-100 and 0.5% saponin in PBS, washed (3x, 5 min; 20°C) in PBS, and Br-U detected using mouse anti-BrdU antibody (4-16 h; 20°C; 1:333), followed by donkey anti-mouse Cy3 (1 h; 20°C; 1:250).

2.13.2 Labelling transcription sites using Br-UTP

This protocol was adapted from a published procedure (Pombo et al., 1998). Cells were incubate on ice (5 min), washed with ice-cold PB+F buffer, permeabilised on ice with permeabilisation buffer (5 min), washed quickly with ice-cold PB+F, and incubated (33ºC; 15 min) with transcription buffer. Cells were then washed (2 x) with ice-cold PB+F, fixed with 4% PFA, 250mM HEPES (pH 8.0; 20ºC; 15 min), washed (3 x) with PBS, permeabilised (20ºC; 5 min) with 0.5% Triton X-100 in PBS, and washed (3 x) with PBS. Immuno-detection was as described in 2.13.1.

2.13.3 Combining detection of Br-RNA with RNA FISH

2.13.3.1 Br-U labelling combined with RNA FISH

RNA FISH was carried out as described in Ch.2.12.3 after the permeabilisation step described in Ch.2.13.1, then Br-U and FISH probes were immuno-detected simultaneously, using mouse anti-BrdU antibody (4 h; 20°C; 1:333) and sheep anti-DIG antibody (4 h; 20°C; 1:250), followed by donkey anti-mouse Alexa 488 (1 h; 20°C; 1:200) and donkey anti-sheep Cy3 (1 h; 20°C; 1:250).

Alternatively, after Br-U immuno-detection as described in Ch.2.13.1, cells were fixed in 4% PFA in PBS (20 min; 20°C). Cells were then permeabilised again in 0.5% Triton X-100 and 0.5% saponin in PBS (5 min; 20°C), before RNA FISH and detection was carried out as described in Ch.2.12.3.

2.13.3.2 Br-UTP labelling combined with RNA FISH

Br-UTP run-on was carried out as described in Ch.2.13.2, and the combination with RNA FISH was as in Ch.2.13.3.1.

2.13.4 Labelling replication sites using Br-dU

Cells were grown in 10 mM Br-dU (37°C; 15 min) in medium. Subsequent fixation and immuno-detection were as described in Ch.2.13.1.

2.14 Cryosections

2.14.1 General methods

The protocol for cryo-sectioning was adapted from (Tokuyasu, 1980; Faro-Trindade and Cook, 2006). Cells were fixed (20°C; 10 min) in 90 mm dishes in a

fixative containing 8% PFA, 250 mM HEPES (pH 8.0), scraped into a 1.5 ml tube, and centrifuged (14,000 rpm; 20°C; 1 min) to collect the pellet. Fresh fixative was now added and changed once every hour three times. The pellet was stored in fixative (4°C; 16 h), prior to transfer to 2.2 M sucrose in PBS (4°C; 72 h); subsequently, fresh sucrose was changed three times every hour.

The pellet was transferred onto copper blocks, excess sucrose removed with filter paper (Whatman), the pellet shaped into a cone under a Zeiss SV-11 dissecting microscope, and stored in liquid nitrogen.

120-500 nm sections were cut using a Reichert Ultravut S with FCS attachment cryomicrotome (Leica) with a cryo-diamond knife (Diatome) or a glass knife. The correct thickness of sections was selected by continuous interference colour. Sections were captured with 2.1 M sucrose in PBS, and transferred onto coverslips [coated with 0.5% gelatine and baked (70°C; 16 h)].

2.14.2 Labelling targets on cryosections

Cells were incubate on ice (5 min), washed with ice cold PB+F buffer, permeabilised on ice with permeabilisation buffer (5 min), washed quickly with ice-cold PB+F, and incubated (33ºC; 15 min) with transcription buffer. Cells were then washed (2 x) with ice-cold PB+F, fixed, and cryosectioned as in Ch.2.14.1. RNA FISH and immuno-detection were performed after cryosectioning as in Ch.2.13.3.2.

2.15 Fluorescence-activated cell sorting (FACS)

2 x 10^5 cells per sample were collected, washed once in PBS, resuspended in 2 ml PBS, and analysed on a FACScan flow cytometer (Becton-Dickinson) equipped with electronic doublet-discrimination module (DDM), which allows discrimination

between single cell and cell clusters. The forward-scatter (FSC) and side-scatter (SSC) light detectors allow size and complexity discrimination. Cell debris and cell clusters were removed from the analyses. EGFP was detected with FL1 (fluorescence detector 1, 535/30 bandpass filter which permits analysis of light at 530 nm), and DsRed with FL3 detector (585/42 bandpass filter, emission between 564 and 606 nm). Mock-transfected cells and cells transfected with a single plasmid (II ^{CMV}, *EGFP*, pA or II ^{CMV}, *DsRed*, pA) were used as controls. Data on 10,000 cells were collected and further analyses performed with CellQuest Software (Becton-Dickinson). The fractions of cells expressing red fluorescence from the reference plasmid (II ^{CMV}, *DsRed*, pA) and green fluorescence from the test plasmid (encoding *EGFP*) were calculated.

2.16 Imaging

Two fluorescence microscopes were used for analysis. For general immuno-fluorescence, a Zeiss Axioskop 40 was used; for more sensitive detection, a Zeiss Axioplan 2 Imaging microscope was used.

2.16.1 Zeiss Axioplan 2

Epifluorescence microscopy was performed using a motorised Zeiss Axioplan 2 Imaging microscope with MetaMorph 7 (Molecular Devices) and a 300W Xenon arc lamp. Images were collected with a Zeiss Plan-APOCHROMAT 63 x (1.4 oil; Ph3; ∞/0.17) lens. The specifics of the filters are listed in Table 2.5.

Table 2.5. Filters used on Zeiss Axioplan 2

	EXCITATION	DICHROIC	EMISSION
DAPI	S360/40X	84100BS	S457/50M
FITC	S490/20X	84100BS	S582/38M
CY3	S555/28X	84100BS	S617/73M
CY5	S635/20X	84100BS	S685/40M
GFP	HQ470/40X	495DCLP	HQ525/50

The 'S' before the first five filter combinations signifies the Chroma Quad Sedat filter set. The first number in the filter designations indicates the centre wavelength of the band of light the filter transmits, and the second the width of the band.
The dichroic used with the filter can transmit four bands of light simultaneously, as opposed to the usual long pass dichroics that reflect light below a certain wavelength and transmit light above that wavelength. The GFP filter is a standard Zeiss filterset 38.

The microscope was equipped with a charge-couple device (CCD) camera, either a Princeton Instruments MicroMax 1024B (resolution: 1024x1024 pixels; 1 pixel = 13 µm x 13 µm; bit depth: 16 bits) or Photmetrics CoolSNAP$_{HQ}$ (resolution 1392 x 1040 pixels; 1 pixel = 6.5 µm x 6.5 µm; bit depth: 12 bits).

2.16.2 Zeiss Axioskop 40

Conventional epifluorescence microscopy was performed using a Zeiss Axioskop 40 microscope with AxioVision 4.2 Software (Zeiss) and a 100 W (HBO 100) mercury arc lamp. Images were collected with a Zeiss Plan-NEOFLUAR 63 x (1.25 oil; Ph3; ∞/0.17) lens. The specifics of the filters are listed in Table 2.6.

Table 2.6. Filters used on Zeiss Axioskop 40

	Zeiss filter set	EXCITATION	DICHROIC	EMISSION
DAPI	49	G 365	FT 395	BP 445/50
FITC	10	BP 450-490	FT 510	BP 515-565
CY3	20	BP 546/12	FT 560	BP 575-640
CY5	26	BP 575-625	FT 645	BP 660-71
GFP	38	BP 470/40	FT 495	BP 525/50

The microscope was equipped with a CCD camera Axio Cam MRm (Bit depth: 12 bits; resolution: 1388 x 1040 pixels; 1 pixel = 6.45 μm x 6.45 μm).

2.17 Image analysis

Image analysis were performed using ImageJ (Rasband, 1997-2007), and images were processed using Adobe Photoshop CS2 and Corel Designer 9.0. Four types of analysis were carried out 1) the number of foci, 2) the intensity of signal, 3) the size of foci, 4) colocalisation. Foci were analysed by human eye or the 'analyse particle' feature of ImageJ; algorithms for defining positive signals are described in Ch 4.4.2. Signal intensities were normalized relative to that of reference beads. Definition of colocalisation is described in Ch. 4.4.3.

2.18 Chromosome conformation capture (3C)

The proximity of two DNA sequences in 3D nuclear space was assessed by 3C (Dekker et al., 2002; Tolhuis et al., 2002). In some cases, transcriptionally-active chromatin was selected using a ChIP-3C (Cai et al., 2006).

2.18.1 Conventional 3C

2.18.1.1 Fixation and lysis

4 x 10^7 cells were fixed (10 min; 20°C) in 50 ml DMEM + 10% FCS supplemented with 2% formaldehyde, before glycine was added to 0.125 M. Cells were scraped from the surface, pelleted (3500 rpm; 4°C; 15 min), and lysed (90 min; 4°C) with vigorous stirring in 50 ml lysis buffer.

2.18.1.2 Digestion

Cells were re-pelleted, incubated (1 h; 37°C) with shaking in 1x NEB buffer 3 + 0.3% SDS, and subsequently in 1.8% Triton-X (1 h; 37°C). 10^6 nuclei were digested (16 h, 37°C) with 600 U restriction enzyme (*Bam*HI or *Pst*I), and incubated (20 min, 65°C) in a final concentration of 1.6% SDS.

2.18.1.3 Ligation

Samples containing 2 μg DNA were incubated (1 h; 37°C) in 1% Triton-X, the volume adjusted to 2 ml, and incubated (4 h, 16°C) in 2,000 U T4 DNA ligase.

2.18.1.4 DNA purification

Cells were incubated (16 h; 65°C) with 100 μg/ml proteinase K, treated (30 min; 37°C) with 0.4 μg/ml RNase A, and DNA purified by phenol extraction and ethanol precipitation. PCR (25 cycles) was performed with serially-diluted samples

using primers complementary to backbone sequences as described in Table 2.7, and PCR products sized by gel electrophoresis.

2.18.2 Chromatin immuno-precipitation – 3C (ChIP-3C)

2.18.2.1 Fixation and lysis

8 x 10^6 cells were washed in PBS (3x; 0°C; 1 min) before 1 ml 0.67 M PFA in PBS (20°C; 10 min) was added to crosslink protein and DNA, and subsequently 670 µl 1 M glycine to quench the cross-linker. Cells were scraped from the surface and transferred to two 1.5 ml tubes, and pelleted (13,000 rpm; 20°C; 1 min). Cells were suspended in 1 ml ice-cold hypotonic buffer, incubated (0°C; 10 min), and nuclei pelleted (5,000 rpm; 4ºC; 5 min). The pellet was dissolved in 1 ml CSK buffer, incubated (0°C; 20 min), and pelleted again (5,000 rpm; 4ºC; 5 min).

2.18.2.2 ChIP

The pellet of nuclei was dissolved in 40 µl 5 M NaCl (0°C; 10 min), and the sample diluted to reduce the NaCl concentration to 150 mM. Antibodies against tri-methyl K4 in histone H3 (16 h; 4°C; 1:1000) were added and incubated (4ºC; 16 h) with rotation.

30 µl protein G Sepharose beads were washed (3x; 2,000 rpm; 20ºC; 3 min) with H_2O and once (2,000 rpm; 20ºC; 3 min) with restriction wash buffer, before addition to samples and incubation (4ºC; 1 h). Beads were collected (1000 rpm; 4ºC; 3 min), and supernatant saved for analysis of the unbound fraction. Beads were washed once (4ºC; 5min) with low stringency buffer, twice (4ºC; 5min) with high

stringency buffer, and once (4ºC; 5min) with low stringency buffer. Between each wash, beads were collected by centrifugation (2000 rpm; 4ºC; 3 min).

2.18.2.3 Digestion and ligation

Beads were then incubated (37°C; 16 h) with *Bam*HI or *Pst*I (600 U) in NEB Buffer 3 in 100 µl, before the reaction was stopped (65ºC; 10 min). Samples were then treated with 200 µg/ml RNase A (37ºC; 30 min), diluted with 400 µl of H_2O, and ligated (16ºC; 4 h) in ligation buffer with 2,000 U T4 DNA ligase, before cross-links were reversed (65ºC; 16 h).

2.18.2.4 DNA purification

Cells were incubated (42°C; 1h) with 100 µg/ml proteinase K, and DNA purified by phenol extraction and ethanol precipitation. PCR (25 cycles) was performed with serially-diluted samples using primers complementary to backbone sequences as in Table 2.7, and PCR products sized by gel electrophoresis.

Table 2.7. 3C primers

3C	primers	Ta	cycles
EGFP: DsRed	**5' TGAAGCGCATGAACTCGGTG**	**60**	**25**
	5' TGGTGCAGATGAACTCAGGG		
ALDOA	**5' AAATTTCCTCTGAAGCACCG**	**59.5**	**25**
(*Bam*HI)	**5' TTAATGGGGCATCAAGCAC**		
ALDOA	**5' TTGCAGTGAGCCGAAATCG**	**59.5**	**25**
(*Pst*I)	**5' AGGGTGTAGTCTCTGTTGGG**		
plasmid backbone:	**5' CCTCTACAAATGTGGTATGGC**	**56.6**	**25**
host *U2*	**5' GGCAGTGGGAGCAAAATTC**		
***Neo*^{r}**	**5'ATTCGGCTATGACTGGGCAC**	**58.7**	**25**
	5'AGTGACAACGTCGAGCACAG		

2.19 Statistics

The Mann-Whitney test, one-way ANOVA, and Student's *t*-test were performed using GraphPad 'Prism' (www.graphpad.com; San Diego, CA). Sample sizes were such that the last 5 samples analysed all lay within 5% of the progressive mean (Williams, 1977).

2.20 Bioinformatics

The bioinformatics tools used in this study were Ensembl (release 45 – Jun 2007, www.ensembl.org), UCSC Genome Browser (http://genome.ucsc.edu/), PCR primer Database (QPPD; http://www.ncifcrf.gov/rtp/gel/primerdb/), National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov) BLAST and OMIM; RIKEN OmicBrowse (http://omicspace.riken.jp/db/genome.html). Sequence alignments, restriction analysis, PCR primer/probes design were carried out using DS Gene 1.5 (Accelrys).

3 Strategy

3.1 Introduction

The main aim of this study is to answer the question: do particular transcription factories specialise in transcribing specific gene sub-sets? My strategy is to use replicating minichromosomes as probes and then examine the distribution of these probes to see if transcription occurs in factories. Next, I use FISH and 3C to investigate if two minichromosomes encoding different transcription units colocalise (i.e., target the same or different factories; Fig 3.1). In this chapter, I describe 1) the behaviour of minichromosomes in cells, and why a replicating system is used, 2) how I optimised transfection conditions, and 3) the design of plasmids encoding different transcription units.

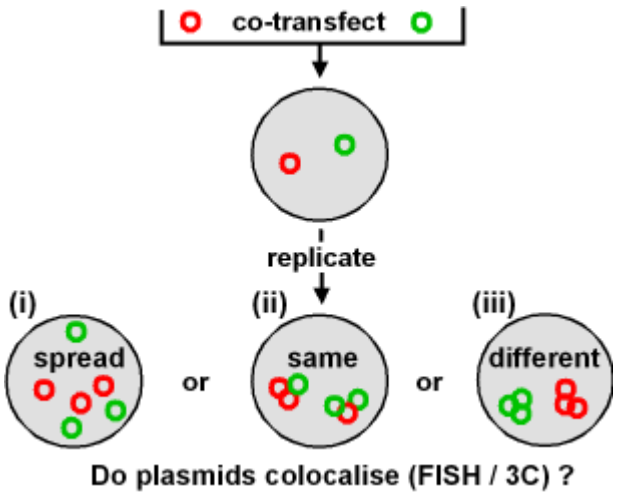


Fig. 3.1. Strategy

Different plasmids are co-transfected into COS-7 cells, and allowed to replicate; then, FISH and 3C are used to see if they are spread throughout nuclei, or targeted to the same or different factories.

3.2 Minichromosomes in cells

3.2.1 Why choose a replicating system?

As factories are small (~50 nm across) and numerous (factory:factory distance ~300 nm), they can only be resolved in the light microscope using confocal imaging of thin cryosections (~100 nm). However, one of the two gene copies in a diploid cell is unlikely to be found in such a thin section (the average HeLa cell nucleus is roughly 10-20 μm across. The actual size and shape is variable). Besides, since the number of factories is large (a typical HeLa cell contains ~8,000 pol II factories), two genes targeting a similar type of factory may not necessarily be in exactly the same factory. A large number of cells will have to be examined before the number becomes significant. For these reasons, replicating plasmids are used to increase copy number.

3.2.2 The SV40 replication system

I introduce plasmids encoding the SV40 origin (*ori*) of replication and early promoter into COS-7 cells, a monkey line that expresses the SV40 large T antigen (Mellon et al., 1981). Once the plasmids are transfected, the large T antigen protein binds to and activates DNA replication from the SV40 *ori* (Fanning and Knippers, 1992; Bullock, 1997; Bochkareva et al., 2006).

The control region of SV40 early RNA synthesis is illustrated in Fig 3.2.A. It contains three 26-bp DNA sites (I, II, III). The early RNA encodes the T antigen. Two 72-bp repeats act as enhancers. A cellular protein factor sp1 binds to the 21-bp repeats and is required for transcription. The T antigen binds to site I initially, but

when the antigen is at higher concentrations it also binds to sites II and III. *In vitro*, binding at both sites I and II will stop transcription from the early promoter. In this study, the T antigen is steadily expressed in the COS-7 cells, and the SV40 early promoter is hence down-regulated (Myers et al., 1981; Dynan and Tjian, 1985).

3.2.3 Minichromosomes are replicated by the cellular machinery

In the 1970s, the 5,224 base-pair SV40 DNA (Fiers et al., 1978) became a popular subject for studies on the structure and function of nucleosomes. In infected cells, SV40 DNA and cellular histones assembled a chromatin-like structure called a "minichromosome" (Griffith et al., 1975). It forms a spherical particle ~300 Å in diameter, and contains all five histones (Fig.3.2.B). This compact form of the SV40 minichromosome is completely resistant to staphylococcal nuclease (Varshavsky et al., 1977).

Similarly, after transfection into COS-7 cells, plasmid DNA encoding the SV40 *ori* and early promoter is assembled into nucleosomes, and the resulting minichromosomes are replicated and transcribed by the cellular machinery (Mellon et al., 1981; Jackson and Cook, 1993; Dean, 1997; Maul, 1998). Among the population of plasmids, >80% plasmid DNA molecules are supercoiled and so derived from the nucleosome-covered compact form, with <5% being replicating intermediates (Jackson and Cook, 1993).

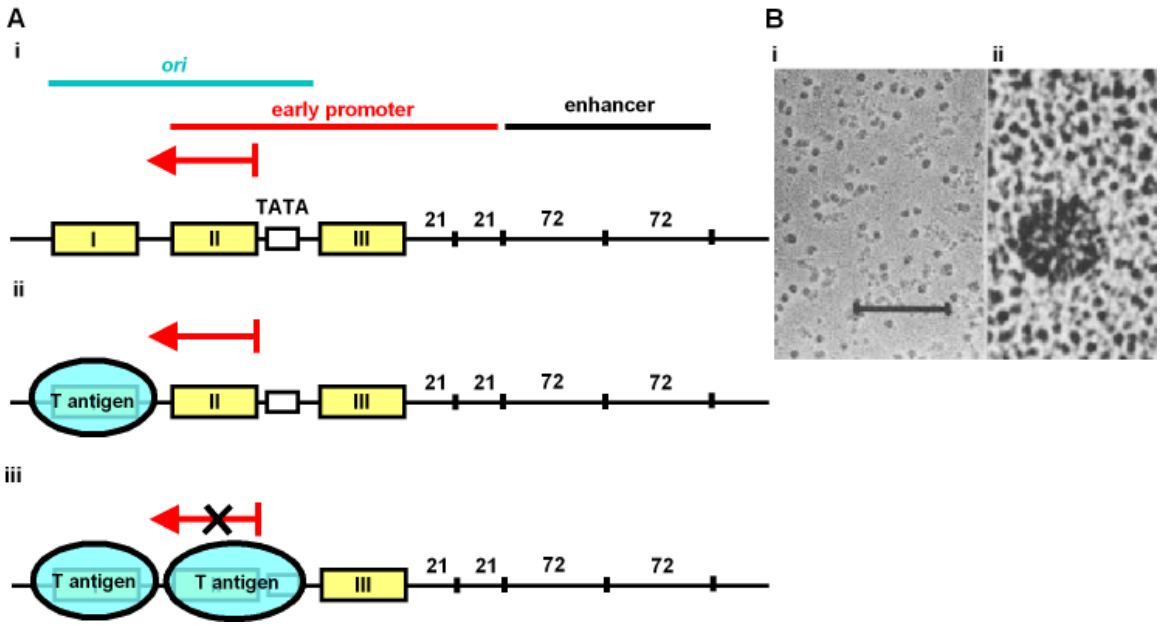


Fig. 3.2. SV40 minichromosomes

A. The transcriptional regulation of the SV40 early promoter and *ori* [adopted from (Myers et al., 1981; Darnell et al., 1986)].

i. The SV40 early promoter and *ori*. Yellow boxes illustrate the 26-bp DNA sites (I, II, III) to which T antigen binds. Two 72-bp repeats act as enhancers. A cellular protein factor Sp1 binds to the 21-bp repeats and is required for transcription. The backbone of my constructs contains this entire cassette, but not the SV40 late promoter downstream (Shaw et al., 1985; Omilli et al., 1986) (Fig. 3.4B).

ii. At an early stage during infection, T antigen preferably binds to site I and replication initiates.

iii. Later, when the T antigen concentration becomes higher, it binds to sites I and II; this inhibits transcription from the early promoter.

B. Electron microscopy of metal-shadowed compact SV40 minichromosomes (from (Varshavsky et al., 1977).

i. Minichromosomes prepared for electron microscopy in 0.13 M NaCl. Bar: 0.5 μm.

ii. 100x magnification of (i).

3.3 Optimising transfection

3.3.1 Choice of reagent

Two transfection reagents were compared; efficiencies were evaluated 24 h after transfection. FuGene6 outperformed Lipofectamine under my conditions (Fig.3.3.A), and a DNA:FuGene6 ratio of 1 μg:3 μl was chosen for further optimisation.

3.3.2 Optimising input plasmid

In order to be detected, sufficient plasmids must be transfected, but high inputs yield many transcriptionally-inert minichromosomes (Fig 3.3.B; (Jackson and Cook, 1993)). Thus, low inputs (i.e. 0.2 μg-1 μg per plate) were used as a compromise between maximising copy number and minimizing the number of inactive chromosomes. I first established the lowest input required to ensure the maximum number of cells were transfected.

Efficiencies were evaluated 24 or 48 h after transfection. Transfection efficiency increased as input increased up to 0.25 μg/coverslip (Fig. 3.3.C). Therefore, an input of 0.25 μg/coverslip (cs), and harvesting after 24 h was used in all future experiments.

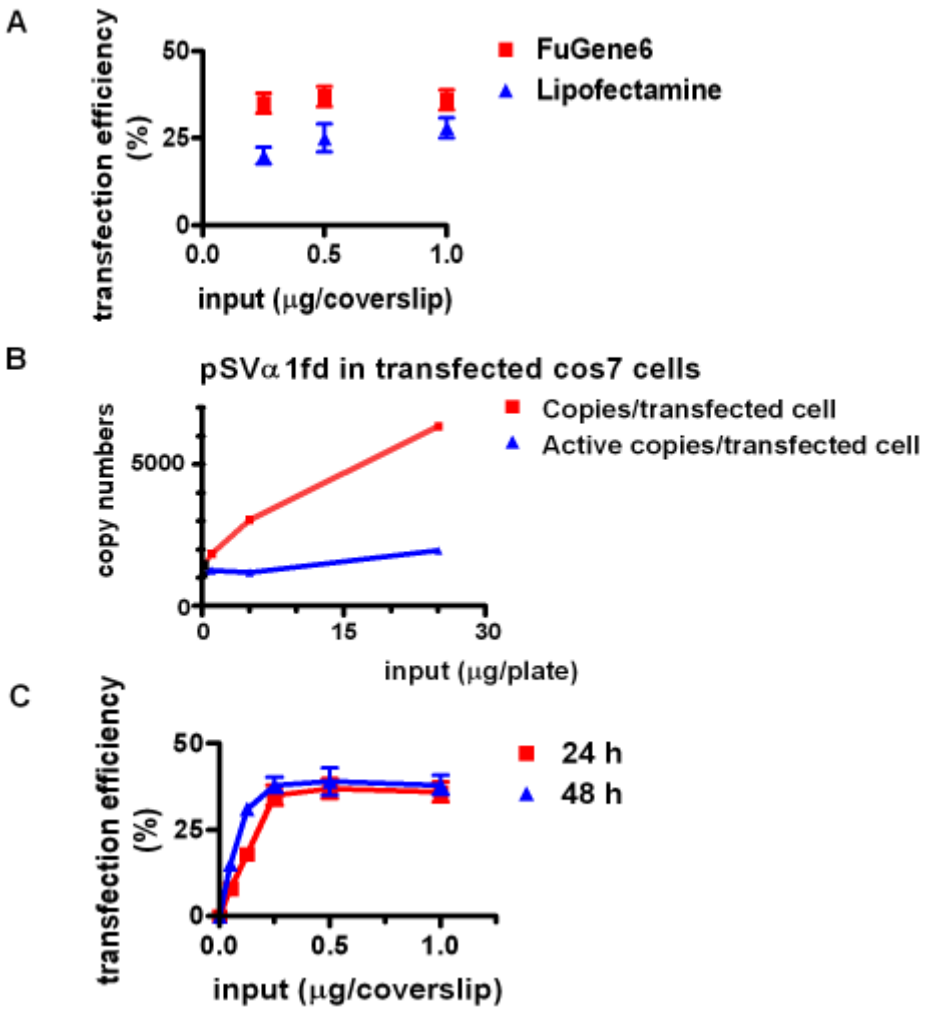


Fig. 3.3. Optimising transfection

A. Determining the best transfection reagent. Transfection efficiencies using FuGene6 and Lipofectamine (input 0.25 – 0.1 μg/cs) were evaluated 24 h after transfection as described in Methods. FuGene6 outperforms Lipofectamine.

B. High inputs result in many transcriptionally-inert minichromosomes. The number of active plasmids was estimated by their ability to resist electroelution from nuclei (as active plasmids attach to underlying nuclear structures, unlike inactive ones). The same numbers/cell were retained irrespective of initial input. Data from (Jackson and Cook, 1993).

C. Optimising input. Transfection efficiencies using FuGene 6 with different inputs (0.06 – 1.0 μg/cs) were evaluated 24 or 48 h after transfection as described in Methods. Transfection efficiency increases as input increases up to 0.25 μg/cs.

3.4 Construction of plasmids

Plasmids based on a common backbone but encoding different transcription units with various promoters, ± introns, and with different 3' ends were constructed (Fig. 3.4). The SV40 origin of replication (*ori)* in the backbone allows the plasmids to replicate in COS-7 cells. *Ori* also encodes the SV40 early promoter. Plasmids are named by promoter type, coding sequence, whether or not they contain an intron, and 3' end. For example, plasmid II^{CMV},$EGFP_V$,pA possess the (polymerase II) CMV promoter upstream of *EGFP* followed by the β-*globin* intron and a poly(A) signal, while 0,*EGFP*,pA lacks a promoter and intron. Details of plasmids construction was described in Ch.2.7.

3.5 Conclusion

In order to investigate if factories specialise in transcribing specific gene sub-sets, I use replicating plasmids encoding different promoters, transcription units, +/- introns, and 3' ends as probes. Plasmid DNA is transfected into cells where it is assembled into nucleosomes, and the resulting minichromosomes are then replicated and transcribed by the cellular machinery. As high inputs yield many transcriptionally-inert minichromosomes, I optimised the transfection method and determined the minimum transfection input giving sufficient minichromosomes for detection without too great an input background. I now go on to determine whether transcribing minichromosomes are spread throughout the nucleus, or concentrated in the same or different factories (Fig. 3.1, possibilities i-iii).

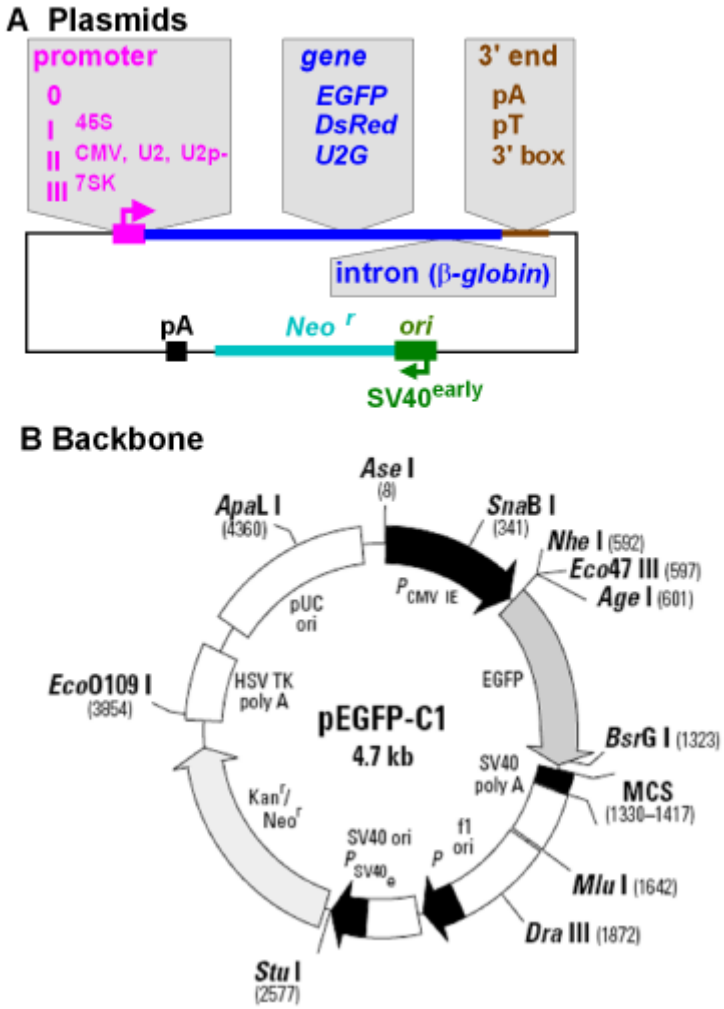


Fig. 3.4. Plasmids.

A. Plasmids encoding different promoters (pol I, II, III), different transcription units (*EGFP, DsRed, U2G*), ± the β-*globin* intron, and ± different 3' ends [poly(A), poly (T), 3' box] are shown. The common backbone is derived from pEGFP-C1, containing the SV40 *ori*, and early promoter driving *Neo*^{r} .

B. Backbone pEGFP-C1 encodes the EGFP which has increased translation efficiency in eukaryotic cells. SV40 polyadenylation signals downstream of the EGFP gene direct proper processing of the 3' end of the EGFP mRNA. The vector backbone also contains an SV40 origin for replication in mammalian cells expressing the SV40 T-antigen. A neomycin resistance cassette (*Neo*^{r}), consisting of the SV40 early promoter (including two 72-bp enhancers, four 21 bp repeats, and an early promoter element), the neomycin/kanamycin resistance gene of Tn5, and polyadenylation signals from the Herpes simplex virus thymidine kinase (HSV TK) gene. A bacterial promoter upstream of this cassette expresses kanamycin resistance in *E. coli*. The pEGFP-C1 backbone also provides a pUC origin of replication for propagation in *E. coli* and an f1 origin for singlestranded DNA production.

4 Minichromosomes are transcribed in factories

4.1 Introduction

In order to investigate whether factories specialise in transcribing different gene subsets, I developed a method using replicating minichromosomes encoding different transcription units as probes.

In this chapter, I describe the establishment and optimisation of this method. First, I assessed the copy number of minichromosomes using 'Southern' blots and quantitative PCR, the copy number of transcripts encoded by the different plasmids by qRT-PCR, and protein expression levels by FACS. Then, I established optimal conditions for DNA and RNA FISH to visualise minichromosomes and their transcripts, and defined algorithms for image analysis.

4.2 Copy numbers of minichromosomes and their transcripts

4.2.1 Copy number of minichromosomes per cell

In order to confirm plasmids are replicated by the cellular machinery to create large numbers of minichromosomes in a cell, I first used 'Southern' blots and quantitative PCR to assess plasmid DNA copy number.

4.2.1.1 Copy number assessed by 'Southern' blot

During transfection, only a fraction of plasmids enter the cell; a large number of input plasmids may be lodged on the cytoplasmic membrane. This will generate a high background which can be confusing. In order to reduce the input background, 8 h after transfection, cells were trypsinised, washed with fresh medium, replated, and regrown for a total of 12, 24, or 36 h.

At each time, samples were used to isolate minichromosomal DNA by 'Hirt' extraction. Transfection efficiencies were determined by FACS or microscopy using EGFP fluorescence (Fig. 4.1C).

Quantitative 'Southern' blots show that minichromosomal copy number increases progressively from the level seen at 8 h (Fig. 4.1A and B). The 8 h sample contains unreplicated input DNA (i.e., naked unsupercoiled and so input DNA, and not supercoiled DNA derived from histone-covered minichromosomes). It provides an indication of the maximum amount of input background that could possibly be present, and so is labelled 'max background'. By 24 h – at least 8,000 new copies have been generated in each cell (Fig. 4.1A and B).

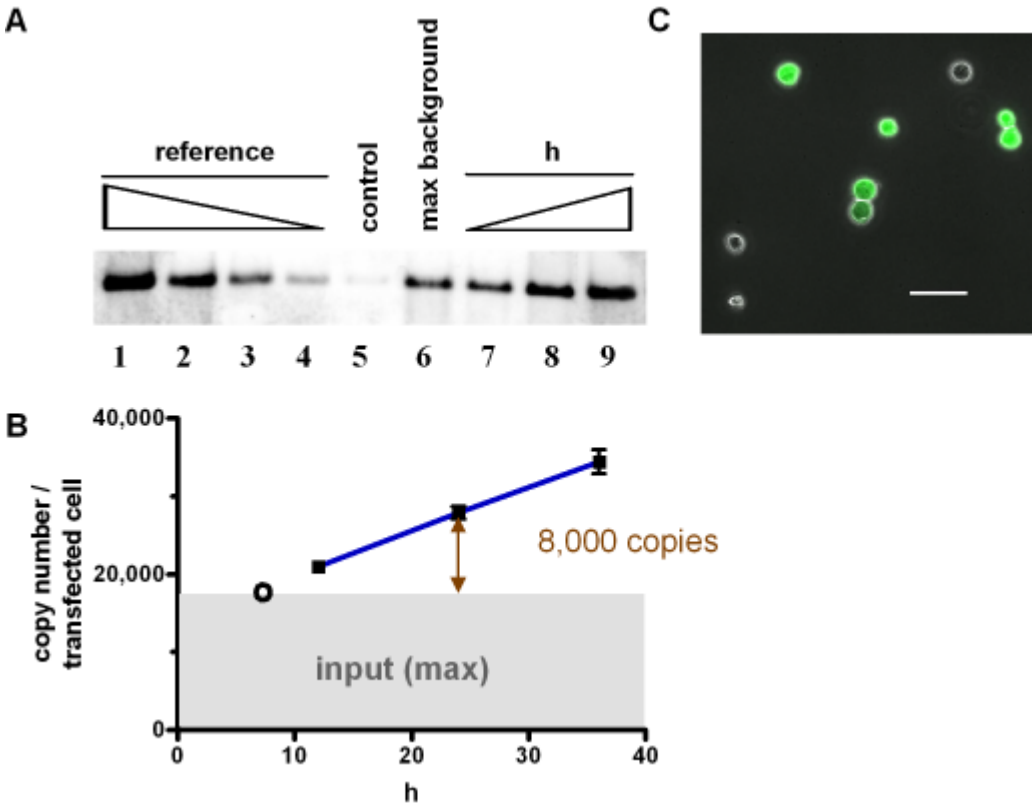


Fig. 4.1. Plasmids are replicating in the nucleus after transfection.

Cells were transfected with II^{CMV},*EGFP*,pA, grown for 8 h, trypsinised, washed, and regrown for 12, 24, or 36 h after transfection. A Hirt extract was prepared to determine plasmid copy number by 'Southern' blotting; the transfection efficiency was determined from the fraction of cells expressing EGFP.

A. Image of a typical 'Southern' blot. Lanes 1-4: three fold dilutions of known numbers of plasmids. Lane 5: Mock-transfected control. Lane 6: 8 h sample (indicating the maximum input background). Lanes 7-9: 12, 24, 36 h samples. Copy number increases with time.

B. Copies per transfected cell. The total number of plasmids present in the population was determined using images like in A, and the number of copies / transfected cell calculated (using the transfection efficiency). The 8 h sample and grey area indicate the maximum background.

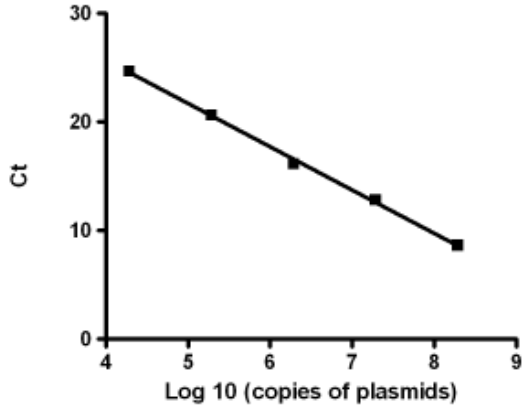
C. A typical image used to determine transfection efficiency. 6 out of 9 cells are expressing EGFP, and so have been successfully transfected. Bar: 60 μm.

4.2.1.2 Copy number assessed by quantitative PCR

Quantitative PCR is carried out to compare the minichromosomal copy number per cell for each different plasmid, by reference to known numbers of pure plasmids. Cells were co-transfected with a reference plasmid (i.e., II^{CMV},*DsRed*,pA), and another test plasmid encoding EGFP. Input background was minimised as described in 4.2.1.1, before total DNA was extracted. Transfection efficiencies were determined by FACS or microscopy using DsRed fluorescence. The number of cells was determined by reference to the DNA copy number given by the cellular gene, *GAPDH*.

Different plasmids replicate to give similar copy number per cell (Fig. 4.2.B). The copy number obtained by quantitative PCR is consistent with the 'Southern' blot result. By 24 h, at least 8,000 new copies (eliminating the maximum background) have been generated in each cell.

A



B

	Ct	Total plasmid copy number	Number of cells	Transfection efficiency	copies per transfected cell
I^{45S},*EGFP*,pA	11.27±0.17	41637825	5924	30.23%	23,000±1,900
II^{CMV},*EGFP*$_V$,pA	9.78±0.39	100194943	15530	36.58%	18,000±3,100
III^{7SK},*EGFP*,pT	10.17±0.30	79416161	13395	27.92%	21,000±2,800

Fig. 4.2. Plasmid DNA copy number determined by qPCR

Cells were co-transfected with a reference plasmid (II^{CMV},*DsRed*,pA) and another test plasmid encoding EGFP as indicated, grown, and replated as in Fig. 4.1. Total DNA was isolated with Trizol, and the transfection efficiency determined from the fraction of cells expressing DsRed.

A. Generation of a standard curve for real-time PCR. Total DNA is extracted from ~10^7 (untransfected) cos7 cells, and mixed with known amount of pure II^{CMV},*EGFP*,pA DNA. Plasmid DNA copy number is determined by real-time PCR using *EGFP* primers. The average threshold cycle number (Ct) from three independent extractions was plotted against the amount of plasmid added. The straight line is a variance weighted linear fit through all the measurements (R^2=0.9985).

B. DNA copy number of different plasmids determined by reference to the standard curve and the level of *GAPDH*. At 24 h, different plasmids yield similar copy numbers (one-way ANOVA; p>0.05). PCR product was not detected in mock transfected cells.

4.2.2 RNA copy number by qRT-PCR

Since different plasmids replicate to give similar copy numbers per cell, do they also yield similar transcript levels? Here, cells were co-transfected with a reference plasmid II^{CMV},*DsRed*,pA, and another test plasmid encoding EGFP as in Ch.4.2.1, before RNA copy number was evaluated using qRT-PCR.

Although transcript copy numbers also increase, promoter type and the presence of an intron affect expression levels significantly (Fig. 4.3). Minichromosomes encoding pol I or III promoters generate far fewer transcripts compared with those encoding a pol II promoter. The transcript level increases two-fold on insertion of an intron. This result is consistent with early observations in transgenic mice, where introducing an intron increased mRNA expression 10-100 fold (Brinster et al., 1988). The mechanism of this intron-induced increase is unclear, but it may be due to more efficient export of intron-containing RNA to the cytoplasm (Ch.1.1.4).

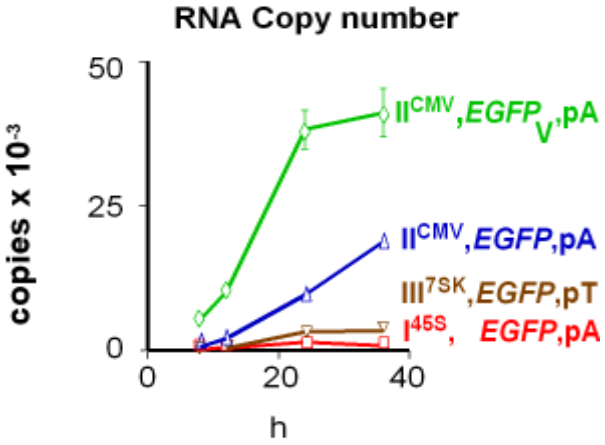


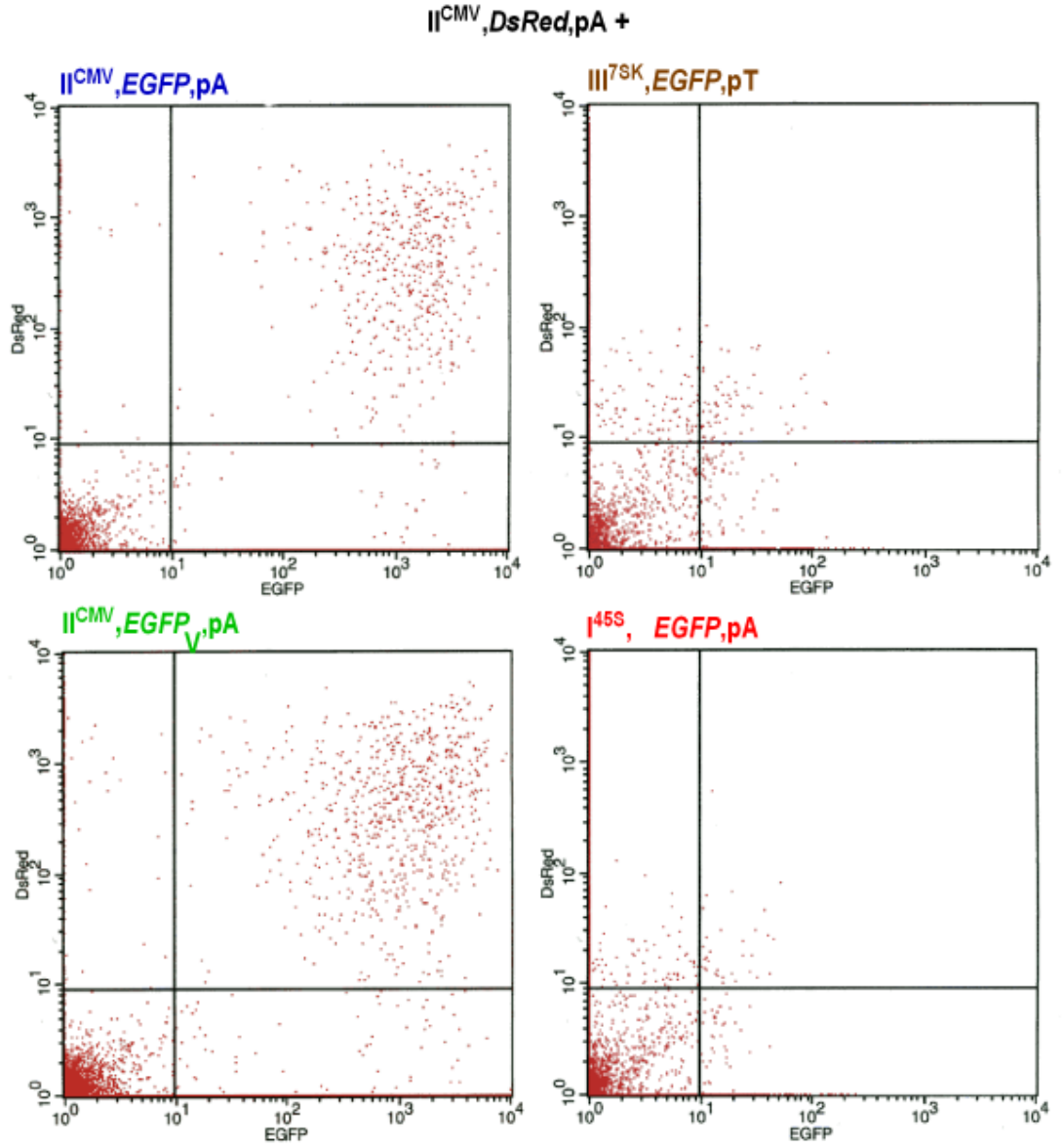
Fig. 4.3. RNA copy number per transfected cell

Cells were co-transfected with a reference plasmid II^{CMV},*DsRed*,pA, and another encoding EGFP as indicated, grown and replated as in Fig. 4.2. Total RNA was isolated with Trizol, and transfection efficiencies determined from the fraction of cells expressing DsRed. Transcript levels were also determined using qRT-PCR by reference to known amounts of RNA prepared by transcription *in vitro*, and normalised to the transcript level given by *GAPDH*. Copy numbers increase during the first 24 h, but promoter type and the presence of an intron affect expression.

4.2.3 Protein expression level by FACS

I went on to see if the protein expression level of EGFP also differs among minichromosomes encoding different promoters and +/- introns. Here, cells were co-transfected with a reference plasmid II^{CMV},*DsRed*,pA, and another test plasmid encoding EGFP as in Ch.4.2.2, before EGFP protein levels were assessed by FACS. The fraction of cells expressing EGFP protein encoded by minichromosomes with pol I and III promoters is significantly lower than that from pol II promoters, while the fraction of cells expressing DsRed remains constant (Fig. 4.4 B). Moreover, minichromosomes encoding an intron yield more EGFP than the others, consistent with the increased mRNA levels seen by qRT-PCR (Fig. 4.4 A).

A



B

II ^{CMV}, *DsRed*, pA +	DsRed (%)	EGFP (%)
II ^{CMV}, *EGFP*, pA	36.9	35.8
II ^{CMV}, $EGFP_V$, pA	36.7	38.6
III ^{7SK}, *EGFP*, pT	37.3	1.2
I ^{45S}, *EGFP*, pA	38.8	0.8

Fig. 4.4. Promoters and introns influence protein expression

Fig.4.4. Promoters and introns influence protein expression

Cells were co-transfected with a reference plasmid II^{CMV},*DsRed*,pA, and another encoding EGFP as indicated, and DsRed and EGFP fluorescence determined by FACS. Debris and aggregates were removed from analysis. Background was determined using mock-transfected cells and cells transfected with a single plasmid (II^{CMV}, *EGFP*, pA or II^{CMV}, *DsRed*, pA).

A. FACS analysis. Each red dot represents one event. The X-axis indicates the intensity of EGFP fluorescence, and the Y-axis that of DsRed. Quadrants were drawn to distinguish different types of cells: lower left, untransfected cells; upper left, cells only expressing DsRed; upper right, cells expressing both DsRed and EGFP; lower right, cells only expressing EGFP. Significantly less EGFP fluorescence was detected from minichromosomes encoding pol I and III. Introducing an intron leads to higher EGFP expression.

B. The fraction of cells expressing DsRed and EGFP was determined from the numbers in the different quadrants in A from three independent experiments. The fraction of cells expressing EGFP from minichromosomes encoding pol I and III promoters is significantly lower than from those encoding a pol II promoter, while the fraction of cells expressing DsRed remains consistent. The fraction of cells expressing EGFP from minichromosomes with or without an intron is essentially the same (Student's *t*-test). This indicates while the transfection efficiencies of different plasmids are the same, EGFP expression levels vary.

4.3 DNA FISH reveals minichromosomes are in factories

I next examined where the minichromosomes are localised in the cell using DNA FISH. I transfected only one plasmid – II^{CMV},*EGFP*,pA – and introduced plastic reference beads into the mounting medium to provide a standard that I could use to normalise intensities.

After replication by the cellular machinery, the resulting minichromosomes are organised into chromatin like the endogenous chromosomes. DNA FISH shows that most minichromosomes are concentrated in ~20 bright nuclear foci against a diffuse background (Fig. 4.5 A ii and iv).

The average intensity (relative to the intensities given by the reference beads; Fig. 4.6 A) and size of foci increase with time, while the intensity outside foci remains constant (Fig. 4.5 B); furthermore, the number of foci per nucleus remains constant (Fig. 4.5 A ii and iv), suggesting newly-replicated minichromosomes are concentrated in the existing foci. These results are consistent with foci being sites of replication (see also (Dean, 1997; Maul, 1998)). Taking copy numbers quantified in Ch.4.2 into account, a typical focus will contain ~200 minichromosomes by 24 h.

Previous studies show that <5% minichromosomes form concatamers and/or are replicating intermediates (Jackson and Cook, 1993), so something must confine them to the foci. As minichromosomes are not diffusely spread, we can eliminate possibility (i) in Fig. 3.1.

The total intensity given by the minichromosomes inside foci also increases with time. The fraction outside foci may represent the transcriptionally-inactive fraction (Jackson and Cook, 1993) as they are not attached to any transcriptional machinery and so able to freely diffuse in the nucleus.

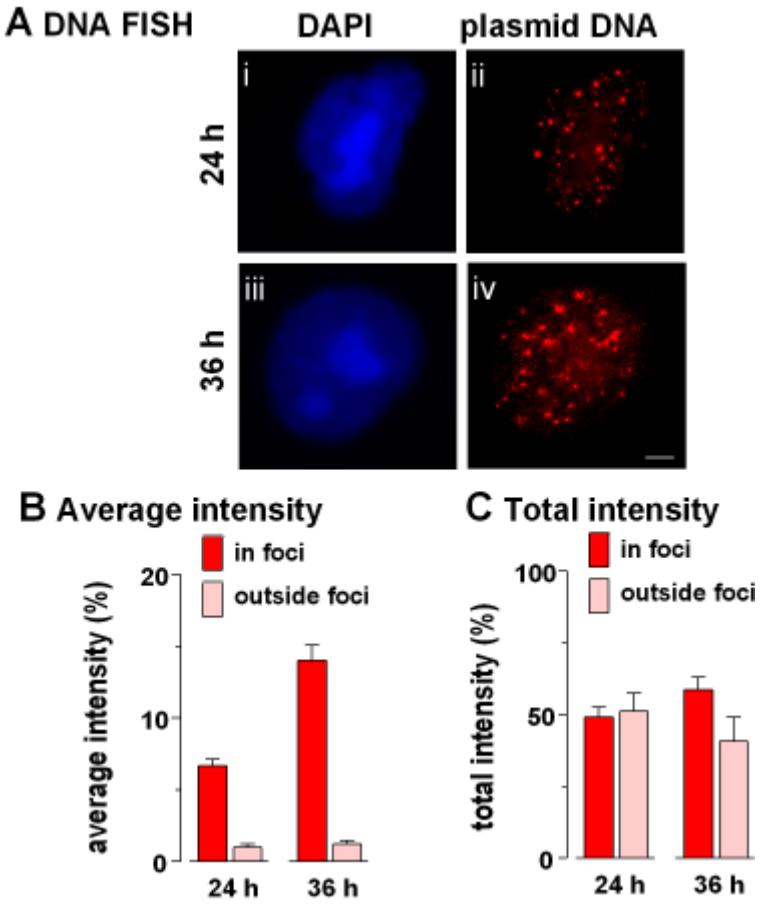


Fig. 4.5. Replicating minichromosomes are concentrated in a few foci.

Cells were transfected with II^{CMV},*EGFP*,pA. After 8 h, cells were replated (to wash away input DNA), and regrown for a total of 24 or 36 h after transfection. At each time, minichromosomes were visualised by DNA FISH.

A. Minichromosome distribution (by DNA FISH). Minichromosomes are concentrated in ~20 bright foci per nucleus, irrespective of time after transfection. Bar: 2.5 µm.

B. Average intensities (by DNA FISH) of pixels in and outside foci (+SD). Intensities like those seen in (A) were normalized relative to reference fluorescent beads (not shown). Most signals are in foci after 24 h, and this increases over time without much change in signal outside foci.

C. Total intensities were determined by the mean intensities in B and the area occupied in and outside foci (+SD). The fraction of plasmids in foci increases with time (Student's *t*-test; $P<0.05$).

4.4 RNA FISH reveals minichromosomes are transcribed in factories

I next examined whether minichromosomes are transcribed in factories. I first established the best RNA FISH conditions – those providing the lowest background and the highest detection sensitivity.

As in Ch. 4.3, only one plasmid II^{CMV},*EGFP*,pA was transfected. It encodes the CMV promoter (driving *EGFP*), plus the SV40 early promoter (driving *Neo*^{r}) on the common backbone (Fig. 3.4); the latter is down-regulated under the conditions used here (Myers et al., 1981; Dynan and Tjian, 1985).

4.4.1 Optimising RNA FISH conditions

4.4.1.1 Selecting probes for RNA FISH

As described in Ch.2.12.2.1, various probe preparation methods can be used, with nick translation (Ch.2.12.2.1.1) being the easiest and most economical. Labelling and purification can be finished within 3 h, and probes can be stored for a relatively long period (~two months). The labelled probes normally have a median length of 150~200 bp, with 1 out of every 10 nucleotides containing a DIG residue; Fig. 4.6B). However, since the probe is double-stranded, it yields a certain degree of negative-strand background. Therefore, nick-translated probes are not as sensitive as chemically-labelled probes (Ch.2.12.2.1.4). PCR-labelled probes (Ch.2.12.2.1.2) have similar pros and cons. Most of my targets are present in high copy numbers, and so I often do not require the highest sensitivity. Therefore, nick-translated and PCR-labelled probes satisfy my requirements for most RNA FISH experiments.

Chemically-labelled probes (Ch.2.12.2.1.4) are the most sensitive. They can detect as few as one RNA molecule (Ch.8.2.2). Since the degree of labelling can be quantified accurately, RNA FISH signals detected using them can also be quantified accurately. As the probe is a single-stranded oligonucleotide, the hybridisation efficiency is higher than for double-stranded DNA probes. Furthermore, using a barcode system, as many as 12 multiple transcripts can be detected simultaneously (Levsky et al., 2002). However, synthesis of these probes is very expensive, while labelling and purification is complex. Thus, chemically-labelled probes are suitable for scarce samples requiring accurate quantification. I use them to visualise a transcription wave in HUVECs (Ch.8).

For RNA FISH targeting U2 transcripts, a short sequence (22 nt) must be targeted in order to distinguish nascent U2 RNA from mature U2 RNA. Thus, a chemically-synthesised oligonucleotide labelled with biotin was used.

RNA probes made by *in vitro* transcription and ULS labelling are unstable, and they can give a high general background (see also (Chakalova et al., 2004)). Terminal transferase-labelled probes do not provide the required sensitivity. Therefore, these two labelling methods were not suitable for my requirements.

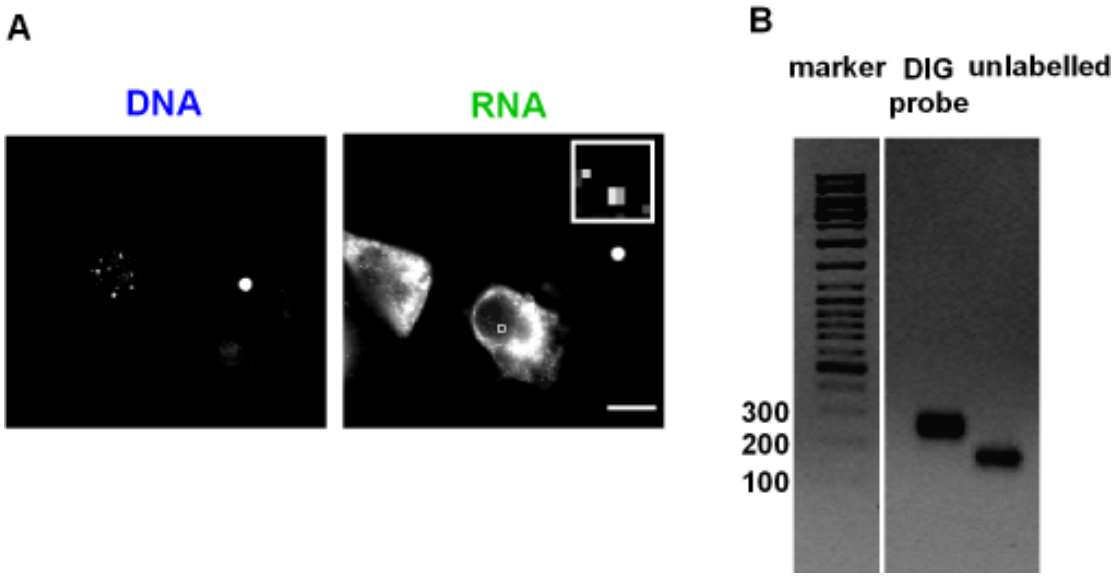


Fig. 4.6. Quantifying FISH signals

A. Intensities of foci are normalized relative to fluorescent beads.

Cells were transfected with II^{CMV},*EGFP*,pA, and *EGFP* DNA or RNA detected by FISH; fluorescent beads were added to the mounting medium. The intensities of nuclear foci were normalized relative to the intensity of fluorescent beads. Inset: a magnified view of typical foci. Bar: 10 μm.

B. The size of a FISH probe prepared by nick translation.

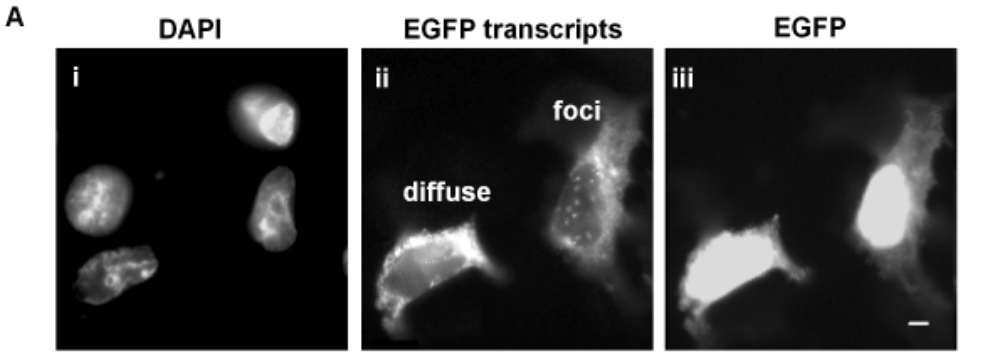
DNA complimentary to the RNA target was first synthesized by PCR, before it was nick translated using a DIG nick-translation kit. The size of the products was determined by gel electrophoresis. A median size of ~150-200 was selected for RNA FISH experiments. DNA labelled with DIG-11-dUTP (DIG probe) migrates more slowly than a control made with dUTP (unlabelled).

4.4.1.2 Selecting permeabilisation and fixation conditions

Different fixation and permeabilisation conditions affect the resulting distribution of polymerases and preserve cellular ultrastructure to different degrees (Guillot et al., 2004). Sufficient permeabilisation is needed to allow probes to access their RNA targets, but cellular structure must be preserved.

Using different combination of fixation and permeabilisation, I observed two patterns of RNA FISH signals. In one, transcripts are evenly distributed throughout nuclei; in the other they are concentrated in ~20 discrete and bright foci per nucleus (Fig. 4.7A). However, the average EGFP protein expression levels (measured by the intensity given by EGFP fluorescence) are similar between the two patterns (not shown). I used progressively harsher fixation and permeabilisation conditions, and found that harsher conditions increased the fraction of nuclei with bright foci (Fig. 4.7 B).

These results can be explained in several ways including: 1) Harsher conditions result in aggregation, so the focal pattern could be an artefact. If this is the case, the total intensity in the nuclear foci seen under harsh conditions should be the same as that seen in the diffuse pattern. 2) Harsher conditions remove some nuclear signal, so only a fraction of the transcript population is seen. If true, the total intensity in foci should be lower than that in the diffuse pattern. 3) The two patterns arise because the harsher condition provides easier access for probes, increasing detection sensitivity. Here, the total intensity of foci pattern should be higher than that of the diffuse pattern.



B

Conditions	Pattern (%)	
Fixation / permeabilization	Foci	Diffuse
1. mild	5 ± 1	95 ± 5
2. moderate	45 ± 2	55 ± 3
3. harsh	64 ± 3	36 ± 2

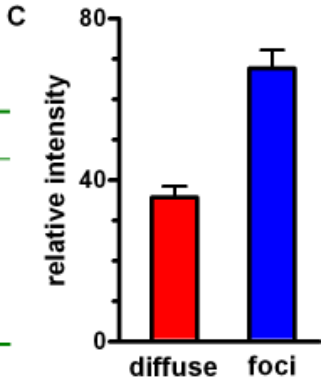


Fig. 4.7. Two patterns of RNA FISH arise from differential detection

A. II^{CMV},*EGFP*,pA was transfected into cells, cells fixed and permeabilised using a moderate procedure, *EGFP* transcripts visualised using RNA FISH (middle), DNA counterstained with DAPI (left), EGFP protein fluorescence visualised (right). Three views of one field are shown. *EGFP* transcripts were found in two patterns, either concentrated in foci or diffusely spread. Bar: 5 µm.

B. Increasing harshness used during fixation and permeabilisation yield different fractions of the two patterns. 1 (mild): fixation (4% PFA in PBS, 20°C, 20 min), permeabilisation (0.5% Triton X-100, 20°C, 5 min). 2 (moderate): fixation (4% PFA in 250 mM HEPES, 20°C, 20 min), permeabilisation (0.5% Triton X-100, 20°C, 10 min). 3 (harsh): fixation (4% PFA, 0.05% acetic acid in 0.15 M NaCl, 20°C, 17 min), permeabilisation (0.01% pepsin, pH 2, 37°C, 5 min). Harsher conditions increased the fraction of nuclei with bright foci.

C. Mean nuclear intensities (+SD) of the two patterns relative to reference beads. Nuclear intensities of the focal pattern is significantly higher than that of the diffuse pattern, suggesting a differential detection (Student's *t*-test; $P<0.05$).

Nuclei with foci were 1.5 fold brighter than those with a diffuse pattern (Fig. 4.7 C); this excludes the first two possibilities. Therefore, I conclude that harsh conditions are required to maximize detection of transcripts by RNA FISH.

In analogous experiments involving "transient" transfection and different fixation conditions, two patterns of nuclear signal were also observed; as plasmid input increased, the fraction with the diffuse pattern increased (Binnie et al., 2006). In Binnie's experiment, the transfection input is higher than mine (equivalent to 6.6 μg/plate vs. 1.2 μg/plate). Previous study indicated higher plasmid input creates more transcriptionally-inert minichromosomes (Fig. 3.3B, (Jackson and Cook, 1993)). Although similar as they appeared, the two patterns observed in Binnie's and my experiments probably resulted from different causes. The diffuse pattern from Binnie's experiment was likely to be composed of inactive minichromosomes (Fig. 3.3B), while my diffuse pattern was mainly a detection problem.

4.4.2 Defining a focus

I distinguish a nuclear focus from the background using two criteria: the focus must (i) have an intensity greater than a threshold defined as below; (ii) occupy ≥4 pixels (where a pixel is 100 x 100 nm and >85% nuclear signal is contained in such foci; Fig 4.8 A).

Some *DsRed* transcripts might be due to read-through from the SV40 promoter on the common backbone (despite the down-regulation and presence of an efficient polyadenylation signal). I constructed a plasmid lacking a promoter (i.e., 0,*EGFP*,pA). Any *EGFP* RNA detected with this minichromosome must result from read-through from the SV40 promoter. *EGFP* signal is normalised relative to the

intensity given by fluorescent reference beads; it is just above (Fig. 4.8B, red bars) the irreducible background due to the autofluorescence seen in mock-transfected cells (Fig. 4.8B; pink bar). I require foci to have intensities above a threshold just above the maximum signal given by this read-through (i.e., by the green dotted line in Fig. 4.8B).

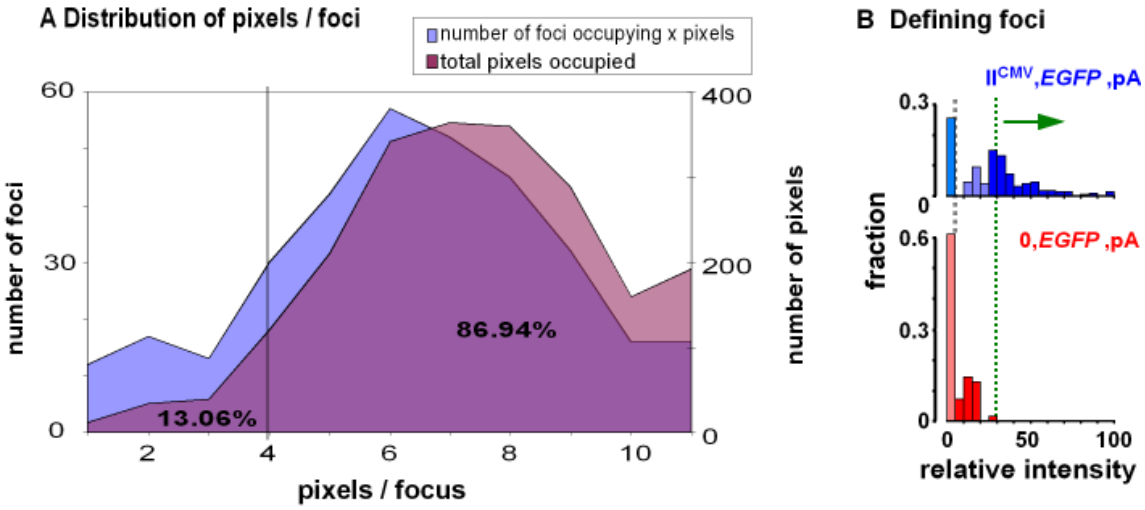


Fig. 4.8. Defining a focus

Plasmids were transfected, cells grown for 24 h, minichromosomal transcripts detected by RNA FISH, and images collected.

A. Cells were transfected with II^{CMV},*DsRed*,pA. The total number of foci in the 200 tested cells occupying 1-11 pixels (no focus is larger than 12 pixels) was binned (blue area); the total pixels occupied in each bin are shown (deep red area). More than 85% of the foci area was occupied by those larger than 4 pixels.

B. A threshold is set to discriminate between nuclear foci and background using cells transfected with promoter-less 0,*EGFP*,pA. Some *EGFP* signal is nevertheless detected by RNA FISH, and the average intensity of pixels in a focus (≥4 pixels) was measured relative to that of reference beads. The fraction of foci with relative intensities of 0-5, 6-10, etc. is indicated. Intensities of mock-transfected and untransfected cells lie in the first bin (i.e., to the left of the grey dotted line). Signal given by 0,*EGFP*,pA to the right of this grey line is due to read-through from the SV40 early promoter into *EGFP*. I set a threshold at the maximum due to this read-through (green dotted line), and with plasmids containing promoters (e.g., II^{CMV},*EGFP*,pA) only foci with intensities greater than this threshold are considered (green arrow).

4.4.3 Minichromosomal transcripts are concentrated in foci

After optimisation of RNA FISH and establishing algorithms for image analysis, I examined whether minichromosomes are transcribed in factories.

RNA FISH shows that both *DsRed* and *Neo*^{r} transcripts are found mainly in the cytoplasm. In more than 60% of the cells, both types of messages are concentrated in nuclear foci against a general nuclear background (Fig. 4.9A-D).

Such nuclear foci (which can only be seen clearly in the insets) mark the high concentration of nascent RNA at transcription sites, while the background represents completed transcripts on their way to the cytoplasm (Dirks et al., 1993; Levsky et al., 2002). Most RNA in such foci is nascent (i.e., still associated with engaged polymerases), as transcripts leave transcription sites quickly (Iborra et al., 1998).

Inspection of the two nuclear foci in the inset in Fig 4.9 A-D reveals that each contains red and green signal. Foci are deemed to colocalise if ≥25% of a focus of either colour also contains a focus of the other colour. 95% *EGFP* RNA foci colocalise with *DsRed* foci, and *vice versa* (Fig. 4.9D, bottom). As expected, these two transcription units – which are both on the same chromosome and driven by strong viral (pol II) promoters – are transcribed in the same place. I will not consider the few transcripts from the (down-regulated) SV40 promoter from now on, as a *Neo*^{r} probe is not used and any readthrough into the test gene is below our threshold.

A nucleus transfected with II^{CMV},*EGFP*,pA typically contains 23 ± 6 foci with an average area 7 ± 3 pixels (Fig. 4.8A). This number is equivalent to the number seen by DNA FISH (Fig. 4.5A).

I conclude a significant fraction of the plasmids are transcribed in foci.

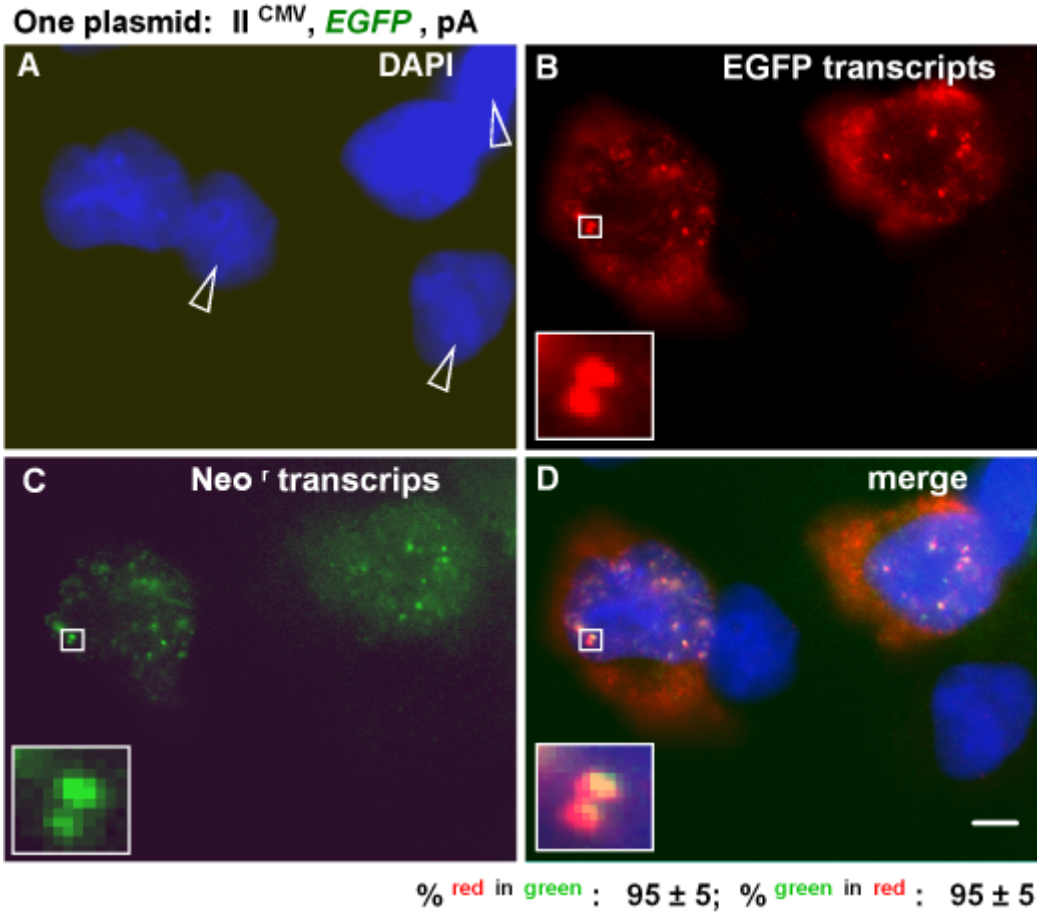


Fig. 4.9. Nascent minichromosomal RNAs are found in nuclear foci

Cells were transfected with II^{CMV},*DsRed*,pA, grown for 24 h, minichromosomal transcripts detected by RNA FISH, DNA counterstained with DAPI, and images collected. Two sets of four views of one field are shown. The percentage (±SD) of green foci that overlap red foci, and *vice versa*, is shown below. Bar: 5 µm.

A. Three of the five cells (arrowhead) were untransfected, and contain no *EGFP* or *Neo*^{r} transcripts.

B-D. Transfected cells contain many cytoplasmic *EGFP* transcripts, but few *Neo*^{r} transcripts. The nuclei contain some red and green foci (marking nascent RNA at transcription sites), against a general background (marking transcripts on their way to the cytoplasm). Inset: two foci with both red and green fluorescence; this result is expected, as the plasmid encodes both *EGFP* and *Neo*^{r}.

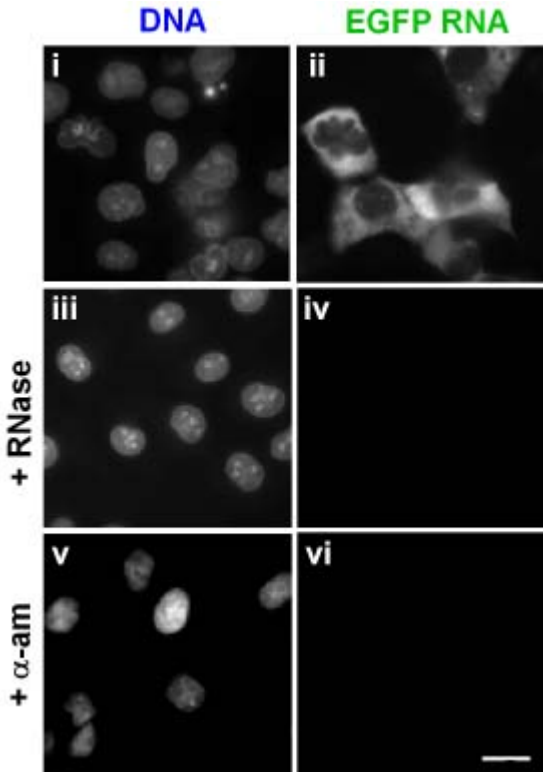


Fig. 4.10. RNA FISH negative controls

Cells were transfected with II^{CMV},*EGFP*,pA, incubated ± α-amanitin or ± RNase, DNA stained with DAPI, and *EGFP* transcripts detected by RNA FISH. Three sets of two views of one field are shown. RNase and α-amanitin abolish *EGFP* signal indicating that RNA made by pol II is detected in untreated cells. Bar: 10 μm.

4.4.4 Negative controls

In order to confirm the signals detected in my RNA FISH experiments are truly from RNA, I carried out two negative controls. In one, after fixation and before hybridisation, cells were treated with RNase. Neither cytoplasmic mRNA nor nuclear foci was seen after this treatment (Fig. 4.10 iii and iv), indicating the signals were from RNA, not DNA or other non-specific signals. In the other, I treated cells with an inhibitor of pol II – α-amanitin (Nguyen et al., 1996; Pombo et al., 1999). Similarly, neither cytoplasmic mRNA nor nuclear foci was seen after this treatment (Fig. 4.10 v and vi), indicating the RNA detected was produced by pol II.

4.5 Discussion

In this chapter, I established the method using replicating minichromosomes as probes to detect transcription factories.

First, using 'Southern' blots and qPCR, I quantified the copy number of minichromosomes and their transcripts. 24 h after transfection, minichromosomes replicate to give more than 8,000 new copies per cell (Fig. 4.1). Plasmids encoding different transcription units give similar numbers of DNA molecules (Fig. 4.2). However, promoter differences and the presence of an intron result in different transcript levels (Fig. 4.3), consistent with the variation in protein expression determined by FACS (Fig. 4.4). The copy number obtained (~8,000) is higher than $CaPO_4$-mediated transfection (~2,000) (Jackson and Cook, 1993), possibly due to the more immediate introduction of plasmid into cells. However, a significant fraction of the 8,000 copies may nevertheless be assembled into minichromosomes but still be transcriptionally inactive.

Second, I visualised minichromosome DNA using DNA FISH. Minichromosomes are concentrated in ~20 nuclear foci against a general background (Fig. 4.5). Although copy number increases with time, the number of foci remains constant, and the intensity inside foci increases, indicating the newly-replicated minichromosomes remain close to their site of synthesis. Quantification of the intensity of foci and the diffuse background revealed ~50% of minichromosomes reside in foci. Therefore, each of the ~20 foci can contain 200 minichromosomes. Several early studies support this finding. For example, (Nakamura et al., 1986; Kennedy et al., 2000) showed nascent DNA is made in a limited number of replication factories in the nucleus. Replicating adenoviruses (Pombo et al., 1994), as well as SV40 viruses (Dean, 1997), are found in clusters. Moreover, SV40 DNA was found in 20-30 foci in the nucleus after microinjection into the cytoplasm (Dean, 1997). Furthermore, Fackelmayer and colleagues found by FRAP that 70% plasmid microinjected into COS-7 cells was immobile, and this fraction was also resistant to salt extraction (Mearini et al., 2004). These findings suggest plasmids are replicated in factories after transfection.

Next, I determined the optimal RNA FISH conditions for the maximum detection sensitivity and specificity. Initially I observed two patterns of signal, one diffusely spread and one in discrete foci. As I used progressively harsher conditions, the fraction with the focal pattern increases. This is neither due to artificial aggregation nor subtraction of RNA targets, since the total intensity in foci is higher than that in the diffuse pattern (Fig. 4.7). I conclude the diffuse pattern arises from limited detection.

I demonstrated nascent minichromosomal RNA is concentrated in ~20 nuclear foci against a general nuclear background (Fig. 4.9). RNase and α-amanitin

controls showed that these RNA FISH signals were due to RNA made by pol II (Fig.4.10).

Although minichromosomal DNA and RNA were not visualised simultaneously, the number of transcription foci is nevertheless equivalent to the number seen by DNA FISH (Fig. 4.5 and 4.9). I conclude most minichromosomes are transcribed in factories. The diffuse background marks the soluble pool of minichromosomes, presumably the transcriptionally-inert fraction (Jackson and Cook, 1993). It is not clear if all the 200 copies in the cluster are transcriptionally active. If all are active, there should be at least 200 engaged pols in a focus. This number is far larger than 8, the average number of active molecules of pol II in a typical factory in human, mouse, and newt cells. The foci visualised here may be a cluster of several such factories, similar to the OPT domain which contains 3 factories (Pombo et al., 1998). After infection, many viral genomes are found to be replicated and transcribed at the periphery of specific nuclear domains (ND10), defined by accumulations of proteins that can be interferon-upregulated (Maul, 1998). ND10 may contain the necessary factors for viral replication, organising the adjacent transcription factories into a large cluster. In addition, active globin genes are organised and transcribed in an “active chromatin hub” (ACH) (Tolhuis et al., 2002), also indicating highly transcribed genes are clustered together to be transcribed, in a large transcription factory.

Do these plasmid replication sites colocalise with their transcription sites? Several early studies indicated that transcription and replication occur in the same sites (reviewed in (Hassan and Cook, 1994)). In this study, replicating plasmids and their nascent transcripts were clustered in similar numbers of bright nuclear foci. This supports the findings of (Hassan et al., 1994) that replication sites are

transcriptionally active and transcription sites are replication-competent. However, (Pombo et al., 1994) suggested replication sites of adenovirus were adjacent to, but do not perfectly colocalise with their transcription sites. The question remains open, and further experiments are needed (e.g. immuno-FISH localising proliferating cell nuclear antigen (PCNA) and transcripts simultaneously).

Finally, RNA FISH indicated two transcription units on the same type of plasmids are always transcribed in the same foci (Fig. 4.9). However, we can not conclude whether the two types of transcripts are from the same minichromosome or different individuals. Previous studies suggested that transcription of a gene can interfere with transcription of another linked gene (Emerman and Temin, 1984; Eszterhas et al., 2002). For example, introduction of a single copy of a bicistronic provirus bearing two adjacent selective markers, each with its own promoter, leads to cells resistant to one or the other selective agent but rarely both (Emerman and Temin, 1984). In this case, *EGFP* and *Neo*^{r} transcription units are on the same plasmid (II^{CMV},*EGFP*,pA), and so are unlikely to be active at the same time. Therefore, different transcription units controlled by viral CMV or SV40 early promoter can be transcribed in the same factories.

In the next chapter, I go on to investigate whether minichromosomes encoding different transcription units are transcribed in the same or different factories.

5 Promoters target genes to specific transcription factories

5.1 Introduction

After transfecting replicating plasmids into COS-7 cells, most of the resulting minichromosomes are transcribed in factories. As reviewed in Ch.1.1, different genes possess different promoters and processing signals. Since much processing occurs co-transcriptionally, appropriate machineries are required to be recruited during transcription. Do all factories contain the same necessary components or do they specialise in transcribing different genes?

Previous studies indicated pols I, II, and III have their own dedicated transcription factories (Pombo et al., 1999). In mammalian cells, active pol I is exclusively located in the nucleolus while active forms of pols II and III are both present in the nucleoplasm. Does each factory contain a mixed population of pol II and III, or is each enzyme found in a different dedicated factory? If pol II and III are present in the same factory, after inhibiting pol II with α-amanitin, the number of foci (detected after immunolabelling nascent RNA) should remain the same while the intensity of foci falls. However, if pol II and III are found in separate factories, site number will fall. Result showed they are found in separate factories (Pombo et al., 1999). Furthermore, an antibody against pol II_O does not block access to nascent Br-RNA made by pol III, and *vice versa*. This result confirmed that active forms of pol II and III are concentrated in spatially separated and dedicated factories (Pombo et al., 1999).

In this chapter, I will test if promoter affects choice of transcription factory using a different method (RNA FISH).

5.2 Similar promoters are targeted to the same factories

I first examined two minichromosomes identical except for coding region – II^{CMV},*EGFP*,pA and II^{CMV},*DsRed*,pA (Fig. 5.1E-H). As before, transfected cells contain high concentrations of mature mRNA of both types in the cytoplasm, as well as ~20 nuclear foci against the nuclear background. 78% *DsRed* (nascent RNA) foci colocalise with *EGFP* foci, and *vice versa*; minichromosomes with identical promoters are transcribed in the same factories.

As most II^{CMV},*EGFP*,pA and II^{CMV},*DsRed*,pA transcription sites colocalise, from now on, II^{CMV},*DsRed*,pA is used as a reference plasmid to compare with various plasmids encoding EGFP.

Since 80% of the two plasmids sequences are identical, the colocalisation might result from "homologous pairing" driven by base pairing (Kleckner, 1996). A recent study using non-replicating plasmids encoding globin genes also made a similar argument (Binnie et al., 2006). If this argument is correct, then substituting the promoter region (~200 bp) with a different promoter – which does not significantly affect homology – will not change the localisation of the two plasmids.

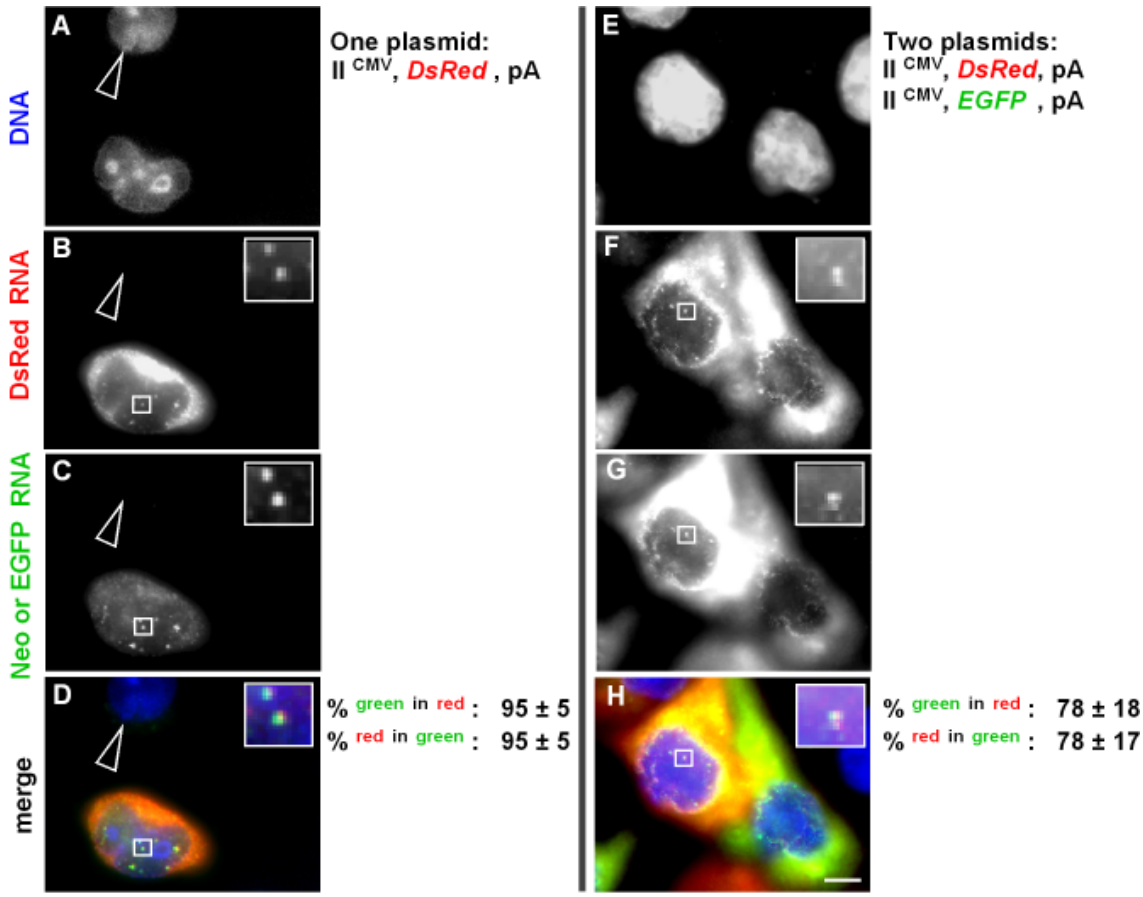


Fig. 5.1. Similar promoters target minichromosomes to similar factories

Cells were transfected, grown for 24 h, minichromosomal transcripts detected by RNA FISH, DNA counterstained with DAPI, and images collected. Two sets of four views of one field are shown. Percentages (±SD) indicate green foci that overlap red foci, and *vice versa*. Bar: 5 μm.

A-D. Transfection only with II^{CMV},*DsRed*,pA. One cell (arrowhead) was untransfected. The other nucleus contains some red (*DsRed*) and green (*EGFP*) transcription foci against a general background. Inset: two foci with both red and green fluorescence.

E-H. Co-transfection with II^{CMV},*DsRed*,pA + II^{CMV},*EGFP*,pA; these differ solely in coding region. The two central cells are transfected; they contain *DsRed* and *EGFP* RNA mainly in the cytoplasm, and some in nuclear foci against a general background. Inset: the nuclear focus contains both types of RNA.

5.3 Polymerase I, II, and III units are transcribed in different factories

Here I substitute only the promoter region of II^{CMV},*EGFP*,pA with a pol I or III promoter. Few transcripts copied using the pol I promoter in I^{45S},*EGFP*,pA are found in prominent nucleoli (Fig. 5.2C) or the cytoplasm; instead, most are found in 38 ± 13 large nuclear foci (area 30 ± 13 pixels). However, few of these nuclear foci colocalise with those containing nascent RNA encoded by II^{CMV},*DsRed*,pA (Fig. 5.2 A-D). Similarly, few transcripts copied from a pol III unit are cytoplasmic (Fig. 5.2G), and nuclei typically contain 16 ± 5 small foci (area 6 ± 3 pixels). Few of these again colocalise with pol II foci (Fig. 5.2E-H). These results indicate active forms of the three nuclear polymerases are each concentrated in their own dedicated factories, consistent with possibility iii in Fig. 3. This confirms earlier results (Pombo et al., 1999).

Pol I transcribes ribosomal cistrons in nucleoli (Ch.1.1.2). Therefore, one might expect transcripts expressed from a pol I promoter to be concentrated in nucleoli. However, most of the pol I minichromosomal foci are nucleoplasmic (Fig. 5.3 E-H). I further investigate the pol I foci using immuno-FISH. Immuno-labelling of UBF, a pol I transcription factor, revealed discrete foci in both nucleoli and the nucleoplasm (Fig. 5.3F). The nuclear pol I transcripts often colocalise with UBF (Fig. 5.3H), suggesting these foci are functionally active. This result is consistent with another study, where artificial arrays of heterologous UBF-binding sequences at ectopic sites on human chromosomes recruit UBF to sites outside the nucleolus, and form pseudo-NORs; these are morphologically similar to the NORs seen during metaphase (Mais et al., 2005).

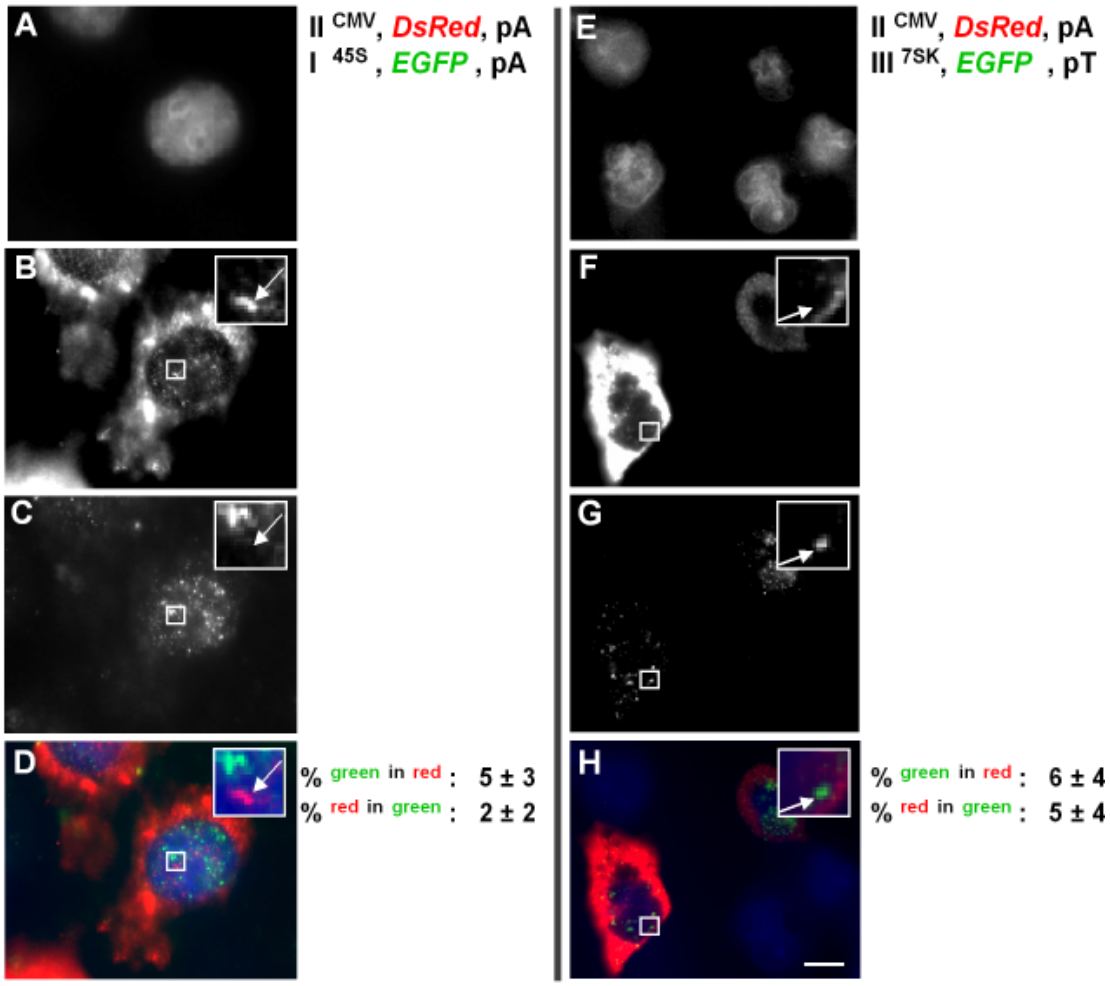


Fig. 5.2. Promoters can target minichromosomes to specific factories

Cells were co-transfected with II^{CMV},*DsRed*,pA and the plasmid indicated, grown for 24 h, transcripts detected by RNA FISH, and DNA counterstained with DAPI. Four views of one field for each pair are shown. Percentages (±SD) indicate green foci that overlap red foci, and *vice versa*. Arrows: non-colocalising foci. Bar: 5 μm.

A-D. Pol I and II promoters target minichromosomes to different factories. The one transfected cell contains *DsRed* (but not *EGFP*) RNA in the cytoplasm, and the nuclear foci do not colocalise. Inset: foci do not colocalise (confirmed by low percentages).

E-H. Pol II and III units are targeted to different factories. The two transfected cells contain both types of RNA in cytoplasm and nucleus. Inset: foci do not colocalise (also low percentages).

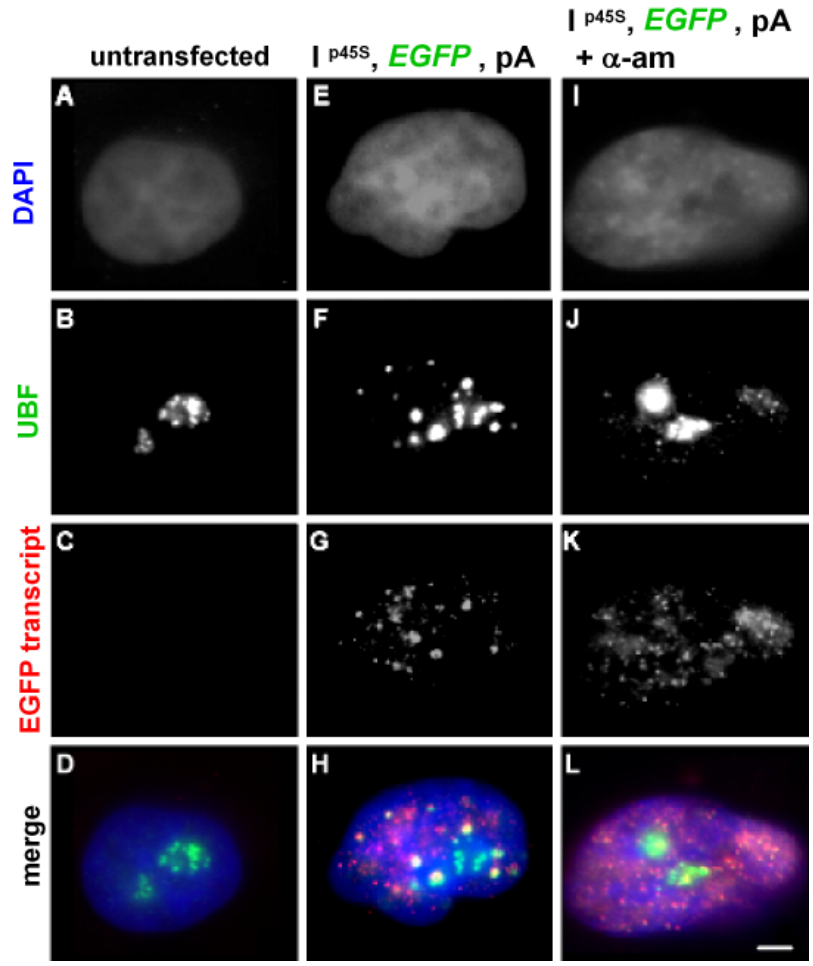


Fig. 5.3. Transcripts made from a polymerase I promoter colocalise with UBF

Cells transfected ± I^{45S},*EGFP*,pA were grown ± α-amanitin, *EGFP* transcripts and UBF detected by immuno-FISH, and DNA stained with DAPI. Four views of three fields are shown. Bar: 5 µm.

A-D. In untransfected cells, UBF is concentrated in nucleoli.

E-H. In transfected cells, UBF is found in additional nucleoplasmic sites that often contain *EGFP* RNA.

I-L. α -amanitin has little effect on *EGFP* RNA, as expression is not controlled by a pol II promoter.

To exclude the possibility that *EGFP* RNA was initiated from a cryptic pol II promoter, I inhibited pol II with α-amanitin. RNA FISH results showed α-amanitin had little effect on the level of *EGFP* transcripts, confirming that they were made by pol I (Fig. 5.3I-L). However, α-amanitin appears to disrupt the pol I foci into many smaller foci.

5.4 Discussion

At the beginning of this chapter, I showed using co-transfection and RNA-FISH that nascent minichromosomal RNAs made from the same promoters driving different genes were nevertheless found in the same transcription foci (Fig.5.1).

Binnie et al. argued that homologous gene sequences mediate 'transcription domain' formation (Binnie et al., 2006), but homologous promoter sequences are not required for this formation. However, I substituted one pol II promoter region (~200 bp) with a pol I or III promoter and found the two plasmids no longer colocalised. Thus, homology between the two transcription units is unlikely to be responsible for the nuclear colocalisation; rather, it is functioning promoters that determine the choice of transcription factory and so colocalisation.

In addition, I confirmed previous antibody blocking and transcription inhibition results. Pol II and III transcription units are transcribed in different factories (Fig. 5.2), strongly supporting hypothesis (iii) in Fig.3.1.

The finding that active forms of pols I, II, and III are found in their own specialised factories can help explain why active pol III units can silence pol II units. Transcription of $tRNA^{Thr}$ (by pol III) silences *URA3* (which is transcribed by pol II); transcription of many other yeast tRNA genes also reduces the transcription of pol II

units lying within ~400 bp between three and fifty-fold (Bolton and Boeke, 2003). Moreover, transcription of tRNA genes in centromeres inhibits transcription of adjacent pol II units, while mutating the pol III promoter abolishes this effect (Scott et al., 2006). We suggest that when pol II genes are flanked by a cluster of pol III genes (e.g., encoding tRNAs), the active *tRNA* genes drive the locus into a pol III factory in order to be transcribed; consequently, the neighbouring pol II genes may be pulled out of pol II factories, and so silenced. Then, when pol III promoters are deleted, the ties between the tRNA genes and the pol III factories disappear, so that the pol II genes can now attach to a pol II factory and be transcribed (Fig.5.4).

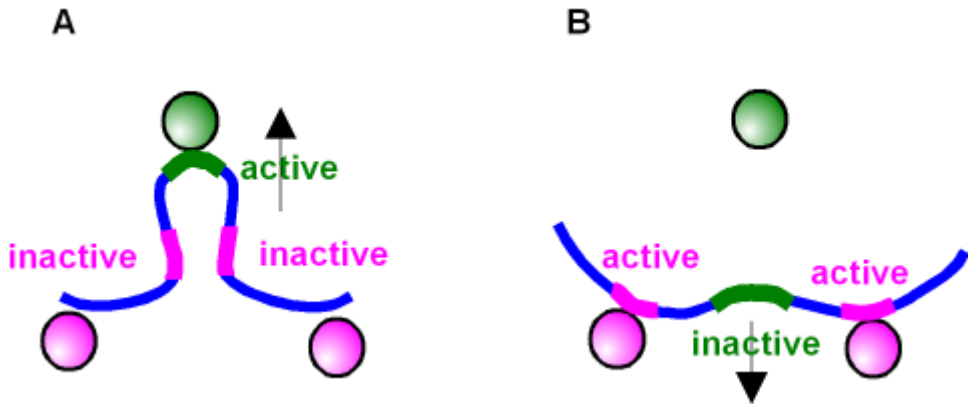


Fig. 5.4. Pol III units can silence pol II units

A pol III transcription unit (green) lies adjacent to two pol II units (purple). Pols II and III have their own specialised transcription factories (purple and green balls).

A. When the pol III unit is transcriptionally active, it attaches (arrow) to the pol III factory. The two neighbouring pol II units are now distant from the pol II factories, and so silenced.

B. When the pol III unit is inactive (e.g., because of a mutation in its promoter), it detaches (arrow) from the pol III factory. The two neighbouring pol II units are now able to reach the pol II factories to be transcribed.

6 Processing signals and introns target genes to specific factories

6.1 Introduction

In the previous chapter, I confirmed that different pol I, II, and III units are transcribed in their own dedicated factories. As different pol II transcripts are being made, they are co-transcriptionally processed in different ways. For example, some are capped and polyadenylated, others are not; some have introns that have to be spliced, others do not; non-coding RNAs also require extensive modifications (Ch.1.1.1). Do transcription factories further specialise in transcribing different groups of pol II genes? Various evidences suggest they do (Ch.1.4.6).

In this chapter, I continue with the same strategy and methods. As ~80% minichromosomes encoding a CMV pol II promoter (regardless of whether it drives *EGFP* or *DsRed*) are transcribed in the same ~20 foci (Fig. 5.1), I use the same reference plasmid (II^{CMV},*DsRed*,pA) in all co-transfections. I test if processing signals affects choice of transcription factory.

6.2 *RNU2* minichromosomes are transcribed in special factories

I first compared two very different pol II units. The *U2G* transcription unit contains a functional U2 promoter including a PSE and DSE and its own enhancer (Cuello et al., 1999; Smith and Lawrence, 2000; Matera et al., 2007); this makes it quite different from most pol II promoters driving protein-coding genes. U2 sequences from +28 to +169 were replaced with β-globin exon 3 sequences (from

151 to 7 bp upstream from the poly(A) site) to enable me to distinguish plasmid sequences from those in endogenous U2 genes (Medlin et al., 2003). The plasmid also contains the full 3' processing sequences of the U2 gene (Cuello et al., 1999).

I co-transfected a reference plasmid II^{CMV},*DsRed*,pA – which encodes a "standard" intron-less unit – and the test plasmid II^{U2},*U2G*,3'box (Fig. 6.1A-D). After co-transfection, most "standard" (*DsRed*) transcripts are cytoplasmic, while most *U2G* transcripts are nuclear. This is expected; "standard" messages are exported to accumulate in the cytoplasm, while *U2G* transcripts are detected using a probe complementary to sequences in the pre-U2 RNA that are removed rapidly and degraded (Medlin et al., 2003).

Most nuclear RNA of both types is again found in foci, but now <6% foci colocalise (Fig. 6.1A-D); *U2G* units are transcribed in different factories from CMV units, consistent with possibility (iii) in Fig. 3.1. There are 23 ± 6 and 15 ± 4 *DsRed* and *U2G* nuclear foci of similar area (7 ± 3 pixels), respectively.

This experiment was repeated using II^{CMV},*DsRed*,pA and II^{U2p-},*U2G*,3'box – the only difference being the replacement of 12 bp in the U2 promoter with irrelevant sequence that destroys promoter activity (Cuello et al., 1999); no U2G transcripts are now detected (not shown), so those seen previously cannot be produced from a cryptic promoter.

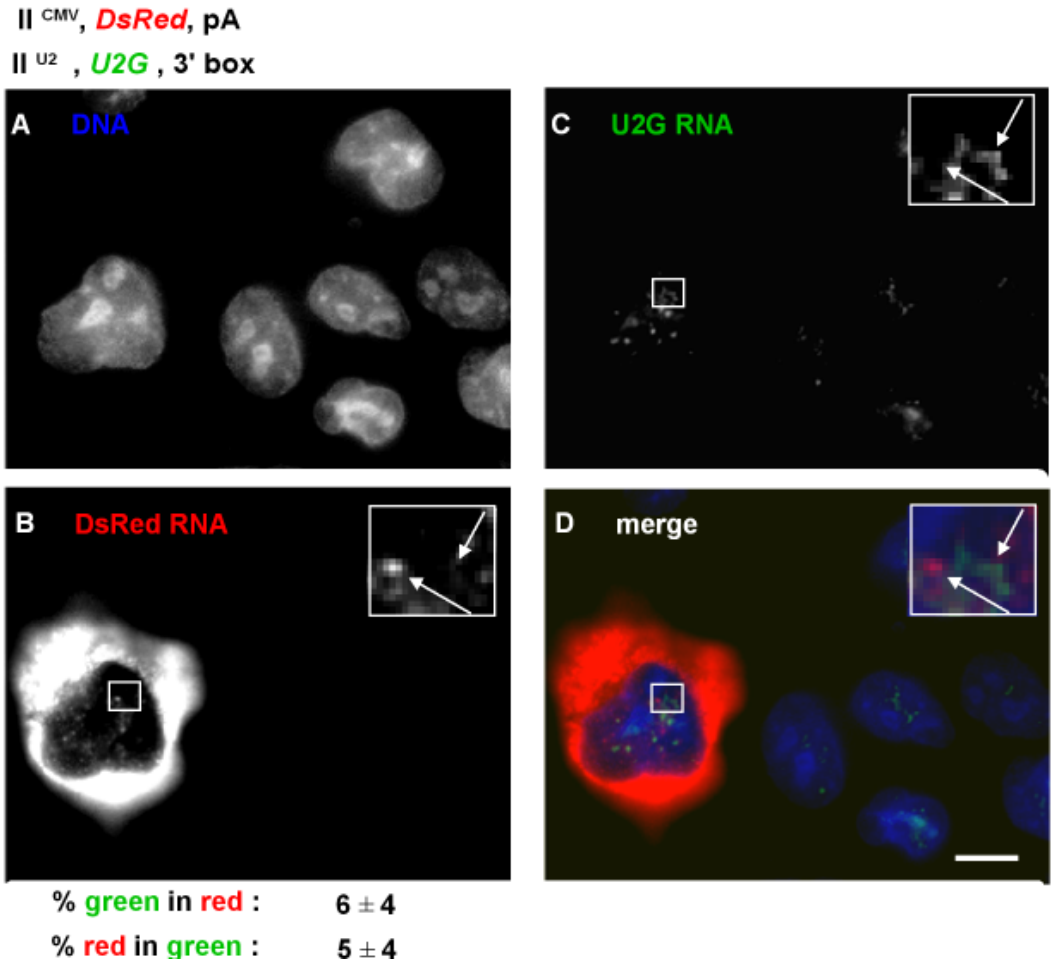


Fig. 6.1. Different pol II units target different factories

Cells were co-transfected with II^{CMV},*DsRed*,pA and II^{U2},*U2G*,3'box, grown for 24 h, transcripts detected by RNA FISH, DNA counterstained with DAPI, and images collected. **A**. Four views of one field are shown. Below: percentages (±SD) of green foci that overlap red foci, and *vice versa*. Arrows: non-colocalising foci. Bar: 5 µm.

B. One transfected cell contains *DsRed* transcripts in the cytoplasm, and in nuclear foci.

C. Pre-*U2* transcripts (host and minichromosomal) are detected by the pre-U2 probe (Smith and Lawrence, 2000). They are found in nuclear foci in both transfected and untransfected cells.

D. Merge of B and C. Inset: nuclear foci do not colocalise (confirmed by low percentages below).

6.3 Introns target genes to special factories

Next I examined if introns also affect localisation. II^{CMV},*EGFP*$_V$,pA, is essentially similar to II^{CMV},*EGFP*,pA, the only difference being the addition of the globin intron 2 to the end of the *EGFP* unit. Expression levels of the two plasmids after introduction into cells is different; quantitative RT-PCR and FACS show that transcripts from the intron-containing unit and the resulting protein are present in higher quantities (Fig. 4.3 and 4.4, see also (Brinster et al., 1988)).

After co-transfection with the reference plasmid II^{CMV},*DsRed*,pA, transcripts copied from II^{CMV},*DsRed*,pA and II^{CMV},*EGFP*$_V$,pA are both found mainly in the cytoplasm (Fig. 6.2B-D). Both intron-containing and intron-less nascent RNA are also found in ~20 nuclear foci of similar size and intensity. There are 21 ± 4 and 23 ± 6 intron-containing and intron-less nuclear foci of 6 ± 3 and 7 ± 3 pixels, respectively.

But although flanked by the same promoter and 3' signals, ~80% of them are found in different factories, and only ~20% colocalise (Fig. 6.2B-D).

If intron-encoding minichromosomes are transcribed in specialised factories, then those factories must contain the necessary splicing machineries. I therefore carried out immuno-FISH to visualise SC-35 (a splicing factor) and *EGFP* RNA simultaneously.

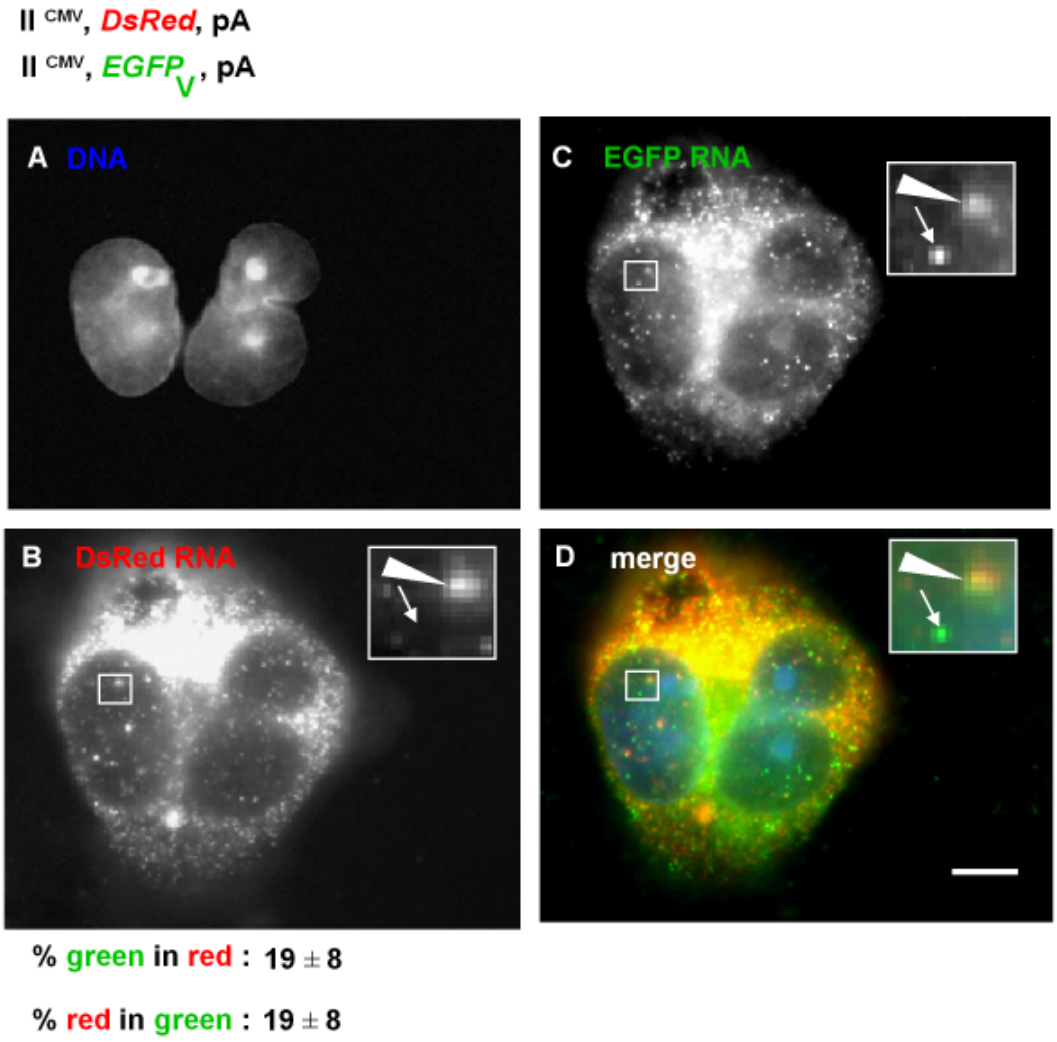


Fig. 6.2. Introns can target minichromosomes to specific factories

Cells were co-transfected with II^{CMV},*DsRed*,pA and II^{CMV},*EGFP*$_V$,pA, grown for 24 h, transcripts detected by RNA FISH, and DNA counterstained with DAPI.

A. Four views of one field are shown. Below: percentages (±SD) of green foci that overlap red foci, and *vice versa*. Arrows: non-colocalising foci. Bar: 5 µm.

B-D. Both cells contain *DsRed* and *EGFP* transcripts in cytoplasm and nucleus. Inset: upper focus contains both types, the lower one only *EGFP* (infrequent colocalisation is reflected by low percentages below).

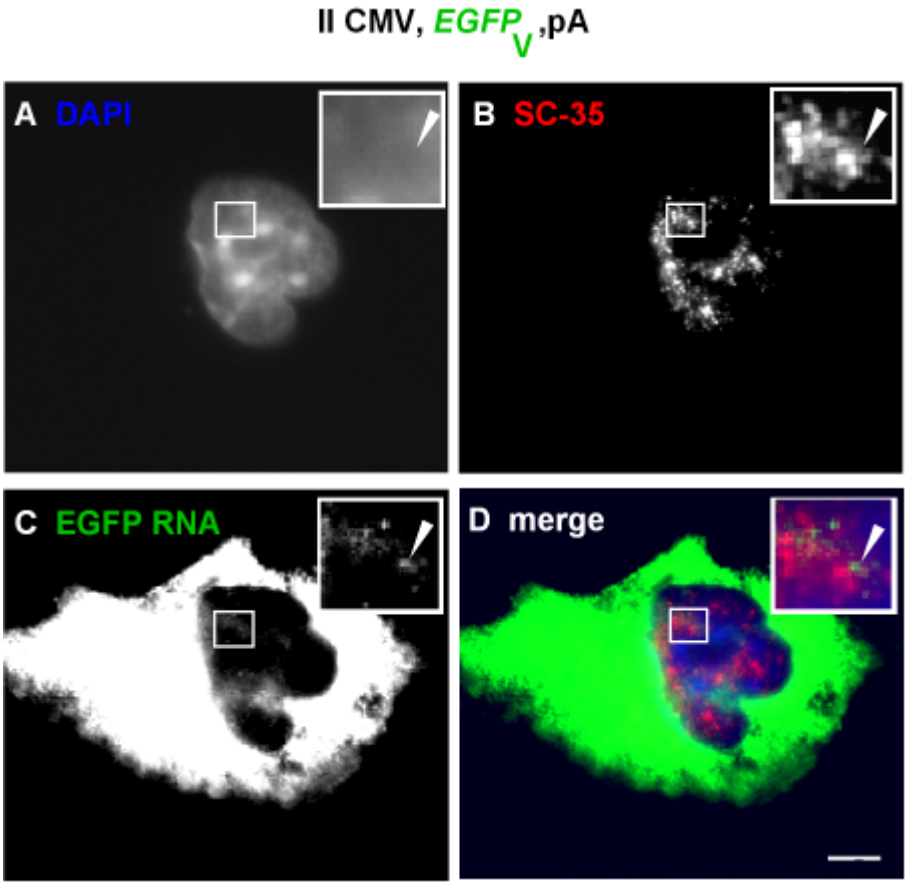


Fig. 6.3. Intron-containing RNA colocalises with faint SC-35 foci

Cells were transfected with II^{CMV},*EGFP*$_V$,pA, grown for 24 h, *EGFP* transcripts and splicing factor SC-35 detected by immuno-FISH, and DNA stained with DAPI. Four views of one field are shown. Inset: *EGFP* transcripts (arrowhead) colocalise with the many faint speckles containing SC-35, but not the few bright ones. Bar: 5 µm.

As expected, SC-35 is diffusely spread in the nucleus, as well as concentrated in a few discrete large “speckles”; it is not found in nucleoli (Lamond and Spector, 2003; Xie et al., 2006). Intron-containing foci contain some SC-35, but they do not generally colocalise with the brightest “speckles” (Fig. 6.3); instead they are often found at the periphery of speckles. This result confirms speckles are not active sites of transcription (Xie et al., 2006); they are more likely to be involved in the storage of splicing components. Factories located close to speckles may then have easier access to these stores.

6.4 Discussion

In this chapter, using co-transfection and RNA-FISH, I found transcription factories containing pol II further specialise.

First, minichromosomal *U2* transcription units are transcribed in different factories from those with CMV promoters (Fig. 6.1), although both promoters are pol II specific. This result suggests that promoter recognition plays an important role in targeting transcription units to the correct factories. Presumably the PIC recognises some ‘code’ in the promoter containing the information about the specific factors needed during transcription and/or processing. If the minichromosome binds to the ‘wrong’ factory, we imagine that the PIC is unstable and productive transcription does not occur; through re-iterative binding and dissociation, the transcription unit will eventually reach the ‘correct’ factory where it can initiate productively. This hypothesis is supported by an earlier study using artificial tandem arrays, where *U2* genes are often found associated with Cajal bodies, and a mutation on the promoter abolish this colocalisation (Frey et al., 1999). Since Cajal bodies are

rich in factors involved in snRNP biogenesis (Ch.1.2.4), factories located close to Cajal bodies ensure a rapid exchange of the necessary factors to process U2 transcripts.

Next, I showed the presence of an intron in the minichromosome also affects localisation. Most minichromosomes with an intron (i.e., 80%) are transcribed in different factories from those with the same promoters and transcription units but without an intron (Fig. 6.2). We now know that splicing is a co-transcriptional event, and that the phosphorylated pol II CTD is responsible for co-regulating transcription and processing (Proudfoot et al., 2002). If introns communicate with promoters through looping on transcriptional initiation, this might stabilise binding of gene locus to factories containing the appropriate processing factors.

Where are the intron-specific factories located? Some transiently transfected HeLa cells contained large clusters of plasmids bearing an intron, which perfectly colocalise with SC-35 domains (Huang and Spector, 1996); while those lacking an intron do not. However, my immuno-FISH result suggested the transcription sites were adjacent to, but do not colocalise with, splicing speckles (Fig. 6.3); this is in accord with a recent study that splicing speckles are not sites of transcription (Xie et al., 2006).

An interesting question remains. How does the polymerase ‘know’ whether the transcription unit contains an intron before transcription takes place? I hypothesise here that minichromosomes carry a certain type of “mark” that directs them to specific factories (Fig. 6.4). Newly-synthesised minichromosomes encoding an intron are unmarked (left). But once transcribed in the ‘correct’ factory (above), the intron is spliced, and the transcribed unit becomes marked (right), ensuring it targets the correct factory during the next transcription cycle. If the minichromosome

reaches the wrong factory (below), abortive transcription occurs; the transcription unit remains unmarked and returns to the unmarked pool (left).

The marking process probably occurs slowly so that transcripts generated on binding to the 'wrong' factory can be detected (i.e., before they are degraded, or exported). This may explain the ~20% colocalisation seen between minichromosomes with and without an intron. Presumably, the "marks" are epigenetic codes (i.e., certain combinations of histone modifications). So far, we know little about histone modifications in intronic regions.

Another explanation of introns targeting specific factories is related to the regulatory elements in the intron. The *HBB* intron 2 sequences contain an intragenic enhancer and a separate potential MAR element (Rubin et al., 2000; Carter et al., 2002). The transcriptional regulatory elements could potentially direct the intron-containing minichromosome to special factories even though it has an identical promoter to the intron-less construct. To conclude that the presence of an intron *per se* is important in transcription localisation, an intron which does not contain regulatory elements should be used as a control. However, the presence of such an intron is worth questioning. Many introns may contain regulatory elements that remain to be identified.

In conclusion, pol II factories further specialise in transcribing genes containing introns, and non-coding U snRNAs. Transcription units must associate with the correct factory before they can be transcribed.

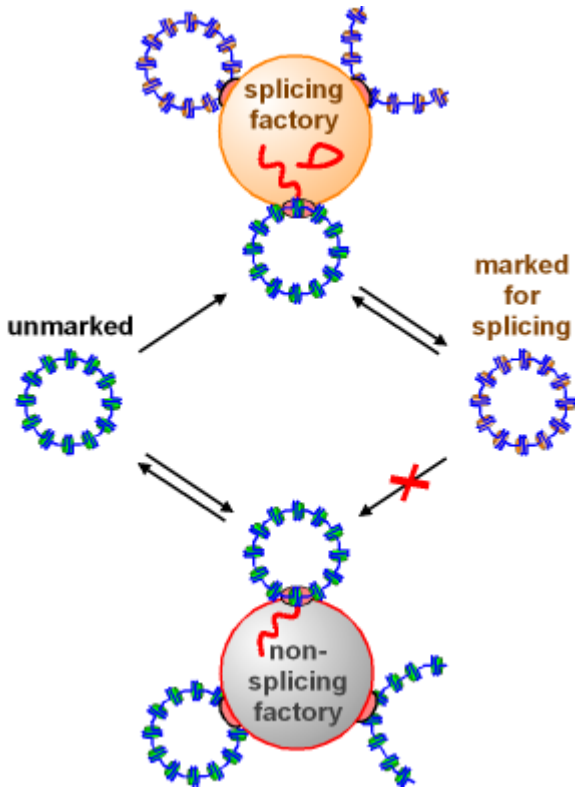


Fig. 6.4. Model: introns target minichromosomes to specific factories

Newly-synthesised minichromosomes are unmarked (left, green nucleosomes). After transcription in the correct factory (above), the intron is spliced, and the transcribed unit marked (right, orange nucleosomes) to ensure it targets the correct factory during the next transcription cycle. If the minichromosome goes to the wrong factory (below), it is transcribed without becoming marked, and so will return to the unmarked pool (left).

7 Minichromosomes with similar promoters are in close proximity

7.1 Introduction

Using RNA FISH, I demonstrated that promoters and the presence of an intron strongly affected nuclear location of minichromosomes (Ch.4 and 5). In this chapter, I use chromosome conformation capture (3C) to verify the FISH result.

3C has been proven a powerful method to investigate chromatin structure *in vivo* (Dekker et al., 2002). The recent advancement of 3C enables a genome wide view of inter- and intra- chromatin interactions (de Laat, 2007; Ohlsson and Gondor, 2007). However, 3C has hardly been used for minichromosomes. Here I first established a ChIP-3C method to evaluate minichromosome proximity. Then I went on to test whether minichromosomes with the same promoters are in closer proximity than the ones with different promoters. Finally I examined if minichromosomes encoding a specific gene (i.e. *RNU2*) lie close to the endogenous gene locus.

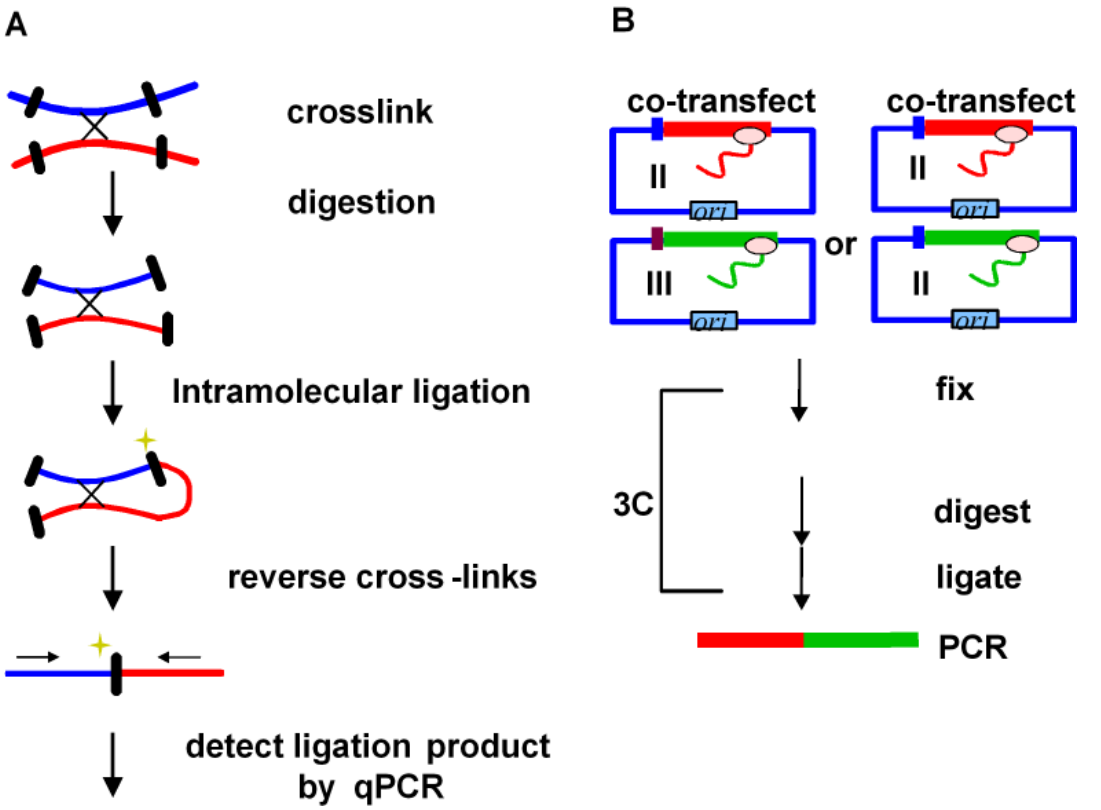


Fig. 7.1. Minichromosome proximity detected by 3C

A. Conventional 3C method. Chromatin is first cross-linked by formaldehyde. After restriction digestion, chromatin is diluted so that intramolecular ligation is favoured. Specific ligation product is detected using qPCR after reversal of the cross-links. The yellow mark indicates restriction sites. Adapted from (Dekker et al., 2002).

B. Detecting proximity of minichromosomes with different promoters using 3C. Two separate co-transfections were carried out: in one, II^{CMV},*DsRed*,pA + II^{CMV},*EGFP*,pA and in another II^{CMV},*DsRed*,pA + III^{7SK},*EGFP*,pT. Standard 3C procedures performed as in A.

7.2 Establishing the ChIP-3C method for plasmids

7.2.1 Preliminary 3C

A scheme of the conventional 3C method is shown in Fig. 7.1A. Dilution of the cross-linked chromatin after restriction digestion is critical. It is essential that the dilution is sufficient to favour an intramolecular ligation in the next step.

I wish to compare if minichromosomes with similar promoters are in closer proximity than the ones with different ones. Therefore, I co-transfected cells with plasmids encoding DsRed and EGFP, either driven by two similar pol II promoters or a pol II and III promoter. I then used 3C to test how close *DsRed* genes are to *EGFP* genes (Fig. 7.1B).

Preliminary experiments revealed that *DsRed* lay significantly closer ($p=0.0321$, unpaired *t*-test) to *EGFP* in a II + II combination (i.e., II^{CMV},*DsRed*,pA + II^{CMV},*EGFP*,pA), than a II + III combination (II^{CMV},*DsRed*,pA + III^{7SK},*EGFP*,pT) (Fig.7.2).

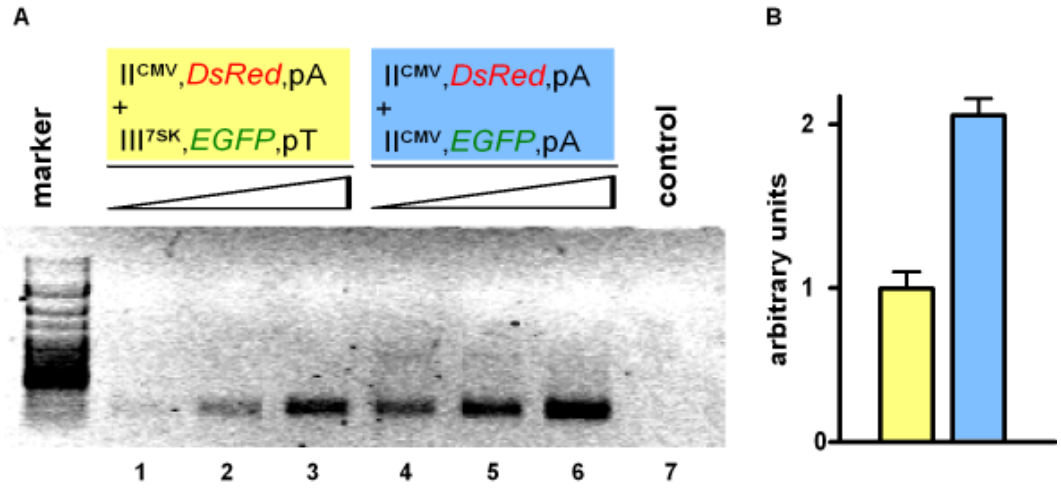


Fig. 7.2. Minichromosomes with similar promoters lie close together

Cells were transfected with (II^{CMV},*DsRed*,pA + III^{7SK},*EGFP*,pT) (yellow) or (II^{CMV},*DsRed*,pA + II^{CMV},*EGFP*,pA) (blue), fixed after 24 h, lysed, and digested with *Bam*HI, ligated, DNA purified, and PCR performed using specific primers.

A. Dilution series of 3C products. Image illustrates gels containing amplimers. Mock-transfected cells served as a negative control (lane 7). More *DsRed*:*EGFP* 3C products are seen when promoters are identical (lanes 4-6; compare with lanes 1-3).

B. Quantification of A. *DsRed*:*EGFP* 3C product between two minichromosomes with identical promoters (blue) is two fold that found with different promoters (yellow).

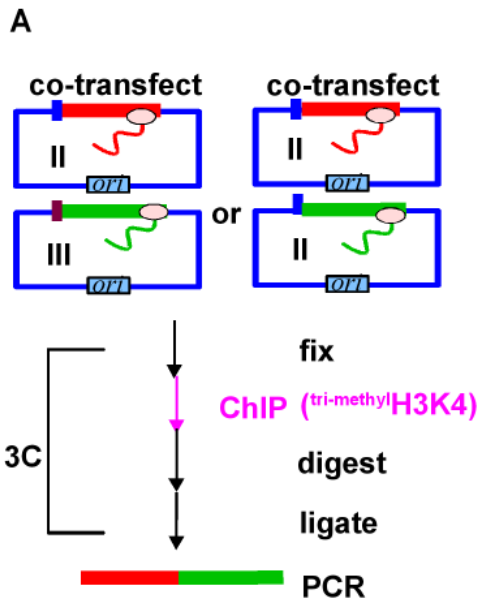
7.2.2 ChIP-3C

As inactive DNA (i.e., naked input and/or minichromosomes in the diffuse nuclear background) probably masks even larger differences, I first select the transcriptionally-active fraction by chromatin immunoprecipitation (ChIP).

The ChIP-3C (Cai et al., 2006) is shown in Fig. 7.3A. Tri-methyl K4 in histone H3 marks the actively-transcribed chromatin (Barski et al., 2007). I added a ChIP step after fixation using a $^{\text{tri-methyl}}$H3K4 antibody, before the active fraction of chromatin (from both endogenous and minichromosomes) is selected using protein G beads. As the stringency of washes directly affects the performance of ChIP, I analyse both the supernatant and pellet.

7.2.3 3C controls

Although the 3C method is in principle straightforward, it nevertheless requires carefully designed controls (Dekker, 2006). The final result of 3C is represented by a relative crosslinking frequency, which corrects for differences in PCR efficiencies, crosslinking and ligation efficiencies, amounts of template, and size of PCR fragments (Fig. 7.3B). All controls are described below.



B

$$[X(\text{mini})] = \frac{[\text{A(mini)} / \text{A}(ALDOA)]\text{cell}}{[\text{A(mini)} / \text{A}(ALDOA)]\text{control} \times \text{A}(Neo^r)}$$

Fig. 7.3. ChIP-3C

A. Detecting proximity of minichromosomes with different promoters using ChIP-3C. Transfections were carried out as in Fig. 7.1B. A ChIP step was added after fixation before restriction digestion to reduce input background.

B. Equation used to calculate relative crosslinking frequency between two minichromosomes [*X*(mini)]. A(mini): intensity of PCR signal (determined with ImageJ) obtained with a given *DsRed*:*EGFP* ligation product. A(*ALDOA*): intensity of PCR signal obtained with the *ALDOA:ALDOA* ligation product. A(mini) and A(*ALDOA*) are determined for both *in vivo* and the PCR control *in vitro*. A(Neo^r): intensity of PCR signal obtained with Neo^r on the common backbone. This controls for plasmid copy number. The calculation gives a relative crosslinking frequency for each sample since it corrects for differences in PCR amplification efficiencies, crosslinking and ligation efficiencies, amounts of template, and size of PCR fragments (Tolhuis et al., 2002).

7.2.3.1 Internal control

The housekeeping gene *ALDOA* (aldolase A; Macaca mulatta; MMUL 1.0, Feb 2006; Chr20 at location 28,031,923-28,050,811) is used as an internal control to validate the 3C method. The primers were designed to flank more than two restriction sites covering > 2,000 bp. If the method is valid, a PCR product of an expected size (~400 bp) is produced.

7.2.3.2 PCR efficiency control

Control ligation products were prepared by mixing equimolar amounts of multiple plasmids containing the restriction sites. The mixture was digested with *Bam*HI or *Pst*I and ligated to generate a library of randomly-ligated control fragments. The specificity of the 3C primers was tested by PCR using 2 ng of the randomly-ligated control fragment with 200 ng of genomic DNA for 30 cycles. Plasmid pTP18 (Pavelitz et al., 1995) and a BAC (RP-11-114A14, BACPAC Resources Center, Oakland, CA) were used as sources of host *U2* and *ALDOA* DNA respectively.

7.2.3.3 Copy number control

The number of plasmids per cell was also taken into account, in order to eliminate differences between each transfection. Copy numbers are normalised to the PCR product given by Neo^r, encoded on the common backbone of all plasmids.

7.3 Minichromosomes with similar promoters target the same factories

I next use ChIP-3C to compare how close *DsRed* genes are to *EGFP* genes in cells co-transfected with plasmids encoding either two similar pol II promoters (i.e., II^{CMV},*DsRed*,pA + II^{CMV},*EGFP*,pA) or a pol II and III promoter (II^{CMV},*DsRed*,pA + III^{7SK},*EGFP*,pT). *DsRed*:*EGFP* 3C products are seen in both supernatant and pellet (Fig. 7.4, rows 1,3), consistent with the presence of inactive DNA in the supernatant. More such products are seen in the pellet (the active fraction) when promoters are similar (row 3). In contrast, *ALDOA* is an active (endogenous) housekeeping gene and no *ALDOA*:*ALDOA* 3C products are seen in the supernatant (row 2); moreover, pellets from both combinations yield similar levels (row 5). The average of 4 independent experiments confirms the II + II combination gives 6-fold more *DsRed*:*EGFP* 3C product than the II + III combination (row 7). I conclude that minichromosomes bearing similar promoters are more likely to be in close proximity. This strongly supports the conclusion from RNA FISH – transcription factories specialise in transcribing genes controlled by different promoters.

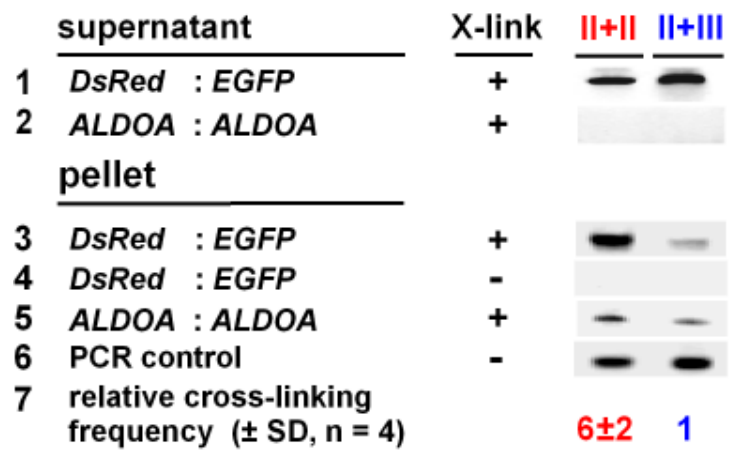


Fig. 7.4. ChIP-3C: minichromosomes with similar promoters lie together

Cells were co-transfected with II^{CMV},*DsRed*,pA + II^{CMV},*EGFP*,pA (II + II) or II^{CMV},*DsRed*,pA + III^{7SK},*EGFP*,pT (II + III), grown for 24 (rows 1-7), treated ± cross-linker, active and inactive chromatin separated by ChIP (using an antibody against tri-methyl K4 in H3), and proximity between *DsRed* and *EGFP* genes assessed by 3C. Images illustrate gels containing amplimers, while row 7 gives the relative cross-linking frequency. Amounts of 3C amplimers in bands were determined by reference to equivalent weights of DNA, and normalized relative to *ALDOA*:*ALDOA* levels (*ALDOA* is an active endogenous housekeeping gene) and amplification efficiency.

DsRed:*EGFP* products are obtained from both supernatant and pellet (rows 1,3), consistent with naked input and/or active minichromosomes in the supernatant; they are seen only after cross-linking (X-link; rows 1,4), and more are seen in the pellet when promoters are identical (row 3, compare left and right lanes). *ALDOA*:*ALDOA* 3C products are seen only in the pellet (rows 2,5) and in similar amounts in both combinations (row 5). Amplimers are obtained in roughly equal amounts using the same primers and synthetic 3C templates made by ligating equimolar amounts of pure DNA from relevant genes (row 6). Minichromosomes bearing identical promoters (i.e., II + II) yield 6-fold more 3C product than those encoding different ones (i.e., II + III; row 7), indicating they lie closer together.

7.4 Similar host- and minichromosomal units lie together

Next, I wished to know if active minichromosomes organise their own factories, or share the same factories with the host genes. Here I transfect plasmids with either the *U2* or CMV units, and assess the proximity between the common plasmid backbone and host *U2* units using ChIP-3C. As both kinds of plasmids replicate to give ~8000 copies at 24 h (Ch.4.2), inactive background may confuse the result. Therefore, I choose an earlier time point (8 h), when there are fewer copies (Fig.4.2).

As Fig. 7.5 shows, 9-fold more plasmid:host 3C product was obtained with the *U2* unit than the CMV unit. This confirms the FISH results, and suggests minichromosomal and host genes bearing similar *U2* units are transcribed in the same factory.

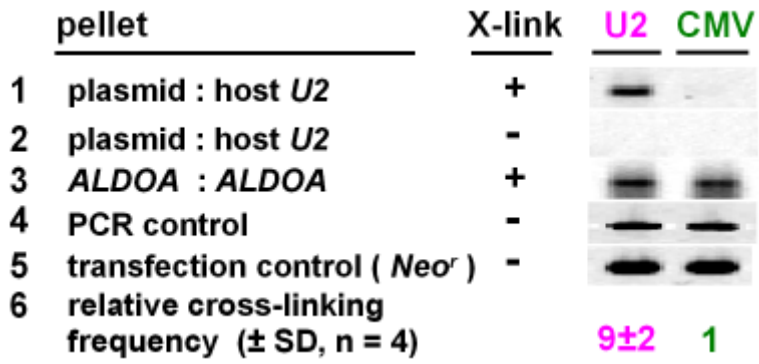


Fig. 7.5. ChIP-3C: similar genes in host and minichromosomes lie together

Cells were co-transfected with II^{U2},*U2G*,3' box (U2) or II^{CMV},*EGFP*,pA (CMV), grown for 8 h (rows 1-6), treated ± cross-linker, active and inactive chromatin separated by ChIP (using an antibody against tri-methyl K4 in H3), and proximity between plasmid backbone and sequences flanking host *U2* genes assessed by 3C. Images illustrate gels containing amplimers, while row 6 gives the relative cross-linking frequency. Amounts of 3C amplimers in bands were determined by reference to equivalent weights of DNA, and normalized relative to *ALDOA*:*ALDOA* levels and amplification efficiency.

More plasmid:host 3C products are seen with the U2 plasmid (row 1, compare left lane with the right). In contrast, amplimers from *ALDOA*:*ALDOA*, from synthetic 3C templates, and from the Neo^r gene on the backbone (a control for transfection efficiency) are all obtained in roughly equal amounts (rows 3-5). Minichromosomes bearing *U2G* units lie closer to host *U2* units than those encoding CMV units, and so yield 9-fold more (plasmid:*U2*) 3C product (row 6).

7.5 Discussion

In this chapter, I first established 3C methods to determine proximity between minichromosomes, as well as between mini- and host-chromosomes (Fig. 7.2). In addition to conventional 3C, a ChIP using an antibody against $^{\text{tri-methyl}}$H3K4 was used to eliminate the transcriptional-inactive background.

ChIP-3C showed minichromosomes bearing identical promoters (i.e., II + II) yield 6-fold more 3C product than those encoding different ones (i.e., II + III; Fig. 7.4); this confirms that minichromosomes with the same promoters lies closer together than those with different promoters. This verified the RNA FISH result (Ch.5.3).

This result with minichromosomes is supported by several studies. For example, artificial tandem arrays of human *U1* and *U2* snRNA genes colocalise in the same factories on Cajal bodies, and the association is abolished by promoter mutations (Frey et al., 1999). Moreover, α- and β-globin transcription units are often found in close proximity, and the interaction is transcription-dependent (Brown et al., 2006).

Moreover, ChIP-3C demonstrated mini- and host-chromosomes bearing similar transcription units (i.e., *RNU2*) lie close to each other. Minichromosomes with a *U2* unit generate 9-fold more plasmid:host 3C product than those with the CMV unit (Fig. 7.5). This is consistent with active minichromosomes sharing the same factories with similar host genes, and further validates the use of minichromosomes as probes to detect specialised transcription factories.

So far, I used two independent methods (RNA FISH and 3C) to demonstrate that pol I, II, and III units are transcribed in their own dedicated factories, and that pol II factories further specialise in transcribing different subsets of genes.

8 Direct visualisation of a transcription wave

8.1 Introduction

The mRNA expression profile of the human genome (i.e., transcriptome) in various tissues has been extensively studied using microarrays (Caron et al., 2001; Purmann et al., 2007). However, little is known about how nascent transcripts are made in real time. As reviewed in Ch.1.1.1, nascent RNA is co-transcriptionally processed, and most introns are rapidly degraded; as a result, nascent RNA is hard to detect. However, an inducible system and the presence of genes with exceptionally long introns provide a window of opportunity to study transcript production in detail. This study represents the collaboration with T Kodama's group in Tokyo (Laboratory for Systems Biology and Medicine, Research Centre for Advanced Science and Technology, The University of Tokyo). They performed all the microarray studies that will be described; they also provided me with fixed cells that I then analysed using RNA FISH.

In human vascular endothelial cells (HUVECs), TNFα induces a cascade of gene expression, mediating the inflammatory response such as the migration, adhesion and transmigration of leukocytes (Aggarwal, 2003; Minami et al., 2004; Janes et al., 2006). Using traditional oligonucleotide microarrays (i.e., an Affymetrix U133 plus 2.0 (U133)), the levels of ~500 transcripts were shown by my collaborators to change significantly when TNFα is added to HUVECs. Among these, *SAMD4* (222,209 bp, Chr13) and *EXT1* (312,453 bp, Chr14) are very long genes containing multiple long (~150 kb) introns.

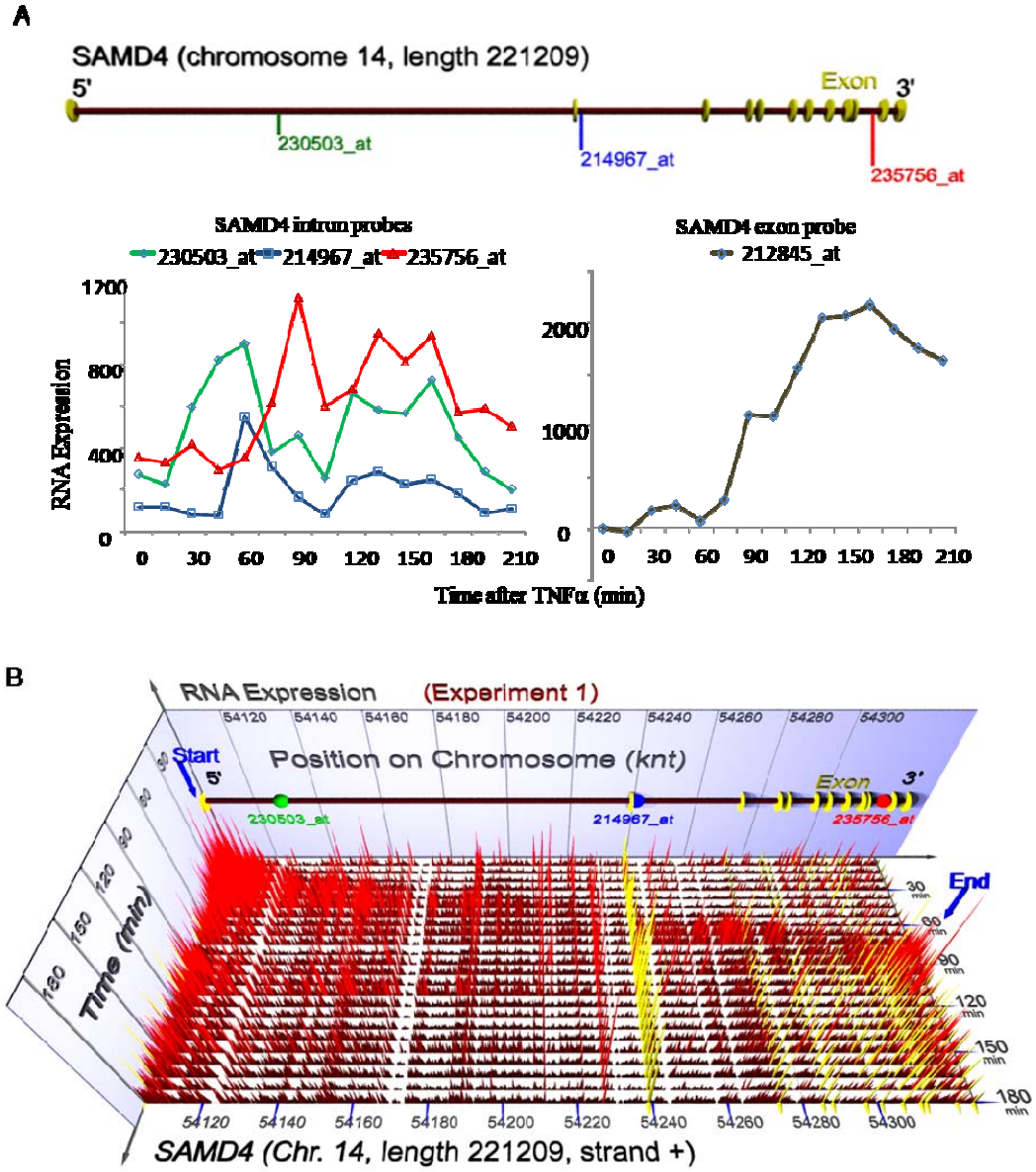


Fig. 8.1. A transcription wave visualised using microarrays

A. Affymetrix U133 plus 2.0 array (samples collected every 15 min after stimulation). Levels of intron 1 (green), 2 (blue), and 7 (red) RNA, plus those of mature mRNA (black), are indicated.

B. Tiling arrays (samples collected every 7.5 min after stimulation). Transcripts detected by intronic and exonic probes are indicated by red and yellow needles, probe positions are shown from left to right, and time after stimulation from top to bottom. Blue arrows: start and end of first wave.

(Adapted from T. Kodama *et al.*, manuscript submitted).

The level of *SAMD4* mRNA increases 2.5 times between 90-150 min after stimulation, indicating a stable accumulation of mature mRNA (Fig. 8.1A). I mainly analyse nascent transcripts during the first 90 min – before the time when mRNA levels begin to rise.

T Kodama and his group used a 25-mer oligonucleotide tiling array covering the whole of *SAMD4*; probes were spaced every 13 nt (except for repeated regions, and those containing high or low GC content). Cells were collected every 7.5 min after stimulation with TNFα, RNA isolated, and applied to the microarray. At t = 0 min, essentially no signal was given by any *SAMD4* probes on the array (targeting either the sense or anti-sense strand). However, soon after stimulation, signal was seen with probes at the immediate 5' end of the gene; a wave of signal then moved progressively down the gene over time (Fig. 8.1B). Similar profiles were seen in other long genes stimulated by TNFα (i.e., *ZFPM2*, *EXT1*, *ALCAM*, and *NFKB1*) (not shown). Signals given by only a few of the many thousands of probes on the array are given in Fig. 8.1A. One probe set targeting an intron one-third of the way down *SAMD4* yielded a wave of signal that appeared between 30-60 min; between 120-180 min, a second wave appeared (Fig. 8.1A, probe 230503_at, green curve). The first wave reaches an intron two-thirds of the way into the gene later (probe 214967_at, blue curve), and a near-terminal intron later still (probe 235756_at, red curve). Movement of the first wave is consistent with passage of a polymerase down the gene coupled to rapid degradation of intronic RNA once the polymerase passes. In contrast, the timing of second wave does not change, and must result from polymerases that initiate at additional promoters and/or complex effects on transcript stability. In this chapter, I used intron-RNA FISH to validate the tiling array results.

8.2 Establishing the intron-RNA FISH method

8.2.1 Designing probes

Five 50-mer probes were synthesised (Gene Design, Tokyo, Japan) for each target (*SAMD4*: intron1a, 1b, 7 (Fig. 8.3A); *EDN1;*) as described (Femino et al., 2003) (see Ch.2 chemical labelling); each set of 5 50-mers targeted ~600 nt. The targets were selected because they provided strong signals in the tiling arrays. Probe sequences, location, and targets are listed in Table 8.1 and 8.2. Every 10 nt, a T was substituted by an amino-modifier C6-dT, and labelled with Cy3 or Alexa 488. The labelling efficiency of *SAMD4* and *EDN1* probes is 23 and 20 per 250 nt respectively.

Table 8.1. FISH probes for *SAMD4*

Intron	Sequence	Start	End	Strand
1a	C(T)AATTACTCTAA(T)CCTGCCATTCC(T)AA GCCAGGTGA(T)GGCCCTTTGA(T)T	54138070	54138119	+
1a	AGAGTG(T)CTGAGGGGT(T)AGAGTATTC(T) CTCCAAGTCA(T)CAGCCTTGGT(T)T	54138250	54138299	+
1a	(T)GGTTGCCAGA(T)GCCTTCGGAG(T)TTTC TTCTT(T)CTCTAAGGAC(T)TAGCAC	54138392	54138441	+
1a	(T)GTCCTGCTT(T)CTCCTTTGTT(T)CTACATT GCT(T)CCCTCAATTT(T)CTCTCC	54138446	54138495	+
1a	GCA(T)TGATTCCACAG(T)TATCTAATCCA(T) GGCCAGCCAG(T)CCCACATTCC(T)	54138589	54138638	+
1b	GTCC(T)GAGATCCAA(T)GCCACTCCCTA(T) TTCATCCCCT(T)CAGGGGTACC(T)G	54219630	54219679	+
1b	C(T)CTATCCTGCC(T)CACAAAGGC(T)CGCA CCTGTA(T)TTTCAGCCA(T)CCCTAT	54219687	54219736	+
1b	GCCA(T)CCACTTGGGTC(T)TCTGGGAGAC(T)CTCACTAACC(T)GGCATGGTCC(T)	54219747	54219796	+
1b	GTCA(T)GATCTTAGAAA(T)CCTATGTGAC(T) GCCTTGGGC(T)CCAGACATCAC(T)	54219797	54219846	+
1b	(T)CCTGATCTCA(T)GACCAGCATT(T)TTCCT CAATC(T)TTCTTGGAAC(T)ATCCC	54219852	54219901	+
1c	ATAG(T)AGCGAGACT(T)TATCAACAG(T)GA GTTAGTG(T)TACAGGTGAC(T)TGGC	54147087	54147136	+
1c	GGAA(T)TTCTGCCCAC(T)CGGAAGGCCAC(T)GGTGTCCTGAG(T)GTTGCTCAT(T)CAGC	54147804	54147857	+
1c	AGGC(T)ACGATTTTTA(T)GTAAAGGCTGA(T)TAAGTGGTT(T)GCTGGGATC(T)GGAG	54147866	54147917	+
1c	CGCC(T)GGCCCCAAC(T)GGCTTATTT(T)CT GTCAAAC(T)CAGTGTTTGA(T)AGAA	54148022	54148071	+
1c	GGTT(T)CACCATGTTA(T)CCAGGCTGG(T)C TCGAACTCT(T)GACCTCAAG(T)CATC	54148121	54148171	+
1d	CTTC(T)GAGACTCATGC(T)ATTGCCCAGGC(T)GAAGTGCAA(T)GGCACAATC(T)CATC	54240283	54240335	+
1d	GCAA(T)ATCAATGCAG(T)ATCTTCAGTG(T)T TGCTTGGTTA(T)TCCACTGAGG(T)ACTG	54240472	54240525	+
1d	CTGT(T)CCTGCTGGA(T)TATAGTCTG(T)GAA AGACAGTG(T)CATCAGAACC(T)CTTA	54240631	54240682	+
1d	AGCC(T)CCAAGTCCTTT(T)CACTCAAGTAT(	54241064	54241118	+

	T)ACATAGGGGT(T)CAGAGGTCAG(T)AAGG			
1d	TGAA(T)GAAACTTTC(T)GGAGGGGAGA(T)C	54241181	54241231	+
	TGAGCTCTT(T)GATCTTCTC(T)GATG			
7	ATTC(T)TCACTGTGT(T)CTTGCTTTC(T)CTTC	54302708	54302757	+
	CTTTG(T)CGCTTTTTCC(T)GGCC			
7	TGCC(T)AATTCTCTCAC(T)ACTGTAACC(T)G	54302779	54302828	+
	CCTAGGTGC(T)CAGTAAGTATC(T)			
7	GCTG(T)GATGGCCTG(T)TTCTTTCTG(T)CTC	54302831	54302880	+
	ACCCCT(T)CTCTCACTAC(T)GTAA			
7	CTGC(T)CCTCTGTGC(T)CCAGGGCAC(T)CA	54302890	54302939	+
	GGGATGACC(T)TATCAGAGGCT(T)T			
7	C(T)CAGCTGTCT(T)GAGGCAAGG(T)GGACA	54303206	54303255	+
	GTACC(T)TGTCCTCCAA(T)ATCAAT			
1a	GGAA(T)CAATGCCCTTT(T)ACTTGGGACA(T	54138628	54138677	-
	)CTTTCCAAG(T)CTGGATTAT(T)TC			
1a	CAGGGA(T)GAAGATACT(T)CCTTTCCTG(T)	54138874	54138923	-
	CTGTTCTAG(T)CTCTTCTGG(T)TGC			
1a	CTTG(T)GGGAATGACA(T)GAGTAGCCCCA(	54138990	54139039	-
	T)GGACAGAAACC(T)TCTCAGCCC(T)			
1a	CTTACA(T)TCTCTCTCAG(T)TCATGCCTG(T)	54139289	54139338	-
	CAGGCTAAA(T)GTCTGGATCC(T)T			
1a	TACT(T)CAAGTGCACAC(T)GGGGACCAG(T)	54139341	54139390	-
	ATCAAGGACA(T)CACCTGAGA(T)CT			
1b	TTTA(T)GACTGCCGCAA(T)ATCCCACTT(T)A	54218373	54218422	-
	GGAGGAAA(T)GCAGAGCCAT(T)GA			
1b	AATAAT(T)ATGCTGTCA(T)AACATGCAG(T)	54218436	54218485	-
	ACCTGGGTC(T)AGGACTGGCC(T)TA			
1b	(T)GAACATGTGCC(T)GAGGTAGGTAT(T)TC	54218509	54218558	-
	ATTAGTGTG(T)GACTCAGGAA(T)CT			
1b	TTCTTA(T)CCCACAATG(T)TTGGGGTTG(T)G	54218570	54218619	-
	GCCTTTAA(T)GTTTGGGAT(T)GTG			
1b	ACTTGC(T)CAGTTCCATG(T)CGGCCTCCTG(	54218723	54218772	-
	T)CTCTTCTGT(T)TGATTCCAT(T)T			
7	TTATTA(T)CCTCAGTTTC(T)AGGAGGCAC(T)	54304865	54304914	-
	GAGTCTTGGG(T)AGAGGGGCA(T)G			
7	CGTGTG(T)AGCTGGACA(T)GCAGGGAGCA(	54305126	54305175	-
	T)CACCAACCAA(T)GAGAGGGAC(T)G			
7	CGCC(T)CCTAGCCCACC(T)TGATGCATA(T)	54305195	54305244	-
	TCTTAATGGCA(T)CTGGCTAAGG(T)			

7	G(T)TTGTAAGTCC(T)CAATGACGTT(T)CTGC CACATGA(T)TAACATCCT(T)CTGC	54305261	54305310	-
7	TTCTCC(T)GGTATTTAGG(T)CCTTAAAGA(T) TCAGCATCAC(T)GGGAGACAAC(T)	54305325	54305374	-

SAMD4 is located on Chr14. (T) is amine-C6-dT. The location of intron 1a, 1b and 7 is illustrated in Fig. 8.3A. 1c and 1d are probes for future experiments (see Ch.8.6)

Table 8.2. FISH probes for *EDN1*

Intron	Sequence	Start	End	Strand
3	C(T)CA(T)AGCCAGAGGGC(T)CTCCAATAAT (T)TCCAATGTG(T)CCATTTTCA(T)TCTC	12401546	12401597	+
3	CTCC(T)TGTGGTTTT(T)GGTGTGGTT(T)GAT ATTGTT(T)GGATTTTGG(T)CCTC	12401445	12401493	+
3	AATG(T)GTGCTTGTT(T)ATGACTGCTCCGA CAGATGA(T)GAACTAGTG(T)CCAG	12401362	12401412	+
3	CTGC(T)ACAAACTCAC(T)CCTGCACAAA(T) GGCTTCAACAC(T)TTGAGCCTAGG(T)TTTT	12401169	12401223	+
3	ATTC(T)AACCCTCTA(T)ATCATCACTTC(T)G CCTCTCAGTC(T)CCACCCTCCCA(T)GAGA	12401095	12401149	+

EDN1 is located on Chr6. (T) is amine-C6-dT.

8.2.2 Can probes detect a single molecule?

Are probes sensitive enough to detect a single transcript? If a single probe can be visualised, then there is a good chance a mixture of 5 probes could detect a single RNA molecule.

I absorbed labelled probes to glass slides as illustrated in Fig.8.2. Using a Zeiss Axioplan 2 microscope equipped with a Photmetrics CoolSNAP$_{HQ}$ camera, and an exposure of 8,000 ms in the FITC channel and 12,000 ms in the Cy3 channel (Alexa 488 is generally brighter than Cy3), I was able to visualise many discrete foci (Fig.8.2A). Dilution of probe correlates with the number of foci seen, and the mean intensity of foci remains the same, indicating the foci detected are single molecules rather than aggregates. I normalised the intensity to that given by 2.5 μm Orange and Green Intensity Calibration beads (0.02 % intensity), and calculated labelling efficiencies from the absorption of bases and bound fluors using published extinction coefficients. For each target region, the probes have between 19-23 fluors. A single *SAMD4* 50-mer had 17 ± 2 % the fluorescence intensity of a bead.

Furthermore, by adding 1 or 2 linking oligonucleotides complementary to the ends of different monomers, followed by hybridisation, one would expect the distribution of the intensity of foci to shift towards higher intensities (as the linker brings together different probe molecules). Theoretically, the ratio of mean intensity without linker: +1 linker: +2 linkers = 1: 1.25: 1.66. Analysis of the result showed a consistent shift of intensity distribution (Fig. 8.2B), and the ratio of mean intensity was 1: 1.28: 1.61, which is similar to the expected value.

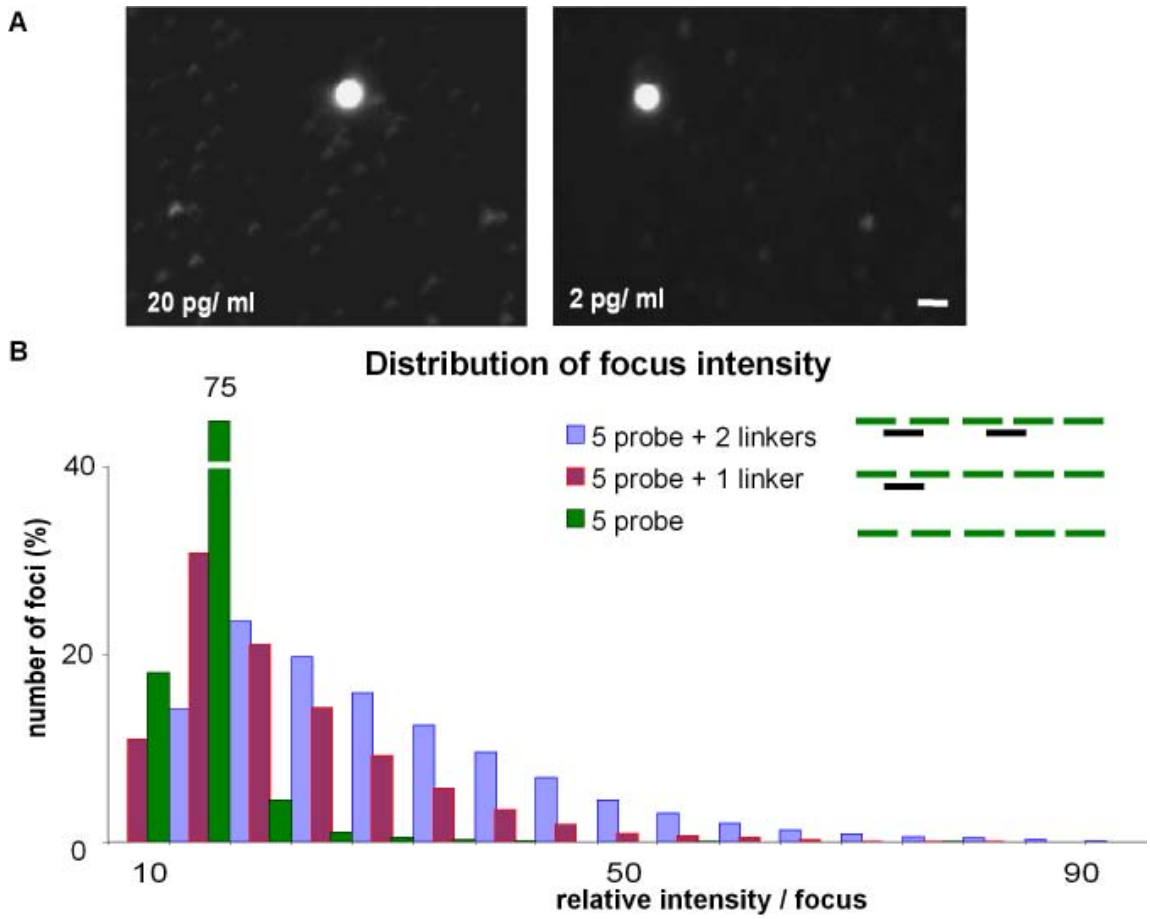


Fig. 8.2. Are probes sensitive enough to detect single molecules?

5 labelled probes (targeted against region 1a and tagged with Alexa 488) were mixed with reference beads, probes of different concentrations absorbed to glass slides, images collected, and intensities analysed.

A. Two concentrations (20 pg/ml and 2 pg/ml) of probes mixed with reference beads (bright circles). When diluted 10x, correspondingly fewer probes are seen, while the mean intensity of each focus remains roughly the same; this indicates there are few aggregates, and that each focus represents a single probe. Bar: 2.5 μm.

B. Distribution of mean intensity of foci shifts when linkers are added. 1 or 2 linkers, linking either 1^{st} + 2^{nd}, and/or 3^{rd} + 4^{th} probes were added to probes before they were absorbed to slides. The intensity of each focus was measured and normalised relative to that of the reference beads, and the fraction of foci in bins of different intensities determined. When 1 or 2 linkers are added, the median intensity increases.

Therefore, I conclude the imaging system is sensitive enough to detect a single probe *in vitro*; therefore, it seems possible that I could use 5 probes to try and detect a single transcript in a cell.

8.3 Visualising the transcription wave using intron-RNA FISH

Using sets of five 50-mers (and their anti-sense equivalents as controls), I carried out RNA FISH to detect nascent *SAMD4* RNA at 0, 30, 52.5, and 75 min after stimulation. Intensities of foci are normalised relative to reference beads.

Immediately after TNFα stimulation (0 min), almost no signal was detected in the middle of the first intron (using probe 1a). By 30 min, some cells contain one or two nuclear foci; there were never more than two foci (Fig. 8.3B). Presumably, such foci mark nascent RNA at the site of transcription on one or both of the two alleles (Femino et al., 1998). At 52.5 and 75 min, the total number of foci in the population declines with time but focus intensity and size remain constant (Table 8.3). A typical focus has an intensity ~50% that of the reference beads, which is equivalent to that of three 30-mers (Ch.8.2.2). I also found two red (or two green) foci in one cell more often than expected if each allele fired randomly. This suggests when one allele fires, the other is more likely to fire in the same cell.

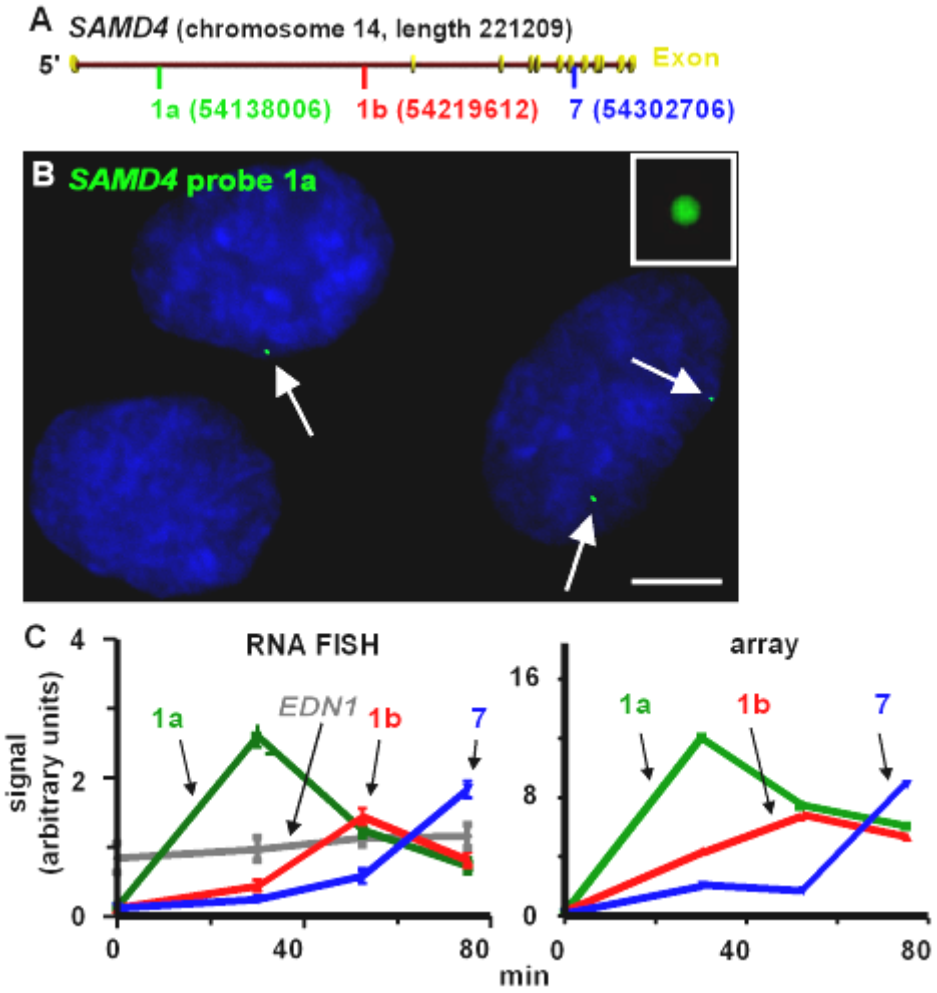


Fig. 8.3. Transcription wave visualised by RNA FISH

HUVECs were fixed 0, 30, 52.5 and 75 min after stimulation with TNFα, nascent *SAMD4* or *EDN1* RNA detected by FISH using intron probes 1a, 1b, or 7, or an intron probe against *EDN1* (labelled with Alexa 488), cells counterstained with DAPI, and images collected.

A. *SAMD4* locus showing probe positions.

B. A typical field 30 min after stimulation obtained using probe 1a. Cells have 0, 1, or 2 green foci/cell (arrows) marking nascent RNA at one or other allele. Intensities are normalised relative to fluorescent beads (inset) to permit comparison between different experiments. Bar: 5 μm.

C. RNA FISH and arrays give similar results. Left: signals (i.e., size in pixels x intensity x number of foci; in arbitrary units) were obtained by RNA FISH. Signals in *SAMD4* intron 1a (green), 1b (red) and 7 (blue) peak as a wave of transcription passes through each region with time, whereas *EDN1* signal (grey) remains constant. Right: similar variations are given by relevant probes in arrays. Array data provided by T. Kodama.

Table 8.3. Size, intensity, and distribution of RNA FISH foci

		Focus					
		size	relative intensity	number per	Distribution		
Intron	min	(pixels ± SD)	(% ± SD)	nucleus	0	1	2
SAMD4							
1a	0	7 ± 0	48 ± 21	0.04	96.5	3	0.5
	30	9 ± 1	42 ± 15	0.73	50	27.5	22.5
	52.5	9 ± 2	51 ± 18	0.27	79.5	14	6.5
	75	10 ± 3	51 ± 24	0.15	87.5	10	2.5
1b	0	8 ± 0	61 ± 96*	0.03	97.5	2.5*	0
	30	10 ± 4	42 ± 18	0.11	89.5	10	0.5
	52.5	9 ± 2	40 ± 14	0.42	66	28	6
	75	9 ± 3	39 ± 18	0.24	77.5	21.5	1
7	0	9 ± 4	65 ± 39*	0.02	98	2*	0
	30	9 ± 4	35 ± 20	0.08	93.5	5.5	1
	52.5	9 ± 4	47 ± 22	0.13	87.5	12	0.5
	75	8 ± 2	35 ± 11	0.63	57.5	27	15.5
EDN1							
2	0	10 ± 4	22 ± 13	0.38	69	22	9
	30	10 ± 4	31 ± 13	0.32	74	20	6
	52.5	10 ± 3	35 ± 13	0.31	77	15	8
	75	9 ± 2	39 ± 20	0.33	72	23	5

Values obtained from images like that shown in Fig. 8.3B. Intensities were normalised relative to fluorescent reference beads (Fig. 8.3B, inset). For each time and region, 400 cells were analysed in two independent experiments for *SAMD4*, and 200 cells from one experiment for *EDN1*. Pixels were 103 x 103 nm.

The fraction of cells with two *SAMD4* foci at 30 min is higher than expected if we assume alleles fire independently; this is consistent with the presence of "responding" cells in the population in which both alleles are likely to fire. *EDN1* relative intensity is lower, as the labelling efficiency of *EDN1* is 20 fluors / 250 nt compared with 23 fluors / 250 nt with *SAMD4*.

*: Few foci were detected and most were probably 'rogue' (background) foci as intensities were much higher than the average.

The integrated intensity per nucleus (equivalent to the total signal in the population measured by microarray) rose from 0 to a peak at 30 min before declining (Fig.8.3C left, green line). Another probe set targeting the end of intron 1 (intron 1b) demonstrated an analogous wave that peaked at 52.5 min (Fig. 8.3B, left, red line); one targeting intron 7 peaked later still (Fig. 8.3B, left, blue line). This pattern of waves is similar to that seen using tiling arrays (Fig. 8.3C right). Anti-sense probes targeting intron 1a, 1b and 7 gives no signal (not shown).

The highly-expressed gene, *EDN1* (endothelin-1), was used as an internal control. Its transcript levels remain stable regardless of time after TNFα stimulation (not shown). No transcription wave was observed by RNA FISH; rather, the expression pattern remained stable (Fig. 8.3C left, grey line).

The nascent transcript profile of *SAMD4* seen by RNA FISH is very similar to that by tiling array (Fig. 8.3C). Thus I conclude intron RNA FISH validated the microarray results.

8.4 Double labelling reveals co-transcriptional degradation

120 min after TNFα stimulation, the tiling array detects a smaller second wave; the peaks given by the two probes (*SAMD4* 1a and 7) are much closer together (Fig. 8.4B right). Several possibilities may explain the appearance of the 2nd wave. 1) Temporal asynchronies: transcription is not as synchronised as in the first wave, perhaps because different cells behave differently. 2) Spatial asynchronies: instead of initiating from one 5' promoter, polymerases may initiate from multiple starting sites in the middle of the gene. In both 1) and 2), the broad wave indicates a collective

result for a population of cells. 3) On each template, several polymerases transcribe simultaneously from different starting sites.

I carried out a double labelling targeting *SAMD4* introns 1a and 7 at 0, 30, 60, 90, 120, 150, 180 min – times that covered both the 1st and 2nd waves.

From 0-112.5 min, double labelling yielded similar results to the single labelling (Ch.8.3). Moreover, intron 1a and 7 never colocalise during the 1st wave, indicating introns are co-transcriptionally degraded (presumably after splicing). In other words, intron 1a is never seen in the presence of intron 7, so that the first intron must be removed by the time that the polymerase is transcribing intron 7.

During the 2nd wave (i.e., between 112.5-180 min) microarray data indicates that regions 1a and 7 are both present in the population; however, FISH shows that they are never together at the same allele (Table 8.4). Thus, some cells contain one or two foci of the same colour (either green or red), and these account for 30-40% of the population. Only a few contain both colours (green and red), and these account for ~5%. But none of the foci contain both colours (and so appear to be yellow). Thus intron 1 must be co-transcriptionally removed before intron 7 is made.

These results also suggest that only one polymerase can be engaged on an allele at any time (which would exclude possibility 3) described above). We calculated the probability of seeing a yellow focus by multiplying the probability of seeing a red one with that of seeing a green one. This value is low but still high enough to be observed when 400 cells are sampled, and so is significantly higher than the experimentally-observed result (Student's *t*-test, $P<0.05$). In other words, transcribing one intron inhibits the transcription of another intron.

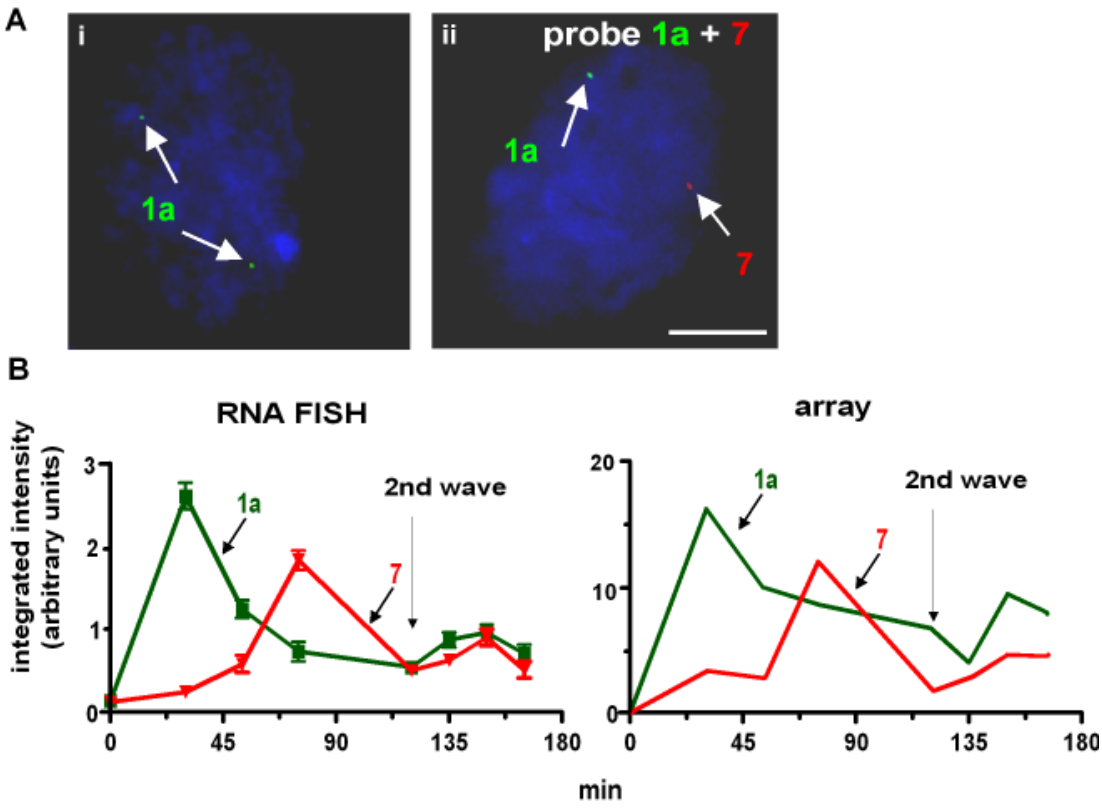


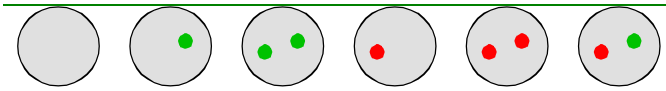
Fig. 8.4. 2nd wave detected by double-labelling RNA FISH and array

HUVECs were fixed after 0, 30, 60, 90, 120, 150, 180 min after TNFα treatment, nascent *SAMD4* RNA detected by RNA FISH using introns probes 1a and 7 (Fig. 8.3A) labelled with Alexa 488 and Cy3 respectively, cells counterstained with DAPI, and images collected. Intensities are normalised relative to fluorescent beads.

A. Two typical images of double labelling at 120 (i) and 150 (ii) min. One cell contains 2 green foci (arrows), the other contains one red and one green focus. Yellow foci are never seen. Bar: 5 μm.

B. A second wave of transcription after TNFα stimulation detected by both RNA FISH and the tiling array. Intron 1a and 7 peak at two distinctive time points during the 1st wave (which lasts between 0-112.5 min). However, during the 2nd wave (which begins at the point marked by the arrow), the peaks are closer together. Array data provided by T. Kodama.

Table 8.4. Fraction of cells after double labelling

time (min)	fraction of cells with pattern indicated (%)						
							yellow focus (any combination) [expected]
30	63	29	6	1	0	1	0 [0.2]
120	70	13	1	14	0	2	0 [0.7]
135	57	18	4	18	1	2	0 [1.5]
150	55	14	6	18	3	4	0 [2.1]
165	70	12	4	10	2	2	0 [0.9]

Fraction of cells with the pattern indicated using *SAMD4* probes 1a (green; Alexa488) and 7 (red; Cy3), as in Fig.8.4, is shown. The expected percentage of yellow foci is the probability of a focus being red multiplied by the probability of a focus being green. Average intensity and size of individual foci are similar at different times (as in Table 8.3; not shown). 200 cells were analysed at each time.

At all times, most cells possess no foci, and the next largest group possesses foci of one or other colour; few cells possess both red and green foci, and no yellow foci were seen. This demonstrates co-transcriptional degradation, as intron 1a must have been removed before intron 7 was transcribed. Moreover, assuming transcriptional initiation occurs randomly, one should see the 'expected' numbers due to two independent polymerases present at one locus. The absence of colocalisation is consistent with transcription of one intron inhibiting transcription of the other, and/or productive transcription occurring very rarely.

8.5 Various data suggest there is one engaged polymerase/allele

Although the fraction of cells containing 1 or 2 foci varied with time, all foci had the same intensity and size (Table. 8.3). The mean intensity of foci given by *SAMD4* and *EDN1* are also similar (Fig. 8.5; if corrected for the labelling efficiency of the probes); this suggests the two genes are generally transcribed by the same number of polymerases.

There are two possible explanations of the above data: a simpler one (i.e., only one polymerase is engaged on an allele at any time), and a more complicated one (i.e., similar numbers of polymerases are packed into a convoy that transcribe both *SAMD4* and *EDN1* at all times and across all regions).

To test the two hypotheses, I analysed the distribution of foci intensities at each region at each time point. If foci contained variable numbers of transcripts, one would expect them to exhibit a wide range of intensities containing more than one peak. However, all profiles appear similar, with intensities clustered around a mean (i.e., *SAMD4* – 48%, and *EDN1* – 36%). Only a single peak is seen in each case (Fig. 8.5). Moreover, comparison of intensities given by a single focus and 50-mer probe indicates that a typical focus contains only three of the five 50-mers in the probe set (mean 50-mer intensity *SAMD4* – 17%, and *EDN1* – 14%). This is consistent with fewer than 5 50-mers hybridising to one transcript.

These different aspects of the data are all consistent with the hypothesis that on one template, only a single polymerase is engaged at any time.

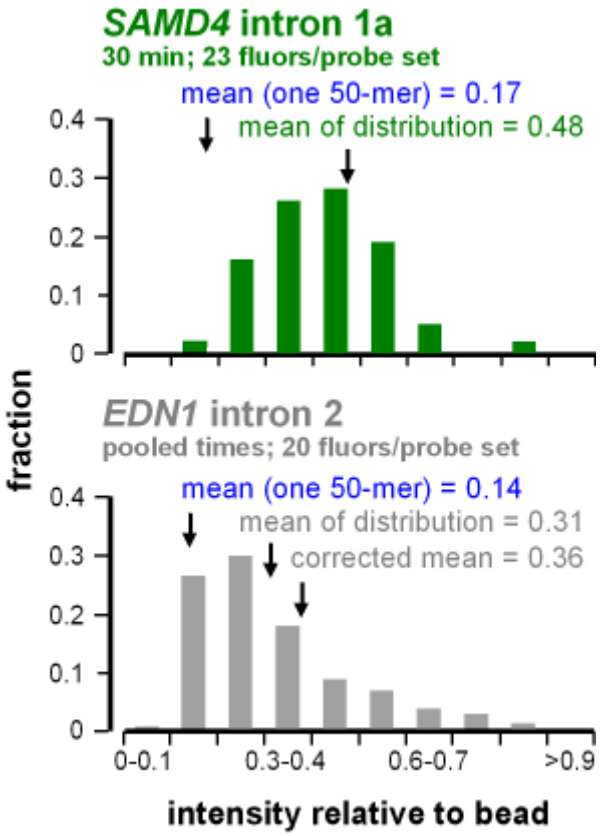


Fig. 8.5. Distributions of intensities of *SAMD4* and *EDN1* foci

The intensity of foci in images like that in Fig. 8.3A were normalised relative to fluorescent beads, and relative intensities binned (0-0.1, 0.1-0.2, etc). Arrows indicate relative intensities of single 50-mers (which had an average of 4.5 and 3.5 fluors per *SAMD4* and *EDN1* 50-mer¸ respectively), mean intensities of distributions, and the mean of the *EDN1* distribution after correction for the difference in intensities of the 50-mers. Distributions are similar.

8.6 Discussion

In this chapter, I used intron-RNA FISH to confirm results obtained by my collaborators using tiling arrays. The two methods yield essentially very similar patterns of *SAMD4* transcripts after TNFα stimulation (Fig. 8.1, 8.3 and 8.4).

Both microarray and RNA FISH data are consistent with a transcription elongation speed of around 3.5 kb/min on *SAMD4* (Fig. 8.1 and 8.3). This value is faster than previous studies, for example 1.1-1.4 kb for *Drosophila* UBX (Shermoen and O'Farrell, 1991), and 1.7-2.5 kb for human dystrophin (Tennyson et al., 1995). Very recently, another sensitive method of detecting polymerase activity was established based on photobleaching and photoactivation of fluorescent MS2 proteins labelling nascent RNA (Darzacq et al., 2007). This study reported an elongation speed of 4.3 kb/min.

In addition to validation of the microarray method, intron-RNA FISH revealed much single-cell information about the transcription of long genes.

First, introns are co-transcriptionally degraded (and spliced), as intron 1a levels fall before the transcription wave reaches intron 7 (Fig. 8.3). This is also supported by double labelling (Fig. 8.4), as intron 1a and 7 RNA are never seen together at the same transcription site (as no yellow foci are seen).

Second, productive transcription is a rare event. Although *SAMD4* is induced by TNFα, most cells possess only one or no transcription foci (Table 8.3). Even at the peak time (intron 1a, 30 min), cells containing 2 foci represent only ~23% of the population. Furthermore, the peak value of transcription seems to decline with time, as Fig. 8.3 and Table 8.3 illustrate (compare integrated intensity of intron 1a at 30min with that of intron 7 at 75 min), as mRNA continues to accumulate (Fig.

8.1A). Tiling array results suggest a continuous declination of nascent transcript level from the 3' to the 5' of the gene. This pattern is also observed in other genes like *ZFPM2* (another TNFα-responsive gene on Chr15, length 485,619) (Kodama *et al.*, manuscript submitted). Probes covering the first ~8,000 nt from the initiation site continuously yield signal (between 15–180 min), and polymerases only seem to escape downstream between 15-30 min to initiate the first wave (Fig. 8.1B). This is not due to the conventional abortive transcription, which occurs before the formation of stable elongation complexes just after initiation (reviewed in Ch.1.1.1.2). The abortive transcripts of this type are <15 nt (Dvir, 2002), and so are too short to be detected in this study. Our results point to a major checkpoint that regulates escape.

Third, I find the transcription event is likely to be performed by a single polymerase. All transcription foci detected for both *SAMD4* and *EDN1* have similar intensity and size (Table. 8.3), and the distribution of foci intensities are similarly narrow (Fig. 8.5).

Fourth, two alleles possibly interact (directly or indirectly) in a cell in response to stimulation. The fraction of cells with two *SAMD4* foci at 30 min is higher than expected if alleles fire independently; this is consistent with the presence of "responding" cells in the population in which both alleles are likely to fire.

A number of studies suggested transcription is a rare event, since most genes are associated with less than 1 polymerase at a time (Jackson et al., 2000). A study using GFP-tagged pol II revealed most active human transcription units are also not being transcribed most of the time (Kimura et al., 2002). Many expressed genes in bacteria and yeast are transcribed only once per cell cycle (Bon et al., 2006).

Moreover, in *Drosophila*, genes encoding tubulin B1 and glyceraldehyde phosphate dehydrogenase associate with <1 pol II (Rougvie and Lis, 1990). In my study, by double labelling, intron-RNA FISH here showed intron 1a and intron 7 RNA never colocalise (Fig. 8.4), indicating that there is only one polymerase engaged at a time even on the long *SAMD4* gene (>200 kb). However, many other studies concluded transcription is a frequent event, especially on hyperactive units – for example, heat shock genes in flies (O'Brien and Lis, 1991), *globin* (Wijgerde et al., 1995) and *actin* (Femino et al., 1998) in mammals. In addition, the ENCODE project reveals most part of the human genome is transcribed (although may not be at the same time, since it is a collective result of many cells), sometimes from both strands (The ENCODE consortium, 2007).

In my opinion, both types of results could fit into the following model: once transcribed, chromatin becomes "open" and frequently contacts transcription factories, resulting in frequent initiation events. However, a checkpoint mechanism is tightly controlling polymerase escape into elongation, to produce a full-length transcript. Many nongenic transcripts are the results of abortive transcription across the whole genome, as detected by the ENCODE project. This model is consistent with several recent findings. A study from Young's group on human ES cells suggested most active genes are associated with pols, but very few produce detectable full length transcripts, other genes experience initiation but not elongation (Guenther et al., 2007). Singer's kinetics study on polymerases in living cells suggested only 1% of polymerase-gene interactions leading to completion of an mRNA (Darzacq et al., 2007). The checkpoint mechanism may be related to cotranscriptional processing. For example, a study in yeast demonstrated an early

elongation checkpoint that facilitates the link to pre-mRNA processing, coordinated by the yeast forkhead transcription factors Fkh1p and Fkh2p (Morillon, 2003).

In addition, many interesting questions still remain.

First, the second wave has a very different pattern from the first. Multiple internal promoters may be present in *SAMD4*. After being transcribed once, the chromatin structure may become open / active, so that polymerases can initiate at alternative sites.

Second, is a long intron being degraded as the polymerase is still transcribing the same intron? If not, the polymerase needs to remain associated with a ~150 knt piece of RNA until it reaches the next exon. This seems unlikely, since the signal detected by probe 1a falls as that detected by 1b rises (Fig.8.3). In *Drosophila*, a large intron can be spliced by a process known as recursive splicing. Here, a splice donor site might be joined to donor site in the same intron if it contains the appropriate signal. Such recursive splicing signals have the sequence YAG|GURAGU, where | marks the exon-intron junction. If a new acceptor site in the same intron is now generated by the polymerase, the old donor may be joined to the new acceptor to give the recursive splicing (Burnette et al., 2005). Examination of the first intron of *SAMD4* reveals two motifs that completely match the recursive splicing sequence described above (none were found in the anti-sense sequence). If less-strict criteria are used (i.e., YAG|GURAG), I could identify 10 such elements in the sense strand. A more comprehensive RNA FISH experiment such as the following is needed. I would need to prepare a series of intron probes ranging from 1 knt away from intron 1a to the end of intron 1 (two examples are intron 1c and 1d, Table 8.1), covering the whole intron. Then, I would perform a double labelling experiment involving a reference (e.g., intron 1a) and a test probe (the rest). Two

probes lying 1 knt apart would serve as a positive control (they would always appear to co-localise as they target the same short piece of nascent RNA). As the distance between the two probes increases, I would expect to see signal from the 3' probe without any signal from the 5' probe. The point at which this occurred could correlate with a recursive splicing element.

Third, are TNFα-induced genes transcribed in specialised factories? As reviewed in Ch.1.4.6, many genes responding to the same inducer are transcribed in dedicated factories. In order to address this question, a double labelling experiment targeting nascent transcripts from *SAMD4* and *EXT1* (or one of the many other genes that respond to TNFα like these two) is required. Thanks to the inducible system, whether colocalisation (if any) is transcription-dependent should be easy to resolve.

In conclusion, the use of tiling arrays and RNA-FISH with intron-specific probes provides a powerful way of analysing transcript production with high temporal and spatial resolution. While arrays provide large amounts of quantitative information on a population of cells, RNA FISH provides detailed single-cell information.

9 Discussion

9.1 Introduction

Different nuclear genes are transcribed by different classes of RNA polymerase (i.e., I, II, or III), and the resulting transcripts are processed differently (e.g., some are capped, others spliced). It is widely assumed that active polymerases track along their templates as they make RNA. However, after allowing engaged polymerases to extend their transcripts in tagged precursors (e.g., Br-U or Br-UTP), and immunolabelling the now-tagged nascent RNA, active transcription units are found to be clustered in nuclei, in small and numerous sites we call “transcription factories” (Cook, 1999; Cook, 2002). Previous research has shown the different classes of polymerases are concentrated in their own dedicated factories (Pombo et al., 1999). This thesis aims to analyse factories involved in transcription of specified groups of genes in eukaryotes.

9.2 Using replicating minichromosomes as probes to detect factories

12 plasmids encoding different promoters, transcription units, +/- introns and 3' ends were constructed; all plasmids contained the SV40 origin of replication (Fig. 3.4). After transfection into COS-7 cells expressing the SV40 T antigen, plasmid DNA is assembled into nucleosomes, and the resulting minichromosomes are replicated and transcribed by the cellular machinery. As some input DNA can persist for many hours as a naked and inactive template, and as replication increases copy numbers well above the input levels, this approach facilitates analysis of active minichromosomes against the inevitable background due to the inactive input (Fig. 3.2). Whether nascent RNA copied from the different transcription units was then found in factories was analysed using RNA FISH. Establishing this method using minichromosomes as probes to detect factories requires careful control, optimisation, and quantification.

Choice of how much input to use for transfection depended on optimising the transfection efficiency, and then selecting the time when replication increased the active fraction of minichromosomes sufficiently (Fig. 3.3). Optimal fixation and permeabilisation conditions were chosen to ensure sufficient access of probes, as well as preserving nuclear structure (Fig. 4.7). Hybridisation conditions were adjusted so that negative controls using RNase and α-amanitin yield essentially no signal (Fig. 4.10). As read-through from neighbouring promoters is always a problem, a promoterless plasmid was constructed so that the extent of read-through from the SV40 promoter into the test transcription unit could be assessed. To enable comparison between results obtained on different days from different experiments,

all signals obtained by RNA FISH were carefully quantified and normalised relative to the fluorescence given by plastic reference beads (Fig. 4.8). Two patterns of RNA FISH signal were observed; a minority of the signal was diffusely spread throughout the nucleoplasm, while the majority was concentrated in discrete foci. Many others have shown that the latter mark the high concentrations of nascent RNA at the site of transcription, and the former completed transcripts on their way to the cytoplasm (Fig. 4.7) (Dirks et al., 1993; Levsky et al., 2002).

9.3 Minichromosomes are transcribed in factories

I first demonstrated that replicating minichromosomes give at least 8,000 copies per cell by 24 h by 'Southern' blots and qRT-PCR (Fig. 4.1 and 4.2). DNA FISH shows these minichromosomes are concentrated in only ~20 foci – or factories (Fig. 4.5). As the copy number increased, the number of plasmids per focus increased, while the number of foci remained the same. This suggests the newly-replicated plasmids remain close to their site of synthesis.

Next, RNA FISH shows nascent minichromosomal transcripts are clustered in a similar number of foci in the nucleus (Fig. 4.9). Although early studies suggests that transcription and replication occur in the same sites (reviewed by (Hassan and Cook, 1994)), there is no direct evidence in this study showing the foci containing minichromosomal RNA are sites where minichromosomes are also being replicated.

9.4 Factories specialise in transcription by pols I, II, and III

Using RNA FISH, I showed that transcripts copied from CMV and SV40 early promoters on the same type of plasmid always colocalise (Fig. 4.9), despite the

well-known phenomenon of transcription interference (Emerman and Temin, 1984; Eszterhas et al., 2002). This makes it likely that these two viral promoters initiate in the same factories.

I also demonstrated that transcription units bearing the same promoters (CMV pol II promoters) are usually transcribed in the same factories (Fig. 5.1). This is unlikely to be due to homologous base-paring as previously described (Binnie et al., 2006), as a substitution in the promoter region dramatically changed the localisation (below), suggesting this association is transcription dependent.

In addition, I showed by RNA FISH that pol II and III transcription units are transcribed in different factories (Fig. 5.2). This is in accord with earlier results showing that some factories contain nascent RNA made by pol II (but not III), while others contain transcripts made by pol III (but not II; (Pombo et al., 1999)). 3C confirms that templates bearing similar promoters are more likely to be together than those with different promoters (Figs 7.2 and 7.4). This result helps explain why active tRNA genes (which are transcribed by pol III) might silence pol II transcription (Scott et al., 2006). The tRNA gene may associate with a pol III factory (as it is transcribed), resulting in the relocation of neighbouring pol II genes distant from a pol II factory (Fig. 5.4).

In contrast to endogenous pol I genes, minichromosomal transcription units driven by pol I create mini-nucleoli in the nucleoplasm; this is supported by the demonstration (by immuno-FISH) that they contain the nucleolar marker, UBF (Fig. 5.3).

9.5 Polymerase II factories further specialise

Here, I demonstrate that transcription units controlled by different pol II promoters are transcribed in their own specialised factories.

When two plasmids encoding different pol II promoters (CMV, U2) are co-transfected, RNA FISH reveals the two kinds of nascent RNA to be in different foci (Fig. 6.1). In addition, 3C showed minichromosomal *U2* genes also lie near their counterparts on host chromosomes (Fig. 7.5); this makes it likely they share the same factory.

Introns also target genes to particular factories, presumably the sub-set containing the splicing machinery (Fig. 6.3); however, such targeting is less precise, as ~20% foci contain nascent transcripts with and without introns (Fig. 6.2).

These results suggest that a promoter – whether on a minichromosome or a host chromosome – must first diffuse to a factory of appropriate type before it can be transcribed.

9.6 Specialised transcription factories

In conclusion, using FISH and 3C, I demonstrated that minichromosomes are transcribed in factories which specialise in transcribing different genes depending on the kind of promoter the contain, whether or not they contain introns, and the type of signals at the 3' end.

Factories can be first categorised into pol I, II, and III classes, according to their polymerase content; then, they can be sub-categorised into different sub-classes, according to promoter type and the processing factors that are required (Fig. 9.1).

Fig. 9.2 summarises what we know of the size and number of the different specialised factories. All nucleoplasmic minichromosomal factories (pol II or III) appear to have a similar size in the light microscope, ~7 pixels (diameter ~250 nm). Of course, if these factories have true dimensions below the resolution of the light microscopes I used, they will nevertheless appear larger. These apparent dimensions are larger than a typical nucleoplasmic factory (seen in mammalian cells in tissue culture by electron microscopy) which has a diameter of ~50 nm. Pol I minichromosomal factories appear much larger, ~30 pixels (diameter ~550 nm). The true diameter of these various minichromosomal factories can be determined by electron microscopy.

All the different plasmids gave a roughly similar number of foci per cell, ~20, although there were fewer U2 snRNA and pol III factories (one-way ANNOVA $P<0.05$). Pol I factories were also more abundant than the others (i.e., ~40 per cell). The number of the different types of factories seems to be related to the different types of polymerase they contain.

These results beg many questions. For example, how many different types of factory are there? What is the recognition mechanism that targets a minichromosome to particular factories? There is some evidence that certain factories contain high concentrations of particular factors (reviewed by (Bartlett et al., 2006)). Then, as a promoter diffuses through the nucleus it might only bind stably to a factory containing the appropriate factors, to increase the frequency of initiation.

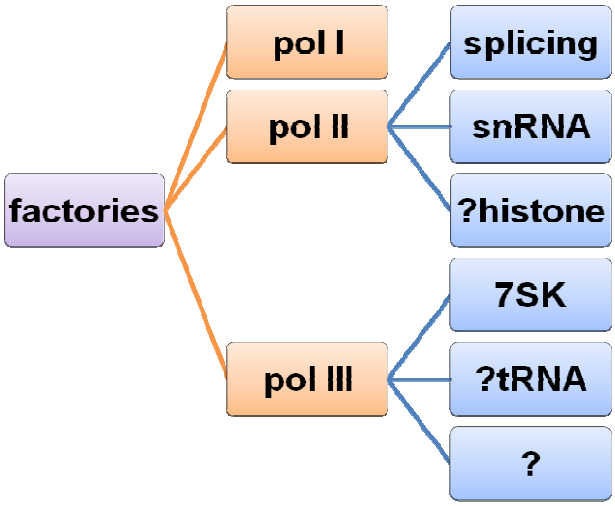


Fig. 9.1. Specialisation of transcription factories

Factories can be sub-divided into class I, II, and III, according to the different types of polymerase they contain. Then pol II factories further specialise in transcribing protein-coding or structural RNAs. Presumably pol III factories also further specialise in producing tRNAs and other structural RNAs.

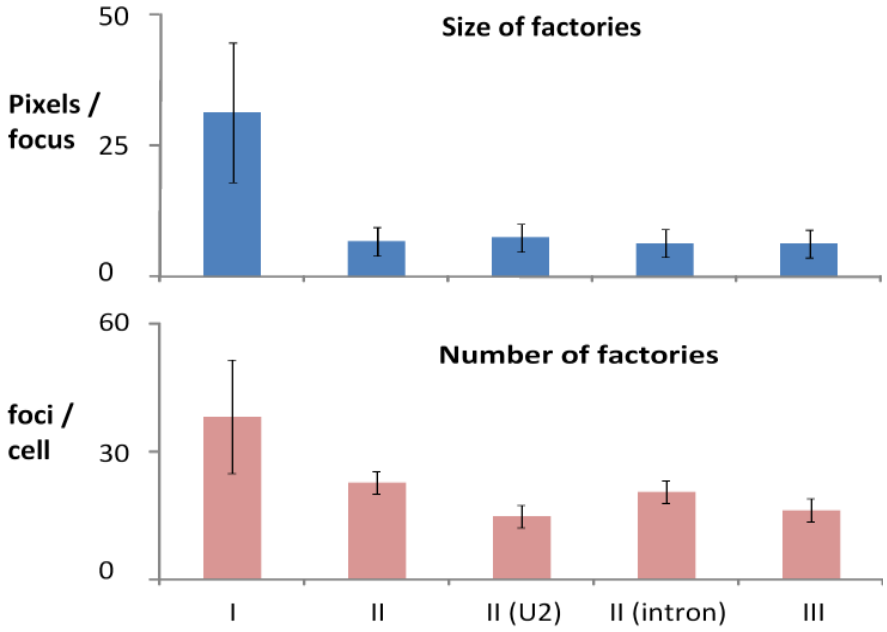


Fig. 9.2. Size and number of different specialised factories

Size and number of factories were measured from images like those shown in Ch.5 and 6. Essentially all nucleoplasmic minichromosomal factories have a roughly similar (apparent) size (~7 pixels) and number per cell (~20).

9.7 Transcription-driven chromosome pairing

These results shed light on the pairing of homologous chromosomes that occurs in both somatic and meiotic cells (Fig. 9.3B). For example, during meiosis, homologs seek out and align with their partners prior to the close synapsis that occurs during recombination. It is now accepted that distinct mechanisms underlie alignment and synapsis, as homologs still align in mutants unable to carry out the later steps (McKee, 2004; Gerton and Hawley, 2005). Chromosomes only pair during the initial alignment when they are transcriptionally active (reviewed in (Cook, 1997)). As a result, homologs will inevitably share similar arrays of promoters, processing signals, etc. along their length. Then, we can imagine that two homologs (which we will call 2m and 2p for chromosome 2 maternal and paternal, respectively) will generally be organised into two similar arrays of factories. If two homologous genes (which we will call *orange*^{m} on the maternal chromosome, and *orange*^{p} on the paternal one) happen to find themselves in close proximity within the nucleus, there is a chance that *orange*^{m} might bind to, and initiate in, a factory mainly organised by 2p. As this factory specialises in transcription of *orange*-like genes, the two chromosomes will be tied together for as long as it takes to transcribe *orange*^{m}. This increases the chances that an adjacent promoter (e.g., driving *grey*^{m}) might bind to the homologous factory on 2p to further stabilize the association between the two homologs. Through a continuation of this process, we can imagine the chromosomes will eventually pair (i.e., become zipped up) with their partners. Note that this alignment is not driven by base pairing between homologs. Subsequently, base-pairing between sequences on the two homologs will drive the

later steps of synapsis, and recombination (McKee, 2004; Gerton and Hawley, 2005)..

Transcription-based pairing of this type can also have direct effects on gene expression. The earliest example is probably provided by the "transvection" seen in the *Ultrabithorax* (*Ubx*) locus in *Drosophila* (Lewis, 1954). One likely explanation of this phenomenon suggests a functional enhancer element on the chromosome with the structural mutation acts *in trans* on a target promoter located on the homologous chromosome (Geyer et al., 1990; Wu, 1993; Cook, 1997; Wu and Morris, 1999). We would suggest that *Ubx* genes are transcribed in a special factory, and that an enhancer on one homolog and a promoter on the other occasionally find themselves in the same factory, and so can communicate effectively with each other. Of course, pairing along the length of the two homologs will increase the chances that the two find themselves in the same factory (as in Fig. 9.3B) and so can communicate, and this is consistent with the loss of transvection seen when pairing is disrupted (e.g., by introducing an inversion).

Another example where pairing affects gene expression, is provided by X chromosome inactivation. When more than one X is present in a diploid mammalian nucleus, one of the two X chromosomes is active. Activity is regulated by the X-inactivation centre (*Xic*). On the active X, transcription of *Xite* activates the linked *Tsix* allele, repressing the antisense counterpart *Xist*; on the other inactive X, *Xist* is transcribed, repressing other genes on the same X (Penny et al., 1996; Lucchesi et al., 2005). X-X pairing is restricted to *Xic*, and occurs just before or during *Xist* up-regulation (Xu et al., 2006). In other words, the two X chromosomes come together only when *Xist* and *Tsix* are actively transcribed, perhaps to share one factory which contains the necessary machinery that allows the two genes *in trans* to be transcribed

in a mutual exclusive manner. Deleting *Xite* and *Tsix* perturbs pairing, whereas ectopic X-autosome interactions inhibit endogenous X-X pairing and block the initiation of X-chromosome inactivation (Xu et al., 2006). Moreover, Pol II inhibition by actinomycin D or α-amanitin resulted in a consistent disruption of X-X pairing (Xu et al., 2007). Once again, a specialised factory brings together genes on two different chromosomes so that they can interact.

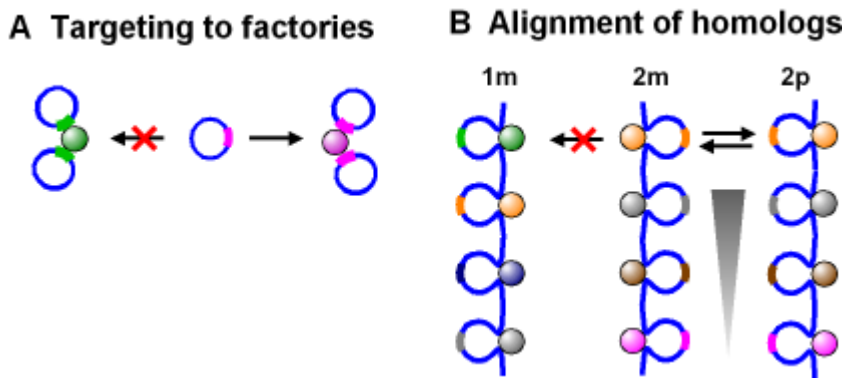


Fig. 9.3. Transcription-driven chromosome pairing

A. Targeting to factories. The promoter encoded by the minichromosome in the centre only initiates in a factory with the appropriate machinery (i.e., with a similar colour). As a result, it forms a cluster with two similar minichromosomes by association with the factory on the right.

B. Chromosome pairing. During alignment, homologs are transcriptionally active (Cook, 1997), so that each chromosome in the haploid set will possess a unique array of active transcription units running from telomere to telomere; only the homolog will possess a similar array. Here, only one of the many loops associated with a factory is shown. The orange promoter on maternal chr2 (2m) is unlikely to bind to the green factory on maternal chr1 (1m). But just as a specialised factory facilitates pairing between minichromosomes bearing similar transcription units, correct alignment begins when the orange (2m) promoter binds to the orange factory on its homolog (2p). Once transcription of the orange promoter on 2m begins, 2m and 2p become temporarily tethered together, and this will increase the chances that adjacent promoters bind to homologous factories (i.e, the grey unit on 2m with the grey factory on 2p, etc.). As a result, 2m and 2p eventually become zipped together (arrowhead), and so aligned.

9.8 Direct observation of a transcription wave

Using TNFα to switch on transcription of selected human genes rapidly and synchronously, T. Kodama's group generated a high temporal (i.e., at 7.5 min intervals) and spatial (i.e., every 13 nt) resolution profile of nascent transcripts in responsive genes using tiling arrays. I then used probes targeting intronic regions that were sufficiently sensitive to detect a single transcript copied from one of the genes (i.e., *SAMD4*; Fig. 8.2) to verify the microarray results, and to monitor a transcription cycle at the level of a single cell. The two methods yield essentially similar results.

RNA FISH gave the following results.

First, soon after the addition of TNFα, a wave of transcription swept along *SAMD4* at around 3.5 kb/min (Fig. 8.1 and 8.3). This wave reflects the movement of a convoy or a single polymerase from 5' to 3' of a transcription unit during the first 90 min; subsequently (i.e., between 90-180 min) a second wave appeared, but this could result from the firing of additional promoters and/or complex effects on transcript stability (Fig. 8.4).

Second, introns are co-transcriptionally degraded (presumably after splicing). Single labelling showed a 5' intron was removed before the wave reached a later intron (Fig. 8.3). Moreover, signals from both an early intron and later intron were never seen at the same allele by double labelling (Fig. 8.4).

Finally, productive transcription is a rare event regulated by a checkpoint. Most cells possess only one or no transcription foci (Table 8.3). Double-labelling also indicates that each allele has a maximum of one transcript associated with it at any time (Fig. 8.4), as all transcription foci detected have such similar intensities

and size (Table. 8.3; Fig. 8.5). Moreover, transcription of one intron seems to suppress transcription of another intron on the same allele.

9.9 Further experiments

Further experiments are required to investigate these specialised factories in more detail. To establish a more mature system, an inducible promoter could be used to observe the changing 'targeting' to a specialised factory, perhaps by substituting the CMV promoter for a heat-shock promoter (Pelham, 1982), or one regulated by the 'Tet-on, Tet-off' system (Gossen and Bujard, 1992). Inducible SV40 (Ameres et al., 2005) and pol III (Meissner et al., 2001) promoters can also be tested .

A simultaneous labelling of transcription sites and RNA FISH on cryosections should enable a higher-resolution study of the anatomy of the transcription foci observed in this study. Ultimately, these factories should be analysed by electron microscopy to determine their size, and whether they are a cluster of several smaller factories like the ones in untransfected cells, or if each focus represents one special factory of larger size that contains many more active genes. Tomographical techniques can also be applied to obtain detailed information at the ultrastructural level.

I showed by ChIP-3C that similar host and minichromosomal genes share the same factories (Fig. 7.5); this can be confirmed by RNA FISH.

Whether minichromosome replication and transcription occur at the same foci remains to be answered by combined DNA and RNA FISH (van Raamsdonk and Tilghman, 2001; Lomvardas et al., 2006), or a RNA FISH combined with a PCNA immuno-fluorescence.

We also wish to know how many different kinds of factories exist in a cell, to complete the preliminary diagram in Fig. 9.1. Plasmids encoding histone genes, microRNAs, tRNAs, and promoters with different types of motifs could be constructed, and similar experiments involving RNA FISH and 3C performed. We might, for example, construct plasmids with different combinations of the 5 major elements found within polymerase II promoters – a TFIIB recognition element (BRE; positions -37 to -32 relative to the transcription start site or TSS), TATA box (positions -31 to -26), initiator (Inr; position -2 to +4), motif 10 element (MTE; precisely positioned at +18 to +27), and downstream core promoter element (DPE; +28 to +32; Smale and Kadonaga, 2003; Gross and Oelgeschlager, 2006). [These motifs are degenerate, and most promoters in the human genome contain only some of them (e.g., Jin *et al*., 2006). Thus, only 9 of >9,000 promoters analysed contain all 5 elements, and the synthetic "super core promoter 1" (SCP1) – which is stronger than the CMV IE1 and AdML core promoters both *in vitro* and *in vivo* (Juven-Gershon *et al*., 2006) – contains 4 (i.e., TATA-Inr-MTE-DPE). Many promoters contain 3 motifs (e.g., TATA-Inr-MTE or BRE-Inr-MTE), many others only 2 (mainly DPE-MTE, TATA-MTE, and Inr-MTE), and 3% contain none (Jin *et al*., 2006). TATA boxes were thought to be present in most promoters, but recent surveys show that >70-80% human promoters lack a TATA box (e.g., Jin *et al*., 2006).]

In the long term, it should be possible to visualise these factories in living cells, using the appropriate fluorescent protein based RNA reporter (Daigle and Ellenberg, 2007; Darzacq et al., 2007).

Following the discovery of specialised factories, the next obvious step is to identify their components. RNA-TRAP (illustrated in Fig. 9.3) could be applied to

purify the proteins associated with a specific minichromosomal factory (Chakalova et al., 2004), prior to analysis using 'western' blots, ChIP, or mass spectrometry. RNA TRAP could also be applied to investigate the mechanism enabling an intron-containing gene to detect the appropriate "splicing" factory (e.g., by comparing the protein content of chromatin pulled down after transfecting plasmids +/- introns using 'western' blots and antibodies directed against specific histone modifications, or by analysing the DNA sequences after selecting different protein fractions using ChIP). Similarly, nascent RNA could be selected after 'run-on' in Br-U.

Details of when and where introns are degraded as the wave of transcription sweeps along *SAMD4* can also be analysed using intronic probes scattered every 1,000 nt along the long intron. Double-labelling FISH and 3C can also be applied to see if different genes stimulated by TNFα come together to share the same factory.

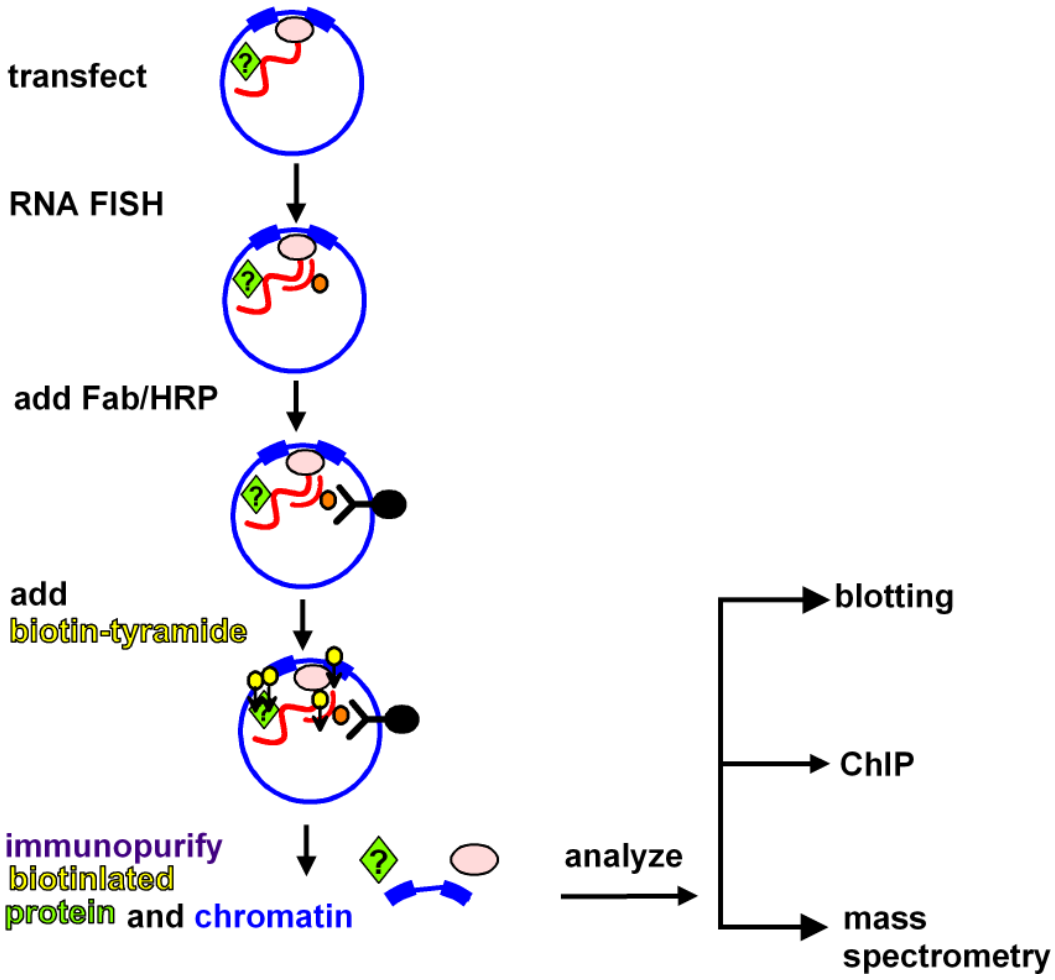


Fig. 9.4. Analysing protein components of specialised factories

Plasmids encoding transcription units with introns are transfected into cells, and RNA FISH performed using oligonucleotides labelled with haptens (orange oval) that target the intron sequence. Next, an HRP-conjugated Fab fragment is added to localise HRP activity to the site of transcription, before biotin-tyramide is added; this allows the HRP to activate the tyramide, which now attaches to proteins in close proximity. Then, biotinylated chromatin and proteins are immuno-purified, to be analysed using blotting, ChIP, or mass spectrometry. Adapted from (Chakalova et al., 2004).

9.10 Concluding remarks

I used replicating minichromosomes as probes to examine whether transcription occurs in factories, and whether factories specialise in transcribing particular sets of genes. Plasmids encoding the SV40 origin of replication are transfected into COS-7 cells, where they are assembled into minichromosomes. Using FISH, I discovered that minichromosomal transcription occurs in discrete foci, and that these foci specialise in transcribing different groups of genes. Pol I, II, and III units all have their own dedicated factories. Different pol II promoters and the presence of introns can also influence the nuclear location of transcription. Finally, using 3C, I confirmed that minichromosomes with similar promoters lie close together in 3D nuclear space, and that they are likely to share factories with similar endogenous genes.

After treating HUVECs with TNFα, I also used RNA FISH to observe a wave of transcription as it swept along a long gene (i.e., *SAMD4*). I found that introns are co-transcriptionally spliced, and that productive transcription is a rare event. It is also likely that only one polymerase is found on the gene at any time, and that a polymerase can either abort at a checkpoint ~8,000 nt after initiating, or continue to transcribe for many thousands of base pairs.

References

Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. and Block, S. M. (2005). Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460-5.

Aggarwal, B. B. (2003). Signalling pathways of the TNF superfamily: a double-edged sword. *Nat Rev Immunol* **3**, 745-56.

Ahn, S. H., Kim, M. and Buratowski, S. (2004). Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* **13**, 67-76.

Akhtar, A. and Matera, A. G. (2006). In and around the nucleus. *Nat Cell Biol* **8**, 3-6.

Alberts, B. (2002). Molecular biology of the cell. New York: Garland Science.

Allen, T. A., Von Kaenel, S., Goodrich, J. A. and Kugel, J. F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* **11**, 816-21.

Ameres, S. L., Drueppel, L., Pfleiderer, K., Schmidt, A., Hillen, W. and Berens, C. (2005). Inducible DNA-loop formation blocks transcriptional activation by an SV40 enhancer. *EMBO J* **24**, 358-367.

Armache, K. J., Kettenberger, H. and Cramer, P. (2003). Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A* **100**, 6964-8.

Arts, G. J., Fornerod, M. and Mattaj, I. W. (1998). Identification of a nuclear export receptor for tRNA. *Curr Biol* **8**, 305-14.

Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N. J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev* **11**, 2494-509.

Bannister, A. J., Schneider, R., Myers, F. A., Thorne, A. W., Crane-Robinson, C. and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. *J Biol Chem* **280**, 17732-6.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C. and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120-4.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-37.

Bartlett, J., Blagojevic, J., Carter, D., Eskiw, C., Fromaget, M., Job, C., Shamsher, M., Trindade, I. F., Xu, M. and Cook, P. R. (2006). Specialized transcription factories. *Biochem Soc Symp*, 67-75.

Batzer, M. A. and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat Rev Genet* **3**, 370-9.

Bazett-Jones, D. P., Leblanc, B., Herfort, M. and Moss, T. (1994). Short-range DNA looping by the Xenopus HMG-box transcription factor, xUBF. *Science* **264**, 1134-7.

Bell, S. P., Learned, R. M., Jantzen, H. M. and Tjian, R. (1988). Functional cooperativity between transcription factors UBF1 and SL1 mediates human ribosomal RNA synthesis. *Science* **241**, 1192-7.

Belmont, A. S. and Bruce, K. (1994). Visualization of G1 chromosomes: a folded, twisted, supercoiled chromonema model of interphase chromatid structure. *J Cell Biol* **127**, 287-302.

Berget, S. M. (1995). Exon recognition in vertebrate splicing. *J Biol Chem* **270**, 2411-4.

Beyer, A. L. and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev* **2**, 754-65.

Binnie, A., Castelo-Branco, P., Monks, J. and Proudfoot, N. J. (2006). Homologous gene sequences mediate transcription-domain formation. *J Cell Sci* **119**, 3876-87.

Bird, A. (1999). DNA methylation de novo. *Science* **286**, 2287-8.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21.

Bird, G., Zorio, D. A. and Bentley, D. L. (2004). RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol* **24**, 8963-9.

Birse, C. E., Minvielle-Sebastia, L., Lee, B. A., Keller, W. and Proudfoot, N. J. (1998). Coupling termination of transcription to messenger RNA maturation in yeast. *Science* **280**, 298-301.

Bjorklund, S. and Kim, Y. J. (1996). Mediator of transcriptional regulation. *Trends Biochem Sci* **21**, 335-7.

Bochkareva, E., Martynowski, D., Seitova, A. and Bochkarev, A. (2006). Structure of the origin-binding domain of simian virus 40 large T antigen bound to DNA. *EMBO J* **25**, 5961-9.

Boisvert, F. M., van Koningsbruggen, S., Navascues, J. and Lamond, A. I. (2007). The multifunctional nucleolus. *Nat Rev Mol Cell Biol* **8**, 574-85.

Bolton, E. C. and Boeke, J. D. (2003). Transcriptional interactions between yeast tRNA genes, flanking genes and Ty elements: a genomic point of view. *Genome Res* **13**, 254-63.

Bon, M., McGowan, S. J. and Cook, P. R. (2006). Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. *Faseb J* **20**, 1721-3.

Borchert, G. M., Lanier, W. and Davidson, B. L. (2006). RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* **13**, 1097-101.

Borsani, G., Tonlorenzi, R., Simmler, M. C., Dandolo, L., Arnaud, D., Capra, V., Grompe, M., Pizzuti, A., Muzny, D., Lawrence, C. et al. (1991). Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**, 325-9.

Boulikas, T. (1995). Chromatin domains and prediction of MAR sequences. *Int Rev Cytol* **162A**, 279-388.

Boyd, D. C., Turner, P. C., Watkins, N. J., Gerster, T. and Murphy, S. (1995). Functional redundancy of promoter elements ensures efficient transcription of the human 7SK gene in vivo. *J Mol Biol* **253**, 677-90.

Branco, M. R. and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol* **4**, e138.

Bregman, D. B., Du, L., van der Zee, S. and Warren, S. L. (1995). Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains. *J Cell Biol* **129**, 287-98.

Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E. and Palmiter, R. D. (1988). Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci U S A* **85**, 836-40.

Brown, J. M., Leach, J., Reittie, J. E., Atzberger, A., Lee-Prudhoe, J., Wood, W. G., Higgs, D. R., Iborra, F. J. and Buckle, V. J. (2006). Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol* **172**, 177-87.

Brown, K. (2002). Visualizing nuclear proteins together with transcribed and inactive genes in structurally preserved cells. *Methods* **26**, 10-8.

Bullock, P. A. (1997). The initiation of simian virus 40 DNA replication in vitro. *Crit Rev Biochem Mol Biol* **32**, 503-68.

Burnette, J. M., Miyamoto-Sato, E., Schaub, M. A., Conklin, J. and Lopez, A. J. (2005). Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements. *Genetics* **170**, 661-74.

Cai, S., Lee, C. C. and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet* **38**, 1278-88.

Cai, X., Hagedorn, C. H. and Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* **10**, 1957-66.

Cairns, J. (1960). The initiation of vaccinia infection. *Virology* **11**, 603-23.

Carmo-Fonseca, M., Mendes-Soares, L. and Campos, I. (2000). To be or not to be in the nucleolus. *Nat Cell Biol* **2**, E107-12.

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M. C., van Asperen, R., Boon, K., Voute, P. A. et al. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289-92.

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. F. and Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nat Genet* **32**, 623-6.

Cavalli, G. (2002). Chromatin as a eukaryotic template of genetic information. *Curr Opin Cell Biol* **14**, 269-78.

Celis, J. E. (1998). Cell biology : a laboratory handbook. San Diego ; London: Academic.

Chakalova, L., Carter, D. and Fraser, P. (2004). RNA fluorescence in situ hybridization tagging and recovery of associated proteins to analyze in vivo chromatin interactions. *Methods Enzymol* **375**, 479-93.

Chakalova, L., Debrand, E., Mitchell, J. A., Osborne, C. S. and Fraser, P. (2005). Replication and transcription: shaping the landscape of the genome. *Nat Rev Genet* **6**, 669-77.

Chen, D., Hinkley, C. S., Henry, R. W. and Huang, S. (2002). TBP dynamics in living human cells: constitutive association of TBP with mitotic chromosomes. *Mol Biol Cell* **13**, 276-84.

Chen, Y., Sinha, K., Perumal, K. and Reddy, R. (2000). Effect of 3' terminal adenylic acid residue on the uridylation of human small RNAs in vitro and in frog oocytes. *Rna* **6**, 1277-88.

Cho, E. J., Kobor, M. S., Kim, M., Greenblatt, J. and Buratowski, S. (2001). Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev* **15**, 3319-29.

Cho, E. J., Takagi, T., Moore, C. R. and Buratowski, S. (1997). mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev* **11**, 3319-26.

Chubb, J. R., Boyle, S., Perry, P. and Bickmore, W. A. (2002). Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol* **12**, 439-45.

Cmarko, D., Verschure, P. J., Rothblum, L. I., Hernandez-Verdun, D., Amalric, F., van Driel, R. and Fakan, S. (2000). Ultrastructural analysis of nucleolar transcription in cells microinjected with 5-bromo-UTP. *Histochem Cell Biol* **113**, 181-7.

Colgan, D. F. and Manley, J. L. (1997). Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11**, 2755-66.

Collins, C. A. and Guthrie, C. (2000). The question remains: is the spliceosome a ribozyme? *Nat Struct Biol* **7**, 850-4.

Conaway, J. W., Shilatifard, A., Dvir, A. and Conaway, R. C. (2000). Control of elongation by RNA polymerase II. *Trends in Biochemical Sciences* **25**, 375-80.

Cook, P. R. (1997). The transcriptional basis of chromosome pairing. *J Cell Sci* **110**, 1033-40.

Cook, P. R. (1999). Molecular biology - The organization of replication and transcription. *Science* **284**, 1790-1795.

Cook, P. R. (2001). Principles of nuclear structure and function. New York, NY ; Chichester: Wiley-Liss.

Cook, P. R. (2002). Predicting three-dimensional genome structure from transcriptional activity. *Nature Genetics* **32**, 347-352.

Cramer, P., Bushnell, D. A. and Kornberg, R. D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863-76.

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

Cuello, P., Boyd, D. C., Dye, M. J., Proudfoot, N. J. and Murphy, S. (1999). Transcription of the human U2 snRNA genes continues beyond the 3' box in vivo. *EMBO J* **18**, 2867-77.

Cullen, B. R. (2003). Nuclear RNA export. *J Cell Sci* **116**, 587-97.

Custodio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H. S., Grosveld, F. and Antoniou, M. (1999). Inefficient processing impairs release of RNA from the site of transcription. *Embo J* **18**, 2855-66.

Custodio, N., Carvalho, C., Condado, I., Antoniou, M., Blencowe, B. J. and Carmo-Fonseca, M. (2004). In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *Rna* **10**, 622-33.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A. and Pirrotta, V. (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185-96.

Daigle, N. and Ellenberg, J. (2007). LambdaN-GFP: an RNA reporter system for live-cell imaging. *Nat Methods* **4**, 633-6.

Darnell, J. E., Baltimore, D. and Lodish, H. F. (1986). Molecular cell biology. New York. New York: Scientific American Books ; Distributed by W.H. Freeman.

Darzacq, X., Jady, B. E., Verheggen, C., Kiss, A. M., Bertrand, E. and Kiss, T. (2002). Cajal body-specific small nuclear RNAs: a novel class of 2'-O-methylation and pseudouridylation guide RNAs. *EMBO J* **21**, 2746-56.

Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S. M., Phair, R. D. and Singer, R. H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol*. **14**, 796 – 806.

Darzacq, X., Singer, R. H. and Shav-Tal, Y. (2005). Dynamics of transcription and mRNA export. *Curr Opin Cell Biol* **17**, 332-9.

Das, R., Yu, J., Zhang, Z., Gygi, M. P., Krainer, A. R., Gygi, S. P. and Reed, R. (2007). SR Proteins Function in Coupling RNAP II Transcription to Pre-mRNA Splicing. *Mol Cell* **26**, 867-881.

Davie, J. R. (1995). The nuclear matrix and the regulation of chromatin organization and function. *Int Rev Cytol* **162A**, 191-250.

de la Mata, M., Alonso, C. R., Kadener, S., Fededa, J. P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D. and Kornblihtt, A. R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* **12**, 525-32.

de Laat, W. (2007). Long-range DNA contacts: romance in the nucleus? *Curr Opin Cell Biol*.

de Laat, W. and Grosveld, F. (2003). Spatial organization of gene expression: the active chromatin hub. *Chromosome Res* **11**, 447-59.

Dean, D. A. (1997). Import of plasmid DNA into the nucleus is sequence specific. *Exp Cell Res* **230**, 293-302.

Dechampesme, A. M., Koroleva, O., Leger-Silvestre, I., Gas, N. and Camier, S. (1999). Assembly of 5S ribosomal RNA is required at a specific step of the pre-rRNA processing pathway. *J Cell Biol* **145**, 1369-80.

Deininger, P. L. and Batzer, M. A. (2002). Mammalian retroelements. *Genome Res* **12**, 1455-65.

Dekker, J. (2006). The three 'C' s of chromosome conformation capture: controls, controls, controls. *Nat Methods* **3**, 17-21.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306-11.

Dewannieux, M., Esnault, C. and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35**, 41-8.

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K. and Zhou, M. M. (1999). Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491-6.

Dickinson, P., Cook, P. R. and Jackson, D. A. (1990). Active RNA polymerase I is fixed within the nucleus of HeLa cells. *EMBO J* **9**, 2207-14.

Dirks, R. W., van de Rijke, F. M., Fujishita, S., van der Ploeg, M. and Raap, A. K. (1993). Methodologies for specific intron and exon RNA localization in cultured cells by haptenized and fluorochromized probes. *J Cell Sci* **104**, 1187-97.

Dominski, Z. and Marzluff, W. F. (1999). Formation of the 3' end of histone mRNA. *Gene* **239**, 1-14.

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. and Mattick, J. S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* **19**, 4008.

Dundr, M., Hoffmann-Rohrer, U., Hu, Q., Grummt, I., Rothblum, L. I., Phair, R. D. and Misteli, T. (2002). A kinetic framework for a mammalian RNA polymerase in vivo. *Science* **298**, 1623-6.

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

Dvir, A. (2002). Promoter escape by RNA polymerase II. *Biochim Biophys Acta* **1577**, 208-223.

Dynan, W. S. and Tjian, R. (1985). Control of eukaryotic messenger RNA synthesis by sequence-specific DNA-binding proteins. *Nature* **316**, 774-8.

Emerman, M. and Temin, H. M. (1984). Genes with promoters in retrovirus vectors can be independently suppressed by an epigenetic mechanism. *Cell* **39**, 449-67.

Eskiw, C. H., Dellaire, G. and Bazett-Jones, D. P. (2004). Chromatin contributes to structural integrity of promyelocytic leukemia bodies through a SUMO-1-independent mechanism. *J Biol Chem* **279**, 9577-85.

Espinoza, C. A., Allen, T. A., Hieb, A. R., Kugel, J. F. and Goodrich, J. A. (2004). B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* **11**, 822-9.

Eszterhas, S. K., Bouhassira, E. E., Martin, D. I. and Fiering, S. (2002). Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Mol Cell Biol* **22**, 469-79.

Falkenberg, M., Gaspari, M., Rantanen, A., Trifunovic, A., Larsson, N.-G. and Gustafsson, C. M. (2002). Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nature Genetics* **31**, 289-294.

Fanning, E. and Knippers, R. (1992). Structure and function of simian virus 40 large tumor antigen. *Annu Rev Biochem* **61**, 55-85.

Farago, M., Nahari, T., Hammel, C., Cole, C. N. and Choder, M. (2003). Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Mol Biol Cell* **14**, 2744-55.

Faro-Trindade, I. and Cook, P. R. (2006). A conserved organization of transcription during embryonic stem cell differentiation and in cells with high C value. *Mol Biol Cell* **17**, 2910-20.

Fatica, A. and Tollervey, D. (2002). Making ribosomes. *Curr Opin Cell Biol* **14**, 313-8.

Fay, F. S., Taneja, K. L., Shenoy, S., Lifshitz, L. and Singer, R. H. (1997). Quantitative digital analysis of diffuse and concentrated nuclear distributions of nascent transcripts, SC35 and poly(A). *Exp Cell Res* **231**, 27-37.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* **132**, 6-13.

Femino, A. M., Fay, F. S., Fogarty, K. and Singer, R. H. (1998). Visualization of single RNA transcripts in situ. *Science* **280**, 585-90.

Femino, A. M., Fogarty, K., Lifshitz, L. M., Carrington, W. and Singer, R. H. (2003). Visualization of single molecules of mRNA in situ. *Methods Enzymol* **361**, 245-304.

Fiers, W., Contreras, R., Haegemann, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. and Ysebaert, M. (1978). Complete nucleotide sequence of SV40 DNA. *Nature* **273**, 113-20.

Finckh, U., Lingenfelter, P. A. and Myerson, D. (1991). Producing single-stranded DNA probes with the Taq DNA polymerase: a high yield protocol. *Biotechniques* **10**, 35-6, 38-9.

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

Fraser, P. and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**, 413-7.

French, S. L., Osheim, Y. N., Cioci, F., Nomura, M. and Beyer, A. L. (2003). In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. *Mol Cell Biol* **23**, 1558-68.

Frey, M. R., Bailey, A. D., Weiner, A. M. and Matera, A. G. (1999). Association of snRNA genes with coiled bodies is mediated by nascent snRNA transcripts. *Curr Biol* **9**, 126-35.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. *Curr Opin Genet Dev* **15**, 490-5.

Gagliardi, D., Stepien, P. P., Temperley, R. J., Lightowlers, R. N. and Chrzanowska-Lightowlers, Z. M. (2004). Messenger RNA stability in mitochondria: different means to an end.[erratum appears in Trends Genet. 2005 Jan;21(1):36]. *Trends in Genetics* **20**, 260-7.

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

Gall, J. G. and Murphy, C. (1998). Assembly of lampbrush chromosomes from sperm chromatin. *Mol Biol Cell* **9**, 733-47.

Gerbi, S. A., Borovjagin, A. V. and Lange, T. S. (2003). The nucleolus: a site of ribonucleoprotein maturation. *Curr Opin Cell Biol* **15**, 318-25.

Gerton, J. L. and Hawley, R. S. (2005). Homologous chromosome interactions in meiosis: diversity amidst conservation. *Nat Rev Genet* **6**, 477-87.

Geyer, P. K., Green, M. M. and Corces, V. G. (1990). Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. *EMBO J* **9**, 2247-56.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. and Kornberg, R. D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* **292**, 1876-82.

Goldman, R. D., Gruenbaum, Y., Moir, R. D., Shumaker, D. K. and Spann, T. P. (2002). Nuclear lamins: building blocks of nuclear architecture. *Genes Dev* **16**, 533-47.

Gossen, M. and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**, 5547-51.

GraphPad Prism version 4.00 for Windows, pp. www.graphpad.com. San Diego California USA: GraphPad Software.

Grande, M. A., van der Kraan, I., de Jong, L. and van Driel, R. (1997). Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *J Cell Sci* **110**, 1781-91.

Greger, I. H., Aranda, A. and Proudfoot, N. (2000). Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **97**, 8415-20.

Greger, I. H. and Proudfoot, N. J. (1998). Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae. *EMBO J* **17**, 4771-9.

Griffith, J., Dieckmann, M. and Berg, P. (1975). Electron microscope localization of a protein bound near the origin of simian virus 40 DNA replication. *J Virol* **15**, 167-72.

Grimm, C., Stefanovic, B. and Schumperli, D. (1993). The low abundance of U7 snRNA is partly determined by its Sm binding site. *EMBO J* **12**, 1229-38.

Gross, P., and Oelgeschlager, T. (2006). Core promoter-selective RNA polymerase II transcription. *Biochem. Soc. Symp.* **73**, 225-236.

Grummt, I. (2003). Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev* **17**, 1691-702.

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. and Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88.

Guillot, P. V., Xie, S. Q., Hollinshead, M. and Pombo, A. (2004). Fixation-induced redistribution of hyperphosphorylated RNA polymerase II in the nucleus of human cells. *Exp Cell Res* **295**, 460-8.

Gunnery, S., Ma, Y. and Mathews, M. B. (1999). Termination sequence requirements vary among genes transcribed by RNA polymerase III. *J Mol Biol* **286**, 745-57.

Haeusler, R. A. and Engelke, D. R. (2006). Spatial organization of transcription by RNA polymerase III. *Nucleic Acids Res*.

Hall, I. M., Shankaranarayana, G. D., Noma, K., Ayoub, N., Cohen, A. and Grewal, S. I. (2002). Establishment and maintenance of a heterochromatin domain. *Science* 297, 2232-7.

Hassan, A. B. and Cook, P. R. (1994). Does transcription by RNA polymerase play a direct role in the initiation of replication? *J Cell Sci* **107**, 1381-7.

Hassan, A. B., Errington, R. J., White, N. S., Jackson, D. A. and Cook, P. R. (1994). Replication and transcription sites are colocalised in human cells. *J Cell Sci* **107**, 425-34.

He, L. and Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**, 522-31.

Hedges, J., Chen, Y. I., West, M., Bussiere, C. and Johnson, A. W. (2006). Mapping the functional domains of yeast NMD3, the nuclear export adapter for the 60 S ribosomal subunit. *J Biol Chem* **281**, 36579-87.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A. et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8.

Henderson, A. S., Warburton, D. and Atwood, K. C. (1972). Location of ribosomal DNA in the human chromosome complement. *Proc Natl Acad Sci U S A* **69**, 3394-8.

Heng, H. H., Goetze, S., Ye, C. J., Liu, G., Stevens, J. B., Bremer, S. W., Wykes, S. M., Bode, J. and Krawetz, S. A. (2004). Chromatin loops are selectively anchored using scaffold/matrix-attachment regions. *J Cell Sci* **117**, 999-1008.

Hernandez, N. (1993). TBP, a universal eukaryotic transcription factor? *Genes Dev* **7**, 1291-308.

Hernandez, N. (2001). Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem* **276**, 26733-6.

Hilleren, P. and Parker, R. (1999). Mechanisms of mRNA surveillance in eukaryotes. *Annu Rev Genet* **33**, 229-60.

Hirose, T., Shu, M. D. and Steitz, J. A. (2003). Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol Cell* **12**, 113-23.

Hirose, Y. and Manley, J. L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev* **14**, 1415-29.

Ho, C. K. and Shuman, S. (1999). Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell* **3**, 405-11.

Hopper, A. K. and Phizicky, E. M. (2003). tRNA transfers to the limelight. *Genes Dev* **17**, 162-80.

Hozak, P., Cook, P. R., Schofer, C., Mosgoller, W. and Wachtler, F. (1994). Site of transcription of ribosomal RNA and intranucleolar structure in HeLa cells. *J Cell Sci* **107**, 639-48.

Hozak, P., Hassan, A. B., Jackson, D. A. and Cook, P. R. (1993). Visualization of replication factories attached to nucleoskeleton. *Cell* **73**, 361-73.

Huang, S. (2002). Building an efficient factory: where is pre-rRNA synthesized in the nucleolus? *J Cell Biol* **157**, 739-41.

Huang, S. and Spector, D. L. (1996). Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. *J Cell Biol* **133**, 719-32.

Huang, Y., Gattoni, R., Stevenin, J. and Steitz, J. A. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell* **11**, 837-43.

Hurt, E., Hannus, S., Schmelzl, B., Lau, D., Tollervey, D. and Simos, G. (1999). A novel in vivo assay reveals inhibition of ribosomal nuclear export in ran-cycle and nucleoporin mutants. *J Cell Biol* **144**, 389-401.

Iborra, F. J. and Cook, P. R. (1998). The size of sites containing SR proteins in human nuclei. Problems associated with characterizing small structures by immunogold labeling. *J Histochem Cytochem* **46**, 985-92.

Iborra, F. J., Escargueil, A. E., Kwek, K. Y., Akoulitchev, A. and Cook, P. R. (2004a). Molecular cross-talk between the transcription, translation, and nonsense-mediated decay machineries. *J Cell Sci* **117**, 899-906.

Iborra, F. J., Jackson, D. A. and Cook, P. R. (1998). The path of transcripts from extra-nucleolar synthetic sites to nuclear pores: transcripts in transit are concentrated in discrete structures containing SR proteins. *J Cell Sci* **111**, 2269-82.

Iborra, F. J., Jackson, D. A. and Cook, P. R. (2004b). The case for nuclear translation. *J Cell Sci* **117**, 5713-5720.

Iborra, F. J., Pombo, A., Jackson, D. A. and Cook, P. R. (1996). Active RNA polymerases are localized within discrete transcription "factories' in human nuclei. *J Cell Sci* **109**, 1427-36.

Jackson, D. A. and Cook, P. R. (1993). Transcriptionally Active Minichromosomes Are Attached Transiently in Nuclei through Transcription Units. *J Cell Sci* **105**, 1143-1150.

Jackson, D. A., Dickinson, P. and Cook, P. R. (1990). The size of chromatin loops in HeLa cells. *EMBO J* **9**, 567-71.

Jackson, D. A., Hassan, A. B., Errington, R. J. and Cook, P. R. (1993). Visualization of focal sites of transcription within human nuclei. *EMBO J* **12**, 1059-65.

Jackson, D. A., Iborra, F. J., Manders, E. M. and Cook, P. R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol Biol Cell* **9**, 1523-36.

Jackson, D. A., Pombo, A. and Iborra, F. (2000). The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells. *Faseb J* **14**, 242-54.

Jady, B. E., Darzacq, X., Tucker, K. E., Matera, A. G., Bertrand, E. and Kiss, T. (2003). Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J* **22**, 1878-88.

Janes, K. A., Gaudet, S., Albeck, J. G., Nielsen, U. B., Lauffenburger, D. A. and Sorger, P. K. (2006). The response of human epithelial cells to TNF involves an inducible autocrine cascade. *Cell* **124**, 1225-39.

Jenuwein, T. and Allis, C. D. (2001). Translating the histone code. *Science* **293**, 1074-80.

Jin, V.X., Singer, G.A., Agosto-Perez, F.J., Liyanarachchi, S., and Davuluri, R.V. (2006). Genome-wide analysis of core promoter elements from conserved human and mouse orthologous pairs. *BMC Bioinformatics* **7**, 114.

Joklik, W. K. (1968). The poxviruses. *Annu Rev Microbiol* **22**, 359-90.

Jolly, C., Metz, A., Govin, J., Vigneron, M., Turner, B. M., Khochbin, S. and Vourc'h, C. (2004). Stress-induced transcription of satellite III repeats. *J Cell Biol* **164**, 25-33.

Juven-Gershon, T., Cheng, S., and Kadonaga, J.T. (2006). Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, 917-922.

Kadonaga, J. T. (1998). Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* **92**, 307-13.

Kadonaga, J. T. (2004). Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors. *Cell* **116**, 247-57.

Kafatos, F. C., Jones, C. W. and Efstratiadis, A. (1979). Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res* **7**, 1541-52.

Kanno, T., Huettel, B., Mette, M. F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D. P., Matzke, M. and Matzke, A. J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* **37**, 761-5.

Kazazian, H. H., Jr. (2004). Mobile elements: drivers of genome evolution. *Science* **303**, 1626-32.

Kennedy, B. K., Barbie, D. A., Classon, M., Dyson, N. and Harlow, E. (2000). Nuclear organization of DNA replication in primary mammalian cells. *Genes Dev* **14**, 2855-68.

Keshet, I., Lieman-Hurwitz, J. and Cedar, H. (1986). DNA methylation affects the formation of active chromatin. *Cell* **44**, 535-43.

Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J. and Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15**, 2654-9.

Kim, T. H. and Ren, B. (2006). Genome-Wide Analysis of Protein-DNA Interactions. *Annu Rev Genomics Hum Genet* **7**, 81-102.

Kimura, H., Sugaya, K. and Cook, P. R. (2002). The transcription cycle of RNA polymerase II in living cells. *J Cell Biol* **159**, 777-82.

Kimura, H., Tao, Y., Roeder, R. G. and Cook, P. R. (1999). Quantitation of RNA polymerase II and its transcription factors in an HeLa cell: little soluble holoenzyme but significant amounts of polymerases attached to the nuclear substructure. *Mol Cell Biol* **19**, 5383-92.

Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**, 145-8.

Kiss, T. (2004). Biogenesis of small nuclear RNPs. *J Cell Sci* **117**, 5949-51.

Kleckner, N. (1996). Meiosis: how could it work? *Proc Natl Acad Sci U S A* **93**, 8167-74.

Koberna, K., Malinsky, J., Pliss, A., Masata, M., Vecerova, J., Fialova, M., Bednar, J. and Raska, I. (2002). Ribosomal genes in focus: new transcripts label the dense fibrillar components and form clusters indicative of "Christmas trees" in situ. *J Cell Biol* **157**, 743-8.

Koleske, A. J. and Young, R. A. (1994). An RNA polymerase II holoenzyme responsive to activators. *Nature* **368**, 466-9.

Komarnitsky, P., Cho, E. J. and Buratowski, S. (2000). Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14**, 2452-60.

Kosak, S. T. and Groudine, M. (2004). Gene order and dynamic domains. *Science* **306**, 644-7.

Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* **65**, 367-409.

Kressler, D., Linder, P. and de La Cruz, J. (1999). Protein trans-acting factors involved in ribosome biogenesis in Saccharomyces cerevisiae. *Mol Cell Biol* **19**, 7897-912.

Kuhn, A., Vente, A., Doree, M. and Grummt, I. (1998). Mitotic phosphorylation of the TBP-containing factor SL1 represses ribosomal gene transcription. *J Mol Biol* **284**, 1-5.

Kumaran, R. I., Muralikrishna, B. and Parnaik, V. K. (2002). Lamin A/C speckles mediate spatial organization of splicing factor compartments and RNA polymerase II transcription. *J Cell Biol* **159**, 783-93.

Lai, E. C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat Genet* **30**, 363-4.

Lamond, A. I. and Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4**, 605-12.

Lee, R. C., Feinbaum, R. L. and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75**, 843-54.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-9.

Lee, Y., Kim, M., Han, J., Yeom, K. H., Lee, S., Baek, S. H. and Kim, V. N. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**, 4051-60.

Levsky, J. M., Shenoy, S. M., Pezo, R. C. and Singer, R. H. (2002). Single-cell gene expression profiling. *Science* **297**, 836-40.

Lewis, E. (1954). The theory and application of a new method of detecting chromosomal rearrangements in Drosophila melanogaster. *American Naturalist* **88**, 225-239.

Lewis, J. D. and Izaurralde, E. (1997). The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem* **247**, 461-9.

Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., Cherry, A. M. and Hoffman, A. R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* **312**, 269-72.

Liu, W. M., Chu, W. M., Choudary, P. V. and Schmid, C. W. (1995). Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts. *Nucleic Acids Res* **23**, 1758-65.

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J. and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403-13.

Lucchesi, J. C., Kelly, W. G. and Panning, B. (2005). Chromatin remodeling in dosage compensation. *Annu Rev Genet* **39**, 615-51.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389**, 251-60.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E. and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* **303**, 95-8.

Mais, C., Wright, J. E., Prieto, J. L., Raggett, S. L. and McStay, B. (2005). UBF-binding site arrays form pseudo-NORs and sequester the RNA polymerase I transcription machinery. *Genes Dev* **19**, 50-64.

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

Margueron, R., Trojer, P. and Reinberg, D. (2005a). The key to development: interpreting the histone code? *Curr Opin Genet Dev* **15**, 163-76.

Margueron, R., Trojer, P. and Reinberg, D. (2005b). The key to development: interpreting the histone code? *Curr Opin Genetics Dev* **15**, 163-76.

Marshall, N. F. and Price, D. H. (1992). Control of formation of two distinct classes of RNA polymerase II elongation complexes. *Mol Cell Biol* **12**, 2078-90.

Marshall, N. F. and Price, D. H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *J Biol Chem* **270**, 12335-8.

Marzluff, W. F. (2005). Metazoan replication-dependent histone mRNAs: a distinct set of RNA polymerase II transcripts. *Curr Opin Cell Biol* **17**, 274-80.

Marzluff, W. F., Gongidi, P., Woods, K. R., Jin, J. and Maltais, L. J. (2002). The human and mouse replication-dependent histone genes. *Genomics* **80**, 487-98.

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

Matera, A. G., Frey, M. R., Margelot, K. and Wolin, S. L. (1995). A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. *J Cell Biol* **129**, 1181-93.

Matera, A. G., Terns, R. M. and Terns, M. P. (2007). Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* **8**, 209-20.

Matera, A. G. and Ward, D. C. (1992). Oligonucleotide probes for the analysis of specific repetitive DNA sequences by fluorescence in situ hybridization. *Hum Mol Genet* **1**, 535-9.

Mattaj, I. W. (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46**, 905-11.

Maul, G. G. (1998). Nuclear domain 10, the site of DNA virus transcription and replication. *Bioessays* **20**, 660-7.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S. and Bentley, D. L. (1997a). 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev* **11,** 3306-18.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M. and Bentley, D. L. (1997b). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357-61.

McCracken, S., Lambermon, M. and Blencowe, B. J. (2002). SRm160 splicing coactivator promotes transcript 3'-end cleavage. *Mol Cell Biol* **22**, 148-60.

McKee, B. D. (2004). Homologous pairing and chromosome dynamics in meiosis and mitosis. *Biochim Biophys Acta* **1677**, 165-80.

Mearini, G., Nielsen, P. E. and Fackelmayer, F. O. (2004). Localization and dynamics of small circular DNA in live mammalian nuclei. *Nucleic Acids Res* **32**, 2642-51.

Medlin, J. E., Uguen, P., Taylor, A., Bentley, D. L. and Murphy, S. (2003). The C-terminal domain of pol II and a DRB-sensitive kinase are required for 3' processing of U2 snRNA. *EMBO J* **22**, 925-34.

Meissner, W., Rothfels, H., Schafer, B. and Seifart, K. (2001). Development of an inducible pol III transcription system essentially requiring a mutated form of the TATA-binding protein. *Nucleic Acids Res* **29**, 1672-82.

Mellon, P., Parker, V., Gluzman, Y. and Maniatis, T. (1981). Identification of DNA-Sequences Required for Transcription of the Human Alpha-1-Globin Gene in a New Sv40 Host-Vector System. *Cell* **27**, 279-288.

Mellor, J. (2006). It takes a PHD to read the histone code. *Cell* **126**, 22-4.

Miller, O. L., Jr. (1981). The nucleolus, chromosomes, and visualization of genetic activity. *J Cell Biol* **91**, 15s-27s.

Miller, O. L., Jr. and Bakken, A. H. (1972). Morphological studies of transcription. *Acta Endocrinol Suppl (Copenh)* **168**, 155-77.

Minami, T., Horiuchi, K., Miura, M., Abid, M. R., Takabe, W., Noguchi, N., Kohro, T., Ge, X., Aburatani, H., Hamakubo, T. et al. (2004). Vascular endothelial growth factor- and thrombin-induced termination factor, Down syndrome critical region-1, attenuates endothelial cell proliferation and angiogenesis. *J Biol Chem* **279**, 50537-54.

Misteli, T. (2000). Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function. *J Cell Sci* **113**, 1841-9.

Misteli, T. (2001). Protein dynamics: implications for nuclear architecture and gene expression. *Science* **291**, 843-7.

Modrek, B. and Lee, C. (2002). A genomic view of alternative splicing. *Nat Genet* **30**, 13-9.

Modrek, B., Resch, A., Grasso, C. and Lee, C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res* **29**, 2850-9.

Moir, R. D., Yoon, M., Khuon, S. and Goldman, R. D. (2000). Nuclear lamins A and B1: different pathways of assembly during nuclear envelope formation in living cells. *J Cell Biol* **151**, 1155-68.

Montoya, J., Ojala, D. and Attardi, G. (1981). Distinctive Features of the 5'-Terminal Sequences of the Human Mitochondrial Messenger-Rnas. *Nature* **290**, 465-470.

Moore, M. J. and Sharp, P. A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**, 364-8.

Morillon, A., J. O'Sullivan, A. Azad, N. Proudfoot, and J. Mellor. (2003). Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. *Science* **300,** 492-495.

Morillon, A., O'Sullivan, J., Azad, A., Proudfoot, N. and Mellor, J. (2003). Regulation of elongating RNA polymerase II by forkhead transcription factors in yeast. *Science* **300**, 492-5.

Moy, T. I. and Silver, P. A. (2002). Requirements for the nuclear export of the small ribosomal subunit. *J Cell Sci* **115**, 2985-95.

Munkel, C., Eils, R., Dietzel, S., Zink, D., Mehring, C., Wedemann, G., Cremer, T. and Langowski, J. (1999). Compartmentalization of interphase chromosomes observed in simulation and experiment. *J Mol Biol* **285**, 1053-65.

Murphy, S., Tripodi, M. and Melli, M. (1986). A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes. *Nucleic Acids Res* **14**, 9243-60.

Myers, R. M., Rio, D. C., Robbins, A. K. and Tjian, R. (1981). SV40 gene expression is modulated by the cooperative binding of T antigen to DNA. *Cell* **25**, 373-84.

Nakamura, H., Morita, T. and Sato, C. (1986). Structural organizations of replicon domains during DNA synthetic phase in the mammalian nucleus. *Exp Cell Res* **165**, 291-7.

Navarro, M. and Gull, K. (2001). A pol I transcriptional body associated with VSG mono-allelic expression in Trypanosoma brucei. *Nature* **414**, 759-63.

Nesic, D., Tanackovic, G. and Kramer, A. (2004). A role for Cajal bodies in the final steps of U2 snRNP biogenesis. *J Cell Sci* **117**, 4423-33.

Neugebauer, K. M. (2002). On the importance of being co-transcriptional. *J Cell Sci* **115**, 3865-71.

Nevins, J. R. (1983). The pathway of eukaryotic mRNA formation. *Annu Rev Biochem* **52**, 441-66.

Newman, A. J. (1997). The role of U5 snRNP in pre-mRNA splicing. *EMBO J* **16**, 5797-800.

Ng, H. H. and Bird, A. (1999). DNA methylation and chromatin modification. *Curr Opin Genet Dev* **9**, 158-63.

Ng, H. H., Robert, F., Young, R. A. and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11**, 709-19.

Nguyen, V. T., Giannoni, F., Dubois, M. F., Seo, S. J., Vigneron, M., Kedinger, C. and Bensaude, O. (1996). In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. *Nucleic Acids Res* **24**, 2924-9.

Nguyen, V. T., Kiss, T., Michels, A. A. and Bensaude, O. (2001). 7SK small nuclear RNA binds to and inhibits the activity of CDK9/cyclin T complexes. *Nature* **414**, 322-5.

O'Brien, T. and Lis, J. T. (1991). RNA polymerase II pauses at the 5' end of the transcriptionally induced Drosophila hsp70 gene. *Mol Cell Biol* **11**, 5285-90.

Ogden, R. C., Lee, M. C. and Knapp, G. (1984). Transfer RNA splicing in Saccharomyces cerevisiae: defining the substrates. *Nucleic Acids Res* **12**, 9367-82.

Ohlsson, R. and Gondor, A. (2007). The 4C technique: the 'Rosetta stone' for genome biology in 3D? *Curr Opin Cell Biol*.

Ohno, M., Segref, A., Bachi, A., Wilm, M. and Mattaj, I. W. (2000). PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187-98.

Ohno, M., Segref, A., Kuersten, S. and Mattaj, I. W. (2002). Identity elements used in export of mRNAs. *Mol Cell* **9**, 659-71.

Okamoto, I., Otte, A. P., Allis, C. D., Reinberg, D. and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* **303**, 644-9.

Omilli, F., Ernoult-Lange, M., Borde, J. and May, E. (1986). Sequences involved in initiation of simian virus 40 late transcription in the absence of T antigen. *Mol Cell Biol* **6**, 1875-85.

Orphanides, G., Lagrange, T. and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev* **10**, 2657-83.

Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W. et al. (2004). Active genes dynamically colocalise to shared sites of ongoing transcription. *Nat Genet* **36**, 1065-71.

Osborne, C. S., Chakalova, L., Mitchell, J. A., Horton, A., Wood, A. L., Bolland, D. J., Corcoran, A. E. and Fraser, P. (2007). Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol* **5**, e192.

Osheim, Y. N., Proudfoot, N. J. and Beyer, A. L. (1999). EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol Cell* **3**, 379-87.

Pabo, C. O. and Sauer, R. T. (1992). Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* **61**, 1053-95.

Palacios, I., Hetzer, M., Adam, S. A. and Mattaj, I. W. (1997). Nuclear import of U snRNPs requires importin beta. *EMBO J* **16**, 6783-92.

Pandey, N. B., Chodchoy, N., Liu, T. J. and Marzluff, W. F. (1990). Introns in histone genes alter the distribution of 3' ends. *Nucleic Acids Res* **18**, 3161-70.

Paushkin, S., Gubitz, A. K., Massenet, S. and Dreyfuss, G. (2002). The SMN complex, an assemblyosome of ribonucleoproteins. *Curr Opin Cell Biol* **14**, 305-12.

Pavelitz, T., Rusche, L., Matera, A. G., Scharf, J. M. and Weiner, A. M. (1995). Concerted evolution of the tandem array encoding primate U2 snRNA occurs in situ, without changing the cytological context of the RNU2 locus. *EMBO J* **14**, 169-77.

Pederson, T. (2000). Half a century of "the nuclear matrix". *Mol Biol Cell* **11**, 799-805.

Pelham, H. R. (1982). A regulatory upstream promoter element in the Drosophila hsp 70 heat-shock gene. *Cell* **30**, 517-28.

Peng, J., Zhu, Y., Milton, J. T. and Price, D. H. (1998). Identification of multiple cyclin subunits of human P-TEFb. *Genes Dev* **12**, 755-62.

Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. and Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-7.

Pestic-Dragovich, L., Stojiljkovic, L., Philimonenko, A. A., Nowak, G., Ke, Y., Settlage, R. E., Shabanowitz, J., Hunt, D. F., Hozak, P. and de Lanerolle, P. (2000). A myosin I isoform in the nucleus. *Science* **290**, 337-41.

Piper, P. W., Bellatin, J. A. and Lockheart, A. (1983). Altered maturation of sequences at the 3' terminus of 5S gene transcripts in a Saccharomyces cerevisiae mutant that lacks a RNA processing endonuclease. *EMBO J* **2**, 353-9.

Pleschka, S., Jaskunas, R., Engelhardt, O. G., Zurcher, T., Palese, P. and Garcia-Sastre, A. (1996). A plasmid-based reverse genetics system for influenza A virus. *J Virol* **70**, 4188-92.

Pombo, A. and Cook, P. R. (1996). The localization of sites containing nascent RNA and splicing factors. *Exp Cell Res* **229**, 201-3.

Pombo, A., Cuello, P., Schul, W., Yoon, J. B., Roeder, R. G., Cook, P. R. and Murphy, S. (1998). Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle. *EMBO J* **17**, 1768-78.

Pombo, A., Ferreira, J., Bridge, E. and Carmo-Fonseca, M. (1994). Adenovirus replication and transcription sites are spatially separated in the nucleus of infected cells. *EMBO J* **13**, 5075-85.

Pombo, A., Jackson, D. A., Hollinshead, M., Wang, Z., Roeder, R. G. and Cook, P. R. (1999). Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J* **18**, 2241-53.

Pontes, O., Li, C. F., Nunes, P. C., Haag, J., Ream, T., Vitins, A., Jacobsen, S. E. and Pikaard, C. S. (2006). The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. *Cell* **126**, 79-92.

Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., Zhang, M. Q. and Spector, D. L. (2005). Regulating Gene Expression through RNA Nuclear Retention. *Cell* **123**, 249-63.

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* **16**, 272-8.

Proudfoot, N. J., Furger, A. and Dye, M. J. (2002). Integrating mRNA processing with transcription. *Cell* **108**, 501-12.

Purmann, A., Toedling, J., Schueler, M., Carninci, P., Lehrach, H., Hayashizaki, Y., Huber, W. and Sperling, S. (2007). Genomic organization of transcriptomes in mammals: Coregulation and cofunctionality. *Genomics* **89**, 580-7.

Rasband, W. S. (1997-2007). ImageJ, http://rsb.info.nih.gov/ij/: U. S. National Institutes of Health, Bethesda, Maryland, USA,.

Richard, P., Darzacq, X., Bertrand, E., Jady, B. E., Verheggen, C. and Kiss, T. (2003). A common sequence motif determines the Cajal body-specific localization of box H/ACA scaRNAs. *EMBO J* **22**, 4283-93.

Roeder, R. G. (1996a). Nuclear RNA polymerases: role of general initiation factors and cofactors in eukaryotic transcription. *Methods Enzymol* **273**, 165-71.

Roeder, R. G. (1996b). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327-35.

Roeder, R. G. (2003). Lasker Basic Medical Research Award. The eukaryotic transcriptional machinery: complexities and mechanisms unforeseen. *Nat Med* **9**, 1239-44.

Rossmanith, W., Tullo, A., Potuschak, T., Karwan, R. and Sbisa, E. (1995). Human mitochondrial tRNA processing. *J Biol Chem* **270**, 12885-91.

Rougvie, A. E. and Lis, J. T. (1990). Postinitiation transcriptional control in Drosophila melanogaster. *Mol Cell Biol* **10**, 6041-5.

Roussel, P., Andre, C., Comai, L. and Hernandez-Verdun, D. (1996). The rDNA transcription machinery is assembled during mitosis in active NORs and absent in inactive NORs. *J Cell Biol* **133**, 235-46.

Rubin, J. E., Pasceri, P., Wu, X., Leboulch, P. and Ellis, J. (2000). Locus control region activity by 5'HS3 requires a functional interaction with beta-globin gene regulatory elements: expression of novel beta/gamma-globin hybrid transgenes. *Blood* **95**, 3242-9.

Ryan, K. J. and Wente, S. R. (2000). The nuclear pore complex: a protein machine bridging the nucleus and cytoplasm. *Curr Opin Cell Biol* **12**, 361-71.

Sachs, R. K., van den Engh, G., Trask, B., Yokota, H. and Hearst, J. E. (1995). A random-walk/giant-loop model for interphase chromosomes. *Proc Natl Acad Sci U S A* **92**, 2710-4.

Saitoh, Y. and Laemmli, U. K. (1993). From the chromosomal loops and the scaffold to the classic bands of metaphase chromosomes. *Cold Spring Harb Symp Quant Biol* **58**, 755-65.

Samarsky, D. A., Fournier, M. J., Singer, R. H. and Bertrand, E. (1998). The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *EMBO J* **17**, 3747-57.

Sambrook, J. and Russell, D. W. (2001). Molecular cloning : a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.* **8,** 424-436.

Santos-Rosa, H., Schneider, R., Bannister, A. J., Sherriff, J., Bernstein, B. E., Emre, N. C., Schreiber, S. L., Mellor, J. and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407-11.

Schramm, L. and Hernandez, N. (2002). Recruitment of RNA polymerase III to its target promoters. *Genes Dev* **16**, 2593-620.

Scott, K. C., Merrett, S. L. and Willard, H. F. (2006). A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. *Curr Biol* **16**, 119-29.

Sedat, J. and Manuelidis, L. (1978). A direct approach to the structure of eukaryotic chromosomes. *Cold Spring Harb Symp Quant Biol* **42 Pt 1**, 331-50.

Shadel, G. S. and Clayton, D. A. (1997). Mitochondrial DNA maintenance in vertebrates. *Annu Rev Biochem,* 409-435.

Sharp, P. A. (1994). Split genes and RNA splicing. *Cell* **77**, 805-15.

Shatkin, A. J. and Manley, J. L. (2000). The ends of the affair: capping and polyadenylation. *Nat Struct Biol* **7**, 838-42.

Shaw, P. E., Bohmann, D. and Sergeant, A. (1985). The SV40 enhancer influences viral late transcription in vitro and in vivo but not on replicating templates. *EMBO J* **4**, 3247-52.

Shedlock, A. M. and Okada, N. (2000). SINE insertions: powerful tools for molecular systematics. *Bioessays* **22**, 148-60.

Shermoen, A. W. and O'Farrell, P. H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* **67**, 303-10.

Shiels, C., Islam, S. A., Vatcheva, R., Sasieni, P., Sternberg, M. J., Freemont, P. S. and Sheer, D. (2001). PML bodies associate specifically with the MHC gene cluster in interphase nuclei. *J Cell Sci* **114**, 3705-16.

Shuman, S. (1995). Capping enzyme in eukaryotic mRNA synthesis. *Prog Nucleic Acid Res Mol Biol* **50**, 101-29.

Shumyatsky, G. P., Tillib, S. V. and Kramerov, D. A. (1990). B2 RNA and 7SK RNA, RNA polymerase III transcripts, have a cap-like structure at their 5' end. *Nucleic Acids Res* **18**, 6347-51.

Shutt, T. E. and Gray, M. W. (2006). Bacteriophage origins of mitochondrial replication and transcription proteins. *Trends in Genetics* **22**, 90-5.

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-54.

Singh, R. and Reddy, R. (1989). Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci U S A* **86**, 8280-3.

Sleeman, J. E. and Lamond, A. I. (1999). Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr Biol* **9**, 1065-74.

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

Smith, K. P. and Lawrence, J. B. (2000). Interactions of U2 gene loci and their nuclear transcripts with Cajal (coiled) bodies: evidence for PreU2 within Cajal bodies. *Mol Biol Cell* **11**, 2987-98.

Sorensen, P. D. and Frederiksen, S. (1991). Characterization of human 5S rRNA genes. *Nucleic Acids Res* **19**, 4147-51.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**, 503-17.

Spann, T. P., Goldman, A. E., Wang, C., Huang, S. and Goldman, R. D. (2002). Alteration of nuclear lamin organization inhibits RNA polymerase II-dependent transcription. *J Cell Biol* **156**, 603-8.

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

Spector, D. L. (2001). Nuclear domains. *J Cell Sci* **114**, 2891-3.

Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. and Flavell, R. A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637-45.

Spiller, M. P., Boon, K. L., Reijns, M. A. and Beggs, J. D. (2007). The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Res* **35**, 923-9.

Splinter, E., Heath, H., Kooren, J., Palstra, R. J., Klous, P., Grosveld, F., Galjart, N. and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* **20**, 2349-54.

Steinmetz, E. J. (1997). Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? *Cell* **89**, 491-4.

Stoffler, D., Fahrenkrog, B. and Aebi, U. (1999). The nuclear pore complex: from molecular architecture to functional dynamics. *Curr Opin Cell Biol* **11**, 391-401.

Struhl, K. (1996). Chromatin structure and RNA polymerase II connection: implications for transcription. *Cell* **84**, 179-82.

Takaku, H., Minagawa, A., Takagi, M. and Nashimoto, M. (2003). A candidate prostate cancer susceptibility gene encodes tRNA 3' processing endoribonuclease. *Nucleic Acids Res* **31**, 2272-8.

Tennyson, C. N., Klamut, H. J. and Worton, R. G. (1995). The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat Genet* **9**, 184-90.

Terns, M. P. and Terns, R. M. (2002). Small nucleolar RNAs: versatile trans-acting molecules of ancient evolutionary origin. *Gene Expr* **10**, 17-39.

The ENCODE consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816.

Thompson, C. M., Koleske, A. J., Chao, D. M. and Young, R. A. (1993). A multisubunit complex associated with the RNA polymerase II CTD and TATA-binding protein in yeast. *Cell* **73**, 1361-75.

Thompson, M., Haeusler, R. A., Good, P. D. and Engelke, D. R. (2003). Nucleolar clustering of dispersed tRNA genes. *Science* **302**, 1399-401.

Tokuyasu, K. T. (1980). Immunochemistry on ultrathin frozen sections. *Histochem J* **12**, 381-403.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453-65.

Tran, D. P., Kim, S. J., Park, N. J., Jew, T. M. and Martinson, H. G. (2001). Mechanism of poly(A) signal transduction to RNA polymerase II in vitro. *Mol Cell Biol* **21**, 7495-508.

Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R. and Abelson, J. N. (1997). The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell* **89**, 849-58.

Turner, B. M. (2000). Histone acetylation and an epigenetic code. *Bioessays* **22**, 836-45.

Turner, B. M. (2007). Defining an epigenetic code. *Nat Cell Biol* **9**, 2-6.

Uguen, P. and Murphy, S. (2003). The 3' ends of human pre-snRNAs are produced by RNA polymerase II CTD-dependent RNA processing. *EMBO J* **22**, 4544-54.

Valadkhan, S. and Manley, J. L. (2000). A tertiary interaction detected in a human U2-U6 snRNA complex assembled in vitro resembles a genetically proven interaction in yeast. *Rna* **6**, 206-19.

van Raamsdonk, C. D. and Tilghman, S. M. (2001). Optimizing the detection of nascent transcripts by RNA fluorescence in situ hybridization. *Nucleic Acids Res* **29**, E42-2.

Varshavsky, A. J., Nedospasov, S. A., Schmatchenko, V. V., Bakayev, V. V., Chumackov, P. M. and Georgiev, G. P. (1977). Compact form of SV40 viral minichromosome is resistant to nuclease: possible implications for chromatin structure. *Nucleic Acids Res* **4**, 3303-25.

Venema, J. and Tollervey, D. (1999). Ribosome synthesis in Saccharomyces cerevisiae. *Annu Rev Genet* **33**, 261-311.

Vlcek, S. and Foisner, R. (2007). Lamins and lamin-associated proteins in aging and disease. *Curr Opin Cell Biol* **19**, 298-304.

Wahle, E. and Ruegsegger, U. (1999). 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev* **23**, 277-95.

Wang, J., Shiels, C., Sasieni, P., Wu, P. J., Islam, S. A., Freemont, P. S. and Sheer, D. (2004). Promyelocytic leukemia nuclear bodies associate with transcriptionally active genomic regions. *J Cell Biol* **164**, 515-26.

Wang, Z. and Roeder, R. G. (1997). Three human RNA polymerase III-specific subunits form a subcomplex with a selective function in specific transcription initiation. *Genes Dev* **11**, 1315-26.

Wang, Z. G., Delva, L., Gaboli, M., Rivi, R., Giorgio, M., Cordon-Cardo, C., Grosveld, F. and Pandolfi, P. P. (1998). Role of PML in cell growth and the retinoic acid pathway. *Science* **279**, 1547-51.

Wansink, D. G., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R. and de Jong, L. (1993). Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J Cell Biol* **122**, 283-93.

Weidemann, T., Wachsmuth, M., Knoch, T. A., Muller, G., Waldeck, W. and Langowski, J. (2003). Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal imaging. *J Mol Biol* **334**, 229-40.

Wetterberg, I., Zhao, J., Masich, S., Wieslander, L. and Skoglund, U. (2001). In situ transcription and splicing in the Balbiani ring 3 gene. *EMBO J* **20**, 2564-74.

White, R. J. (2004). RNA polymerase III transcription and cancer. *Oncogene* **23**, 3208-16.

White, R. J. (2005). RNA polymerases I and III, growth control and cancer. *Nat Rev Mol Cell Biol* **6**, 69-78.

Wijgerde, M., Grosveld, F. and Fraser, P. (1995). Transcription complex stability and chromatin dynamics in vivo. *Nature* **377**, 209-13.

Williams, M. A. (1977). Quantitative methods in biology. Amsterdam ; Oxford: North-Holland.

Wu, C. T. (1993). Transvection, nuclear structure, and chromatin proteins. *J Cell Biol* **120**, 587-90.

Wu, C. T. and Morris, J. R. (1999). Transvection and other homology effects. *Curr Opin Genet Dev* **9**, 237-46.

Wurtele, H. and Chartrand, P. (2006). Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology. *Chromosome Res* **14**, 477-95.

Xiao, S., Scott, F., Fierke, C. A. and Engelke, D. R. (2002). Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes. *Annu Rev Biochem* **71**, 165-89.

Xie, S. Q., Martin, S., Guillot, P. V., Bentley, D. L. and Pombo, A. (2006). Splicing speckles are not reservoirs of RNA polymerase II, but contain an inactive form, phosphorylated on serine2 residues of the C-terminal domain. *Mol Biol Cell* **17**, 1723-33.

Xing, Y., Johnson, C. V., Moen, P. T., Jr., McNeil, J. A. and Lawrence, J. (1995). Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J Cell Biol* **131**, 1635-47.

Xu, N., Donohoe, M. E., Silva, S. S. and Lee, J. T. (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet* **39**, 1390-6.

Xu, N., Tsai, C. L. and Lee, J. T. (2006). Transient homologous chromosome pairing marks the onset of x inactivation. *Science* **311**, 1149-52.

Yang, Z., Zhu, Q., Luo, K. and Zhou, Q. (2001). The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **414**, 317-22.

Yekta, S., Shih, I. H. and Bartel, D. P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. *Science* **304**, 594-6.

Yik, J. H., Chen, R., Nishimura, R., Jennings, J. L., Link, A. J. and Zhou, Q. (2003). Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. *Mol Cell* **12**, 971-82.

Yonaha, M. and Proudfoot, N. J. (2000). Transcriptional termination and coupled polyadenylation in vitro. *EMBO J* **19**, 3770-7.

Yue, Z., Maldonado, E., Pillutla, R., Cho, H., Reinberg, D. and Shatkin, A. J. (1997). Mammalian capping enzyme complements mutant Saccharomyces cerevisiae lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II. *Proc Natl Acad Sci U S A* **94**, 12898-903.

Zeng, Y., Wagner, E. J. and Cullen, B. R. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* **9**, 1327-33.

Zhang, L. F., Huynh, K. D. and Lee, J. T. (2007). Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell* **129**, 693-706.

Zhao, J., Hyman, L. and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63**, 405-45.

Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K. S., Singh, U. et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* **38**, 1341-7.

Zheng, L., Dominski, Z., Yang, X. C., Elms, P., Raska, C. S., Borchers, C. H. and Marzluff, W. F. (2003). Phosphorylation of stem-loop binding protein (SLBP) on two threonines triggers degradation of SLBP, the sole cell cycle-regulated factor required for regulation of histone mRNA processing, at the end of S phase. *Mol Cell Biol* **23**, 1590-601.

Zhou, G. L., Xin, L., Song, W., Di, L. J., Liu, G., Wu, X. S., Liu, D. P. and Liang, C. C. (2006). Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes. *Mol Cell Biol* **26**, 5096-105.

Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E. and Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**, 401-5.

Zirbel, R. M., Mathieu, U. R., Kurz, A., Cremer, T. and Lichter, P. (1993). Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Res* **1**, 93-106