ABSTRACT

CONSEQUENCES OF STOCHASTIC mRNA SYNTHESIS
IN A GENE REGULATORY PATHWAY

by
Khyati Shah

Gene expression is a stochastic process, with elements of randomness present in both transcription and translation. This stochasticity results in cell-to-cell variation in the amounts of gene products, mRNAs and proteins, and is observed in organisms ranging from bacteria and yeast to higher eukaryotes. Randomness in the activation and inactivation of a gene is the preliminary cause of this variation. At the level of proteins, these variations are buffered compared to levels of mRNA, due to the longer lifespan of proteins. Nevertheless, there is substantial variation observed at the level of proteins, resulting in phenotypic diversity among genetically identical cells.

In higher eukaryotes, sets of genes are often expressed in a coordinated manner, and function together in response to extracellular stimuli. If the expression of such genes is indeed stochastic, how can a given cell produce a coherent response? Additionally, during multi-subunit protein assembly, how does variation in levels of the component proteins affect their assembly and impact their function? Furthermore, how does this variation propagate in a gene regulatory pathway, when protein products of an upstream gene, or a pair of upstream genes, aids in the expression of downstream genes? Does variation in the expression of upstream genes affect the expression of downstream genes?

These questions are addressed using the serum-mediated induction of c-Fos and c-Jun as a model. c-Fos and c-Jun are transcription factors that together form heterodimers and induce the expression of downstream genes. With the aid of single-molecule fluorescence *in situ* hybridization for the detection of individual mRNA molecules, cell-to-cell variation in the expression of c-Fos and c-Jun mRNAs, and variations in the expression of mRNAs from a pair of downstream genes, collagenase and cox-2 were studied. Cell-to-cell variation in the number of c-Fos-c-Jun protein heterodimers in the nucleus was also studied. It was found that, although c-Fos and c-Jun mRNA expression is highly variable and not correlated, the number of the c-Fos-c-Jun protein hetrodimers did not vary as much from cell to cell. Despite relatively invariant heterodimer numbers, the downstream mRNAs, collagenase and cox-2, were expressed in a highly stochastic manner. These results suggest that, despite the buffering of variation in intermediate steps, the downstream steps in a gene regulatory pathway are noisy. These results are consistent with the view that noisy expression is an inherent property of the transcriptional machinery.

As a second project, where in the nucleus, and at what step during mRNA biogenesis, does mRNA splicing occur was explored. It is believed that splicing generally occurs co-transcriptionally at the gene locus. Introns are removed before the mRNA is released. However, during alternative splicing it is important that processing be delayed until all of the exons and introns involved in the splice choice are synthesized. Is processing just delayed briefly until the alternative splice sites are synthesized, or does alternative splicing involve the uncoupling of splicing from transcription, so that splicing occurs post-transcriptionally?

The intracellular distribution and dynamics of individual molecules of pre-mRNAs and their spliced products were imaged utilizing a set of synthetic reporter genes, as well as a classically well-studied alternatively spliced gene: Sex-lethal (Sxl) in *Drosophila*. The normally tight coupling between transcription and splicing was found to be broken in situations where an intron's polypyrimidine tract is sequestered within a strong secondary structure. Furthermore, it was also found that, in the case of the alternative splicing of Sxl mRNA in female *Drosophila* cells, particular exon is removed from the transcript, due to the activity of the RNA binding protein Sxl, which binds to nearby introns, causing splicing in those regions to be uncoupled from transcription. This uncoupling occurs only on the perturbed introns, while the preceding introns are removed co-transcriptionally.

CONSEQUENCES OF STOCHASTIC mRNA SYNTHESIS
IN A GENE REGULATORY PATHWAY

by

Khyati Shah

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
and Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Biology

Federated Biological Sciences Department

January 2012

APPROVAL PAGE

**CONSEQUENCES OF STOCHASTIC mRNA SYNTHESIS
IN A GENE REGULATORY PATHWAY**

Khyati Shah

Dr. Sanjay Tyagi, Dissertation Co-Advisor   Date
Associate Professor, Department of Medicine,
Public Health Research Institute, UMDNJ – New Jersey Medical School
Graduate Faculty, Department of Biological Sciences, NJIT

Dr. Farzan Nadim, Dissertation Co-Advisor   Date
Professor, Department of Biological Sciences, NJIT

Dr. Edward M. Bonder, Committee Member   Date
Associate Professor and Chair, Department of Biological Sciences, Rutgers
Chair and Graduate Program Director, Federated Activities

Dr. Alexis J. Rodriguez, Committee Member   Date
Assistant Professor, Department of Biological Sciences, Rutgers

Dr. Arjun Raj, Committee Member   Date
Assistant Professor, Department of Bioengineering, University of Pennsylvania

BIOGRAPHICAL SKETCH

Author: Khyati Shah

Degree: Doctor of Philosophy

Date: January 2012

Undergraduate and Graduate Education:

- Doctor of Philosophy in Cell and Molecular Biology
  New Jersey Institute of Technology, Newark, New Jersey, January 2012

- Master of Science in Biochemistry
  South Indian Education Society College of Arts, Science and Commerce Mumbai, India, May 2005

- Bachelor of Science in Microbiology and Biochemistry,
  South Indian Education Society College of Arts, Science and Commerce, Mumbai, India, May 2003

Major: Cell and Molecular Biology

Presentations and Publications:

D. Y. Vargas, K. Shah, M. Batish, M. Levandoski, S. Sinha, S. A. Marras, P. Schedl, S.Tyagi, Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147**, 1054 (Nov 23, 2011).

8th Cold Spring Harbor Laboratory Meeting on Eukaryotic mRNA Processing (August 2011).
Title: Alternative Splicing that Skips Exon-3 of *Drosophila* Sex-Lethal Pre-mRNA Occurs Post-Transcriptionally.
Author: Khyati Shah, Sanjay Tyagi

3rd Annual UMDNJ Inter-School Technology Symposium on Advances in Imaging & Genomics – Innovative Research Technologies to Bridge Bench to Bedside (April 2011)
Title: Alternative Splicing in Sxl Pre-mRNA Occurs Post-Transcriptionally.
Author: Khyati Shah, Sanjay Tyagi

I dedicate my dissertation to my parents, Mr. Girish S. Shah and Mrs. Chhaya G. Shah, and to my sister, Mrs. Toral T. Shah, who encouraged and supported me to come to the United States to pursue my doctoral studies, and to my husband, Dhaval, who was my greatest strength throughout graduate school, and who showed unconditional love and trust in me. I also dedicate this work to my advisor, Dr. Sanjay Tyagi, for developing my interest in the field of science and research.

ACKNOWLEDGEMENT

I would never have been able to finish my dissertation without the guidance of my advisor, Dr. Sanjay Tyagi, and the support of my committee members, my friends, and my family.

I would like to express my deepest gratitude to Dr. Tyagi, for his excellent guidance, caring, patience, and providing me with an excellent friendly atmosphere for doing research. Most importantly, I am grateful that he gave me the opportunity to develop my own individuality and self-sufficiency, by encouraging me to work independently. I will always remember his words "you can do it," which provided inspiration at times when things did not fall in place. Just a thank you is not enough for all the things he has done for me.

I also wish to thank my dissertation committee members, Drs. Edward Bonder, Farzan Nadim, Alexis Rodriguez, and Arjun Raj, for their guidance and helpful suggestions during each of my committee meetings.

I am grateful to members of the Kramer-Tyagi-Marras laboratory, past and present: Dr. Fred Russell Kramer, for helping me become a good presenter, reading this dissertation, providing many valuable comments that improved the presentation and contents of this dissertation, and helping me build a solid and effective resume – I am also thankful to him for his encouragement and practical advice, commenting on my views and helping me understand and enrich my ideas; Dr. Salvatore A.E. Marras, "our lab's lifeline," for technically helping me in everything, whether it was related to experiments, computer software, or just any problem one can come across in a lab, he has

an answer for almost everything – most importantly, he taught me how to label and purify fluorescent probes that I have used for most of my experiments, and how to make beautiful Illustrator figures – thank you Salvatore for all your help; Diana Vargas (Dr. Vargas, as she likes to be called ☺), for being my teacher, friend, and most importantly, my elder sister. She was the person who taught me many techniques, from how to load a gel, to how to run a PCR, carry out cloning, and how to use the fluorescence microscope – I think practically everything – thank you for making me what I am today; Susan Rae for making amazing cakes and cookies for everyone's birthdays; and Dr. Patrick van den Bogaard, Dr. Hiyam El-Hajj, Wei Yang, Mona Batish, and Michael Levandoski, for their support and help. I am specially going to miss my lunch and coffee hour talks with all the members of my laboratory.

I also thank the faculty of the Rutgers-NJIT Biological Science Department for teaching excellent courses that helped me develop more insight into the field of molecular biology. I especially wish to thank Dr. G. Miller Jonakait for her helpful discussions during my qualifying examination, and for proofreading my proposals. And I am grateful to the administrative staffs of the Federated Department of Biological Science and the Public Health Research Institute for assisting me with all of the administrative tasks that were necessary for completing my doctoral program: in particular, Karen Roach, Shandell Rivera, and Alla Rabinovich.

I thank my friends at Rutgers, NK, Kavya, Anand, Darshan, Neerja, and Pradeepa, for sharing their enthusiasm for my work and for providing helpful comments. In addition to my friends at school, this acknowledgement would be incomplete without mentioning my invaluable network of supportive, forgiving, generous, and loving friends,

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF SYMBOLS

ERK   Extracellular signal regulated kinase

MAPK   Mitogen-activated protein kinase

JNK   Jun N-terminal kinase

RTK   Receptor tyrosine kinase

CHX   Cyclohexamide

AMD   Actinomycin-D

FISH   Fluorescence *in situ* hybridization

PLA   Proximity ligation assay

CHO   Chinese hamster ovaries

snRNPs   Small nuclear ribonucleoprotein

CHAPTER 1

INTRODUCTION

Gene expression is a process by which information encoded within the sequence of a gene is used for the synthesis of a functional gene product. Most often, the gene product is a protein. Transcription, mRNA processing, translation, and post-translational modification are the fundamental steps involved in the synthesis of a protein. These processes need to be coordinated and regulated appropriately for a cell to function normally, to establish its polarity during early development, and to maintain its phenotype. It has been found recently, through several studies in bacteria, yeast, and higher eukaryotes, that within a genetically identical population, there is great variation in gene expression from cell to cell (*1-7*).

Variations in gene expression have been studied in detail using both reporter genes and natural genes at the single-cell level, and have been found to arise as the result of randomly initiated bursts of mRNA synthesis (*4, 8, 9*). Random bursts of mRNA synthesis, and the short half-life of transcripts, gives rise to variations in the amounts of the encoded proteins, causing phenotypic variation (*10*). Given this cell-to-cell variation, questions arise as to how coordinately expressed genes function in individual cells? Do they show any correlation in their expression relative to each other? Furthermore, if two proteins are involved in the formation of a heterodimer, how does variation in the amounts of the individual protein components of the heterodimer affect its assembly and function? Lastly, in a gene regulatory network, if there is variation in the expression of upstream genes, will this variation propagate and become amplified in the expression of downstream genes?

In the current study, these questions are addressed, utilizing a gene regulatory pathway that involves the immediate early response genes c-Fos and c-Jun, and delayed response genes (collagenase and cyclooxygenase-2) that they control. These studies were carried out with the aid of single-molecule fluorescence *in situ* hybridization technique (to detect individual mRNA molecules), and with the aid of a proximity ligation assay (to directly visualize individual protein heterodimers).

In addition to transcription, pre-mRNA splicing is another step that contributes to variations in gene expression in eukaryotes (*11*). Apart from co-transcriptional splicing, which occurs on the chromosome as a transcript is synthesized, alternative splicing can create different proteins from the same pre-mRNA by varying the exon composition of the spliced mRNA product. Alternative splicing contributes to the generation of complex proteomes. Various microarray data estimate that 73% of human genes are alternatively spliced, making alternative splicing the rule, rather than the exception. Thus, alternative splicing is a fundamental aspect of post-transcriptional gene regulation that has significant functional and biological implications (*12, 13*).

Where in the nucleus, and at what stage during mRNA synthesis, does splicing take place? It is currently believed that introns are spliced out from pre-mRNAs during transcription while the pre-mRNA is still tethered by RNA polymerase to the gene locus. This is an efficient way for constitutively spliced exons to be joined to each other in sequential order. However, when alternative splicing occurs, splicing must be delayed until all of the splice sites involved in the alternative choice have been synthesized. A key question concerning alternative splicing, is whether splicing is simply delayed until the alternative splice sites are transcribed, or does the alternative splicing

mechanism require the uncoupling of splicing from transcription, so that alternative splicing occurs after transcription is completed?

 In the current study, the question of whether splicing is coupled to, or uncoupled from transcription was addressed using a single-molecule *in situ* hybridization technique. The intracellular distribution and dynamics of individual molecules of pre-mRNAs and their spliced products were imaged, utilizing a set of reporter genes, and also utilizing the classically well-studied alternatively spliced gene: Sex-lethal (Sxl) in *Drosophila*.

CHAPTER 2

CONSEQUENCES OF STOCHASTIC mRNA SYNTHESIS
IN A GENE REGULATORY PATHWAY

2.1   Introduction

As cells divide and produce tissues, different cells within the same tissue have to perform the same set of tasks. To perform these tasks well it is expected that the expression of a given gene in different cells would be maintained at similar levels. Similarly, as bacteria grow in a liquid culture they are expected to maintain similar expression levels. However, despite of identical genotypes and similar phenotypes, great variations in the expression of the same genes from cell to cell have been observed (*6, 7*).

2.1.1 Origin of Gene Expression Variation in Prokaryotes

In order to understand the origin of these variations in gene expression from cell to cell, we have to first understand the initial steps involved in gene expression – the synthesis pre-mRNAs. For expression to occur, several factors are assembled at the promoter region of the gene. Some of these factors, such as RNA polymerase and sigma factors, are global, i.e., they operate on all or most of the genes. While others, such as transcriptional activators, are gene-specific or operate on only a few genes. Both of these kinds of factors can be present in very low-copy number in individual cells (*14, 15*). Furthermore, most of the genes are present in a single copy in bacterial genomes. Therefore, it is less than certain that in a given cell the relevant factors will be able to assemble at the gene during any particular short time interval. During this time interval, one cell may experience the productive assembly of these factors and produce mRNAs,

while the other cells may not, purely based on probability. Thus, different cells will exhibit variation in mRNA synthesis over time (*16*).

Early studies of cell-to-cell variation in gene expression were carried out in *Escherichia coli*, using fluorescent protein reporters. These studies documented substantial cell-to-cell variation in the amount of proteins being produced in cell populations that were genetically homogeneous. This variation was referred to as "noise" in gene expression. When this noise was quantified, it was found that it had two components. The first component, referred to as "extrinsic" variation, came from variations due to global factors, such as the shape and size of the cell, the particular stage of the cell cycle or fluctuations in the amount of global factors, such as the abundance of RNA polymerase. The second component, referred to as "intrinsic" variation, came from fluctuations in the expression of the gene itself. Intrinsic variation is likely due to a low number of gene-specific transcriptional activators (*6*). By performing time-lapse measurements to determine the time scale over which such fluctuation persist in bacteria, it was shown that the time scale for intrinsic fluctuations was less then nine minutes, whereas the time scale of extrinsic fluctuations was about 40 minutes, the later corresponds to the doubling time of bacteria (*17*).

These studies, along with several other studies (*3, 18, 19*), relied upon reporter constructs to obtain a picture of cell-to-cell variation. In order to explore how the expression of particular genes vary, Taniguchi et al. counted the individual mRNA and protein molecules present in single *E. coli* cells for more than a thousand genes. They found that, at any given time, there is cell-to-cell variation in gene expression of all the genes in *E. coli* cells (*20*).

2.1.2 Quantification of Variation

In order to understand this variation, several different mathematical models were generated (*21*). The initial model suggested that mRNAs are produced and degraded according to the statistics of Poisson distribution. This mean that the synthesis and degradation of mRNAs might occur at random but the probability of a transcript produced within any given time is constant and does not change in time. Hence according to the Poisson model, as the mean mRNA number increases, the variability about that mean should decrease. However, if the variability in gene expression is found to be much higher then predicted by the Poisson model, the mRNA synthesis occurs in the form of random bursts. Hence to experimentally distinguish between transcriptional bursts from poissonian transcription, one needs to measure the mRNA number per cell. Mathematically, one can calculate a fano factor for each distribution. Fano factor is defined as the variance of a distribution divided by the mean and is exactly one for a poisson distribution and much larger than one for transcriptional bursts (*7, 22, 23*).

2.1.3 Origin of Gene Expression Variation in Eukaryotes

After these studies with bacteria, researchers began to investigate whether similar cell-to-cell variations in gene expression occur in eukaryotes, and they found that variation in gene expression in these cells was even higher (*2, 8*).

Initially studies were performed in yeast, using a pair of fluorescent reporters and fluorescence activated cell sorting (FACS). This allowed precise measurement of protein levels at single-cell resolution. The striking result of these studies was the observation that noise in protein expression was due to the random synthesis of mRNAs, and not due to the presence of low numbers of global factors (*10, 24*). Even in the presence of a

saturating amount of transcription factors that turn on the gene, gene expression turned on and off in a stochastic manner.

The next set of studies examined gene expression variation in higher eukaryotes. Chubb et al. (*4*) studied the dynamics of mRNA synthesis from an engineered gene locus in *Dictyostelium discoideum*. They used an MS2-GFP fusion to visualize mRNA synthesis. They found that this gene is not expressed in a continuous and steady manner, but rather, it is expressed in a pulsatile manner. The pulses of expression began and ended randomly in individual cells, and the time interval between each pulse was irregular. This study was the first to provide direct, *in vivo* visualization of pulses in transcriptional activity (*4*). Similar observations were made in prokaryotes using the MS2-GFP approach (*5*). Raj et al. (*8*) studied the variation in fixed cultured mammalian cells by fluorescence *in situ* hybridization. They counted the number of mRNA molecules synthesized from reporter genes in individual cells, and they similarly observed large cell-to-cell variation in the expression of these genes.

With their larger cell size, longer cell division time, and higher number of polymerase and transcriptional activator molecules, one would have expected the eukaryotes to display a lower variation in gene expression, as compared to prokaryotes. To explain eukaryotic gene expression variation, a two-stage gene activation model has been proposed (*25*). The physical basis of this two-stage gene activation model resides in the structure of chromatin. In higher eukaryotes, genes are sequestered in a tight chromatin structure. In its quiescent state, the chromatin surrounding the gene is so compact that activator proteins cannot gain access to the promoter region of the gene, even when they are present in high numbers. It was proposed that random breathing

events in the chromatin allow the activator proteins to bind to the promoter region of the gene. The initial binding of these activators leads to the recruitment of chromatin remodeling factors that unfold the chromatin further, enabling the recruitment of the RNA polymerase machinery. Hence, according to this model, randomness in access to the promoter region is the basis of stochastic gene expression (*8*).

2.1.4 Intrinsic and Extrinsic Variation in Gene Expression

In order to determine whether sources extrinsic or intrinsic to the gene are the cause of the variation, Raj et al. (*8*) integrated two different reporter genes, either at the same genomic location, or at different genomic locations. These reporters could be turned on by the same transactivator protein (abundantly present in the cell). They found that the expression of two reporters present at the same genomic location was correlated, but when present at distant genomic loci, expression of the two genes was not correlated. If the transactivator, which is a global factor for this pair of reporter genes, was the main cause of variation, one would expect the expression of these genes to be correlated irrespective of their locations within the genome. The observation that reporter genes located at different genomic loci are not correlated suggests that the main component of variation is intrinsic to the gene locus (*8*). Similar conclusions were reached in studies of gene expression with yeast, which indicate that intrinsic variation dominates extrinsic variation in eukaryotes (*2, 10, 24*).

The dominance of intrinsic variation in eukaryotes contrasts with the dominance of extrinsic variation in prokaryotes. Since most prokaryotic genes are present in single copy, and since there is no chromatin in prokaryotes, the accessibility of the transcription factors is not a significant factor, though their low copy number is. Furthermore, in

prokaryotes cell size is small, and their doubling time is short, further enhancing extrinsic variation.

2.1.5 Variation as a Consequence of Stochastic Synthesis and the Steady Decay of mRNA

The variation in gene expression is thought to arise from random activation and deactivation of eukaryotic genes. Several studies indeed observed bursts of mRNA synthesis corresponding to the "on" state of a gene, followed by a much longer period during which the gene is inactive. Golding et al. observed bursts of mRNA synthesis in *E. coli* (*5*), and using additional techniques, transcriptional bursts were shown to be the prominent mode of gene expression in higher eukaryotes (*4, 8, 26-29*). Although the occurrence of transcriptional bursts was random, the average length of these bursts was about nine minutes, and on average, about 250 mRNA molecules were made from a gene during each burst. After being synthesized in transcriptional bursts, mRNAs decay with steady first order kinetics. Since the half-life of different mRNAs varies from several minutes to hours, snapshots of cells by *in situ* hybridization techniques show only the mRNAs produced in recent bursts (*8*). Furthermore, the kinetics of the transcriptional bursts is gene specific and the presence of multiple *cis*-acting regulatory elements, and/or the presence of a large number of transactivator proteins increases the average number of transcripts made during each burst, but does not affect the duration of the "active" state of the gene (*26*).

2.1.6 Propagation of Variation in mRNA Expression into Variation in Protein Levels

As compared to mRNAs, the half-life of proteins varies from a couple of hours to a day or more (*30, 31*). Since proteins have longer half-life, proteins made from a new burst of mRNA synthesis are added to the pool of proteins made from previous bursts of mRNA synthesis. Hence, proteins show less variation than their parent mRNAs. Furthermore, noise in protein expression should inversely correlate with its half-life (*8, 10, 32, 33*). Due to the short doubling times of bacteria and the long half-life of proteins in *E. coli*, the number of mRNAs and their corresponding proteins do not correlate (*20*).

2.1.7 Variability in Gene Expression of Natural Genes

Studies discussed so far utilized reporter gene constructs. Several groups have studied the variable expression of natural genes in diverse biological contexts. A study performed in *Saccharomyces cerevisiae* revealed that some genes have very high variability, whereas other genes are expressed at relatively uniform levels. These studies analyzed gene expression of constitutively active housekeeping genes and inducible regulatory genes, using single-molecule fluorescence *in situ* hybridization. They found that the variation in the expression of constitutively active genes is very small, and is due to irregular single transcription-initiation events, as compared to larger variation in the expression of regulated genes, characterized by transcriptional bursts (*34*). The same group in another study found that the expression of functionally related constitutively active genes is not coordinated in individual cells (*35*).

Two distinct modes of gene expression were also found in *Drosophila* embryos, using quantitative *in situ* hybridization. The expression of 14 developmental genes was

visualized in early embryos. Some of these genes contained RNA polymerase II bound to their promoter regions before their induction, and the former group showed uniform expression in all cells of an embryonic tissue. Other genes that did not contain RNA polymerase II pre-loaded on their promoter regions, showed asynchronous, stochastic expression (*9*).

The effect of stochastic gene expression on phenotypic variation in the nematode, *Caenorhabditis elegans*, was studied in a gene regulatory network for intestinal development. By comparing transcripts of the genes in this network in individual wild type or mutant embryos, it was shown that the expression of a redundant gene from this network becomes highly variable in mutant strains. Because of this variation, this gene fails to reach a threshold of expression that is required for the expression of its downstream master regulatory gene in certain mutant causing phenotypic variation. Hence, redundant pathways have evolved to mitigate the effect of gene expression in one pathway (*36*).

Furthermore, to study the effect of tissue development on gene expression, Featherstone et al. used bioluminescence imaging to study the expression of the prolactin gene of the pituitary gland in a transgenic rat. They observed that the transcription pattern changed during tissue development. During early development of fetal tissues, prolactin gene expression was pulsatile, which later becomes more continuous and stable in neonatal tissue (*37*).

2.1.8 Questions Raised by High Cell-to-Cell Variation

Cells express a number of genes simultaneously in response to extracellular stimuli for a short window of time. For example, in response to serum stimulation, approximately 100 genes are expressed simultaneously within a few minutes (*38, 39*). The coordinated expression of these genes is necessary for the overall response of the cell. Given the knowledge of variation in gene expression from cell to cell, one of the questions raised is that if the level of expression of any given gene is different from cell to cell, how are individual cells able to produce a coherent response?

Another question raised by high cell-to-cell variation in gene expression is how the assembly of multi-subunit proteins is accomplished, and how variation in the levels of component proteins affects variations in the levels of the composite proteins. For example, if two individual proteins need to form a heterodimer to carry out a particular function, how does variation in the amounts of the individual proteins affect their assembly and function?

A third question is whether variation propagates and is amplified in a gene regulatory pathway. For example, the genetic program of a living cell is determined by a complex set of gene regulatory networks. The effective functioning of these networks relies on faithful signal propagation from one gene to another. In a gene network, the protein product of an upstream gene, or a pair of upstream genes, is required to induce a downstream gene. Although the large scale cell-to-cell variations observed at the mRNA level are buffered at the protein level, nonetheless there exists considerable variation in the levels of different proteins in cells (*10*). When such a protein is a part of a gene

regulatory network, does variation in the amount of this protein amplify the expression of downstream genes?

In the current study, these questions were addressed, utilizing c-Fos and c-Jun and the delayed response genes that they control as models. c-Fos and c-Jun are a pair of genes induced in a coordinated manner within 15-30 minutes of the addition of serum in quiescent cells. The expression of their mRNAs returns to the basal level after about one to two hours. However, their protein products are more stable, and work together by forming heterodimers that function as transcription factors. These heterodimers induce the expression of several downstream genes such as collagenase, cox-2, cyclin-D1, and IL-1β. The variation in the expression of c-Fos and c-Jun mRNAs, variation in the number of heterodimers, and variation in the expression of a pair of downstream genes (collagenase and cox-2) were studied (Figure 2.1).
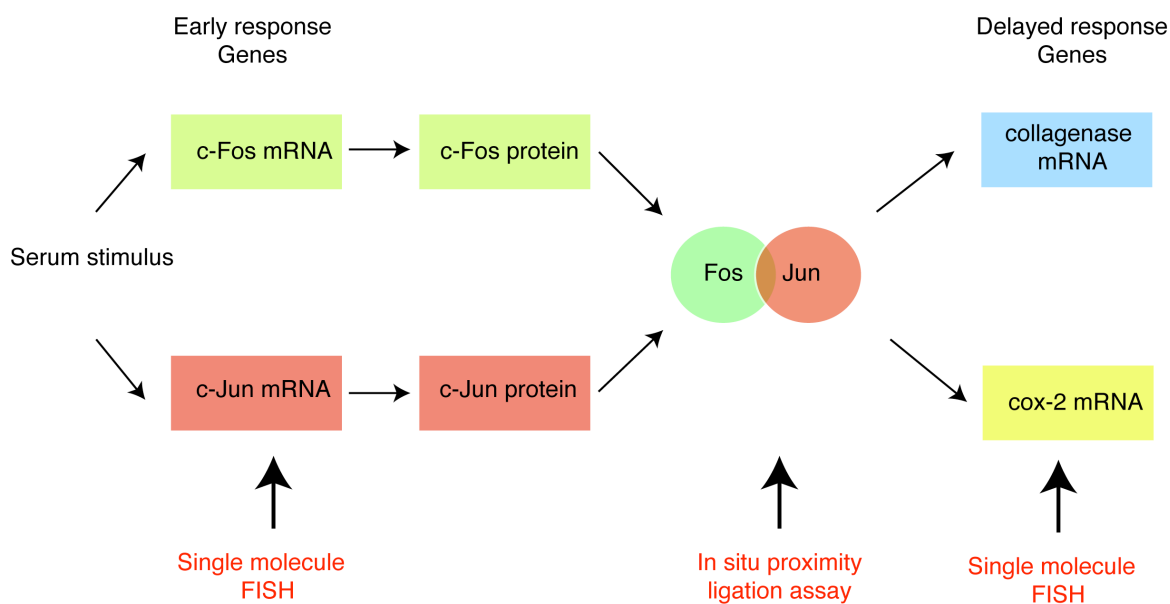
Figure 2.1 Experimental outline to address the proposed questions.

2.1.9 Immediate Early and Delayed Secondary Response Genes in Cell Proliferation and Differentiation

Cell proliferation and differentiation occurs when an extracellular signaling molecule activates a cellular receptor. This interaction leads to series of biochemical changes within the cytoplasm. Afterwards, the activation signals cross the nuclear membrane and alter the expression of genes encoding proteins that dictate tissue or stimulus-specific functional responses. Some of the genes are induced soon after the response and are referred to as "immediate early genes," whereas others are induced only after the expression of the first set of genes, and they are referred to as "delayed secondary response genes" (*38*).

The initial transcriptional response to growth factor stimulation leads to induction of approximately 100 immediate early response genes. Their expression reaches a maximum within 30 minutes of growth factor stimulation, and returns to pre-stimulated levels within 60 to 120 minutes. The expression of these genes does not require *de novo* protein synthesis, and they are over expressed in the presence of protein synthesis inhibitors. Some of these early response genes perform structural functions within the cell (such as actin and tropomyosin). Another important subset of early response genes encodes transcription factors. These genes propagate their activation signals downstream by inducing the expression of delayed secondary response gene (*38-40*). Figure 2.2 shows the time course of early and delayed response genes in G_0-arrested mammalian cells after the addition of serum. Components of serum that are important for this response are growth factors.
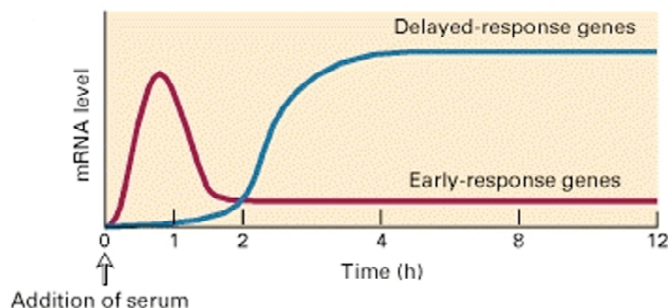
Figure 2.2 The time course of early and delayed response genes in G_0-arrested mammalian cells after the addition of serum containing growth factors. Source: (*41*)

Among the most well studied early response transcription factors are the Fos and Jun family of transcription factors. The Fos family includes four genes (c-fos, fosB, fra-1, and fra-2) and the Jun family includes three genes (c-jun, junB, and junD). These genes are induced in various tissues and cell lines in response to various stimuli, such as cytokines, growth factors, serum, UV irradiation, stress, etc., through Ras-mitogen-activated protein kinase (Ras-MAPK) signaling pathways. The protein products of these genes are required for cell-cycle progression of serum-stimulated or asynchronously growing cells (*42*). Their expression is regulated by both transcriptional and post-transcriptional mechanisms. Alteration in their expression by mutation or deregulation leads to tumorigenesis, hence they are also referred to as proto-oncogenes (*40, 43-45*).

The proteins encoded by the Fos and Jun family members function by forming homodimers (in the case of the Jun protein family alone) and heterodimers (in the case of both the Fos and Jun protein families). They bind to the heptamer consensus sequence, 5'-TGA(C/G)TCA-3', which is known as the "TPA response element" in their target promoters, and these proteins are generally referred to as "activator protein-1 (AP-1) transcription factors." They belong to a class of basic leucine zipper transcription factors

and function in the transcription of several secondary response genes involved in cell proliferation, differentiation, inflammation, and apoptosis, in a cell-type and tissue-specific manner. These genes require the protein product of the primary transcripts in order to express themselves, and consequently, in the presences of a protein synthesis inhibitor, their transcription is inhibited (*44, 46-48*). Various Jun-Fos dimers, in spite of containing similar DNA binding sites, differ in their transcriptional activity, due to regulated phosphorylation at specific sites of non-conserved domains located outside the leucine zipper domain. Thus, AP-1 dimers of different composition execute different cellular functions by inducing dimer-specific target genes. The targets of c-Fos and c-Jun heterodimers include two delayed response genes, collagenase and Cox-2 (*49*).

2.1.10 c-Fos-c-Jun Heterodimers

Three different types of MAPKs, the ERKs, JNKs, and FRKs induce expression of Fos and Jun proto-oncogenes in response to growth factor stimuli (*50, 51*). Among them, the expression of c-Fos is induced by phosphorylation of ERKs, whereas the expression of c-Jun is induced by phosphorylation of JNKs, both occurring concurrently (*52*). Several northern blot, real-time PCR, western blot, and immunoprecipitation analyses have shown that c-Fos and c-Jun mRNAs and proteins are produced in a coordinated manner in various cell lines on a population basis (*53-55*).

Although, the half-life of individual c-Fos and c-Jun proteins is approximately 45 minutes and 90 minutes, respectively, several *in vitro* and *in vivo* association studies have shown that once they form heterodimers, the heterodimers become more stable, with half-lives of about four hours. This is because they remain highly phosphorylated in their heterodimeric state. In their heterodimer form, c-Jun binds to the AP-1 target region, and

c-Fos provides a transactivation function (*56*). Furthermore, the deletion of certain regions of c-Fos protein, or inhibition of its synthesis by antisense RNA, prevents the induction of downstream target genes, suggesting that their association is not only required, but it is essential for them to function (*47, 48, 57-59*).

2.1.11 Induction of Collagenase and Cyclooxygenase-2 (Cox-2) Genes by c-Fos-c-Jun Heterodimers

AP-1 transcription factors induce the expression of several genes that are involved in cell cycle progression and proliferation (*57*). Among the genes that are known to be specifically induced by c-Fos-c-Jun heterodimers are Cyclin-D1, IL1β, collagenase, and cox-2 (*60-62*). The collagenase and Cox-2 genes were used as downstream targets of c-Fos-Jun heterodimers. Collagenase is a member of the metalloproteases family, and its expression is elevated in certain tumor cells, whereas cox-2 is an enzyme required for the synthesis of prostaglandins from arachidonic acid, and it is responsible for the production of elevated levels of prostaglandins during inflammation and carcinogenesis.

There is strong evidence that c-Fos-Jun heterodimers directly control the expression of collagenase and Cox-2 (*55, 62-65*). First, c-Fos-Jun heterodimers bind to the collagenase promoter and lead to its expression (*55*). Second, in a pair of studies in which c-Fos and c-Jun were mutagenized, it was shown that these mutations lead to the abrogation of both collagenase and Cox-2 expression (*58, 61, 62, 64*). Finally, when c-Fos and c-Jun proteins were fused into one protein, the fused protein was able to bind to the collagenase promoter and induce the expression of mRNA from the gene, whereas, fusions between other members of the Fos and Jun family of proteins did not yield any expression (*49*).

2.1.12 Single-Molecule Imaging for mRNA Detection and Quantification

Single cell analysis of gene expression has shown great variability in gene expression from cell to cell with significant biological consequences that were not observed in the population-based analysis. A reliable and sensitive method to count individual mRNA molecules to determine actual count of mRNA copy numbers in individual cell and their localization is needed (*66*). One such method is *in situ* hybridization, where labeled nucleotide probes bind to their complimentary sequences in fixed cells and renders their detection. Initially, the probes were labeled either with radioisotopes or linked to an enzymatic reaction for their detection (*67, 68*). Unfortunately, these reactions generated molecules that diffused away from the probe itself making it difficult to determine the spatial location of the target with poor sensitivity. Alternatively, one can label probes directly with a fluorophore to achieve spatial information but this approach showed poor sensitivity as individual probes give rise to low and diffused fluorescence signal.

To over come these problems, a modification of fluorescent in situ hybridization technique was developed by Robert Singer and colleagues (*69*). Their approach was to have five, 50 nucleotide long probes, each labeled with five fluorophores hybridize to each mRNA target. This approach was sensitive enough to image single mRNA molecule. Further, by using probe sets labeled with different fluorophores, different mRNA targets can be imaged simultaneously. However, this method has two shortcomings: 1. The efficient synthesis and purification of multiple labeled probes is expensive and cumbersome and 2. The multiple dye molecules have the potential to quench each other due to close proximity, hence the signals generated using these probes were subject to more variability (*69*).

Raj et al. (*66*), elaborated the existing single-molecule detection method to over come both the sensitivity and specificity issue. The basic idea was to use 48 probes, each 20-nucleotides long, and each labeled singly at their 3' end. When 48 singly labeled probes bind to a target mRNA, sufficient fluorescence is generated for the target mRNA to be visualized as a single diffraction-limited spot under a fluorescence microscope. Since each probe is singly labeled, the fluorophores on two adjacent bound probes are at a sufficient distance from the each other that quenching does not occur (*66*). This method has proven to be particularly reliable in yielding single-molecule sensitivity and has successfully been used in a diversity of biological contexts (*36, 70, 71*).

The extremely high specificity of our system arises from the fact that when all or most of the probes bind simultaneously to the same mRNA molecule a spot-like signal is generated, whereas, the binding of one or a few probes to non-specific sites only generates a diffused signal. Image processing algorithms designed to detect diffraction-limited spots, and to neglect the diffused signals, can thus be used with high confidence. The high specificity that is achieved has been demonstrated in a number of different ways.

When specific probes were used for mRNAs coding for an artificial, inducible gene, the cells yielded spots only when such mRNAs were expressed (*66, 72*). In the case of endogenous mRNAs localized in specific subcellular zones, signals were detected only in the appropriate subcellular zones (*66, 70*). In the case of inducible genes, spots were detected only upon induction, and their numbers correlated with the extent of induction (*66*). Probe-based controls, such as antisense or irrelevant probe sets, do not generate any signals. Finally, signals from two or more distinctly labeled probe sets

co-localize when the sets are complementary to the same mRNA, but do not co-localize when the probe sets are complementary to different mRNAs (*66*).

Several different lines of evidence indicate that the spots in our method arise from single mRNA molecules. The numbers of spots per cell corresponds with the number of mRNA molecules per cell obtained by real-time PCR (*66, 70, 73*). The intensities of spots exhibit a unimodal distribution (*73*), and the magnitude of intensities scale with the number of probes used (*66, 73*). Two isoforms of alternatively spliced mRNAs can be separately detected using one set of probes for the common region and a pair of distinctly labeled probes for the alternatively chosen region (*74*). The most compelling evidence for single-molecule detection occurred in one of our studies, wherein before splicing, a larger number of intron spots co-localize with exon spots, than co-localize after splicing. It is conceivable that the spots arise from conglomerates of mRNA molecules, or from association with nuclear structures, but in those cases splicing would not have resulted in the segregation of spots (*75*). Finally, the intensity of spots is the sum of intensities of all dye molecules that are tethered to the mRNA molecules (*69*). In the current study, the expression of c-Fos and c-Jun mRNAs, and the expression of the downstream genes (collagenase and Cox-2) were studied with the aid of single-molecule fluorescence *in situ* hybridization.

2.1.13 Imaging and Quantification of Individual Protein-Protein Interactions

Most of the gene expression analysis is predominantly performed at the level of mRNAs. However, specific interactions between proteins to form multi-subunit complex or their post-translational modifications are the key requirement for proper execution of a gene regulatory pathway. Recently, it has been shown that the number of mRNAs and proteins

for any given gene does not correlate in individual cells (*20*). Hence, in order to study the propagation of noise in a gene regulatory pathway, we need to count the amount of individual proteins from cell-to-cell. Techniques most commonly used for detection of individual or multiple proteins are protein microarray, co-immuno precipitation, 2-D gels and mass spectrometry. These techniques are efficient and specific to detect protein-protein interactions. However, they provide information at population level and do not detect transient interactions. Furthermore, by performing immunofluorescence for two different proteins using different secondary antibodies, one can determine the amount of individual protein present from cell-to-cell but it does not provide information regarding their association.

Fluorescence resonance energy transfer (FRET) is another method to determine protein-protein interactions in vivo. In this technique, either the proteins of interest are directly fluorescently labeled or two different antibodies specific to individual proteins are labeled with cy3 and cy5 as the donor and acceptor fluorophores. It involves the nonradioactive transfer of energy from an excited state donor of the fluorophore (cy3) to a nearby acceptor (cy5). The efficiency of energy transfer is related to the distance separating a given donor and acceptor pair, which is usually 1-10 nm. When this distance is more then 20 nm, no FRET occurs (*76*). This is an efficient technique to visualize protein-protein interactions and it also provides spatial information of the proteins. However, this technique has certain limitations: 1. It requires either constructing reporter genes where in your protein of interests are fused to fluorescent proteins, or labeling of primary or secondary antibodies with fluorophores. 2. The signal to noise ratio generated is very low, giving rise to high background, and 3. If two proteins are at a distance more

then 20 nm due to their different conformation, but still in the same complex, FRET will not work efficiently. In addition, in the current project, to address the question of effect of stochastic mRNA synthesis in a gene regulatory pathway, two of the fluorescence channels will be used to image individual downstream genes mRNAs, and using another two FRET pairs of fluorophore to determine c-Fos-c-Jun heterodimers would further interfere with the mRNA detection.

Proximity ligation assay (PLA) is a technique capable of detecting single endogenous protein events, such as protein expression, dimerization, and modifications, such as protein phosphorylation in fixed culture cells or tissue sections. In this method, initially two different primary antibodies raised in different species specific to individual endogenous proteins are used. Thereafter, secondary antibodies containing a unique DNA probe are added. If these two DNA probes are in close proximity, they hybridize to a connector oligonucleotide, and a ligation reaction occurs, that is followed by rolling circle amplification (RCA) of the ligated sequence, generating 1000 of amplified copies of single stranded DNA, which is then detected with fluorescent probes targeted against the amplified sequence. Individual proteins that are not part of a complex are not able to elicit this reaction. Figure 2.3 shows the steps in PLA. PLA generates localized, distinct signal, which remain anchored to one of the proximity probes, thereby revealing the exact location of the proteins. Also, since 100-1000s of amplified copies of template DNA are made, the signals generated from multiple probes binding to the amplified DNA gives higher signal to noise ratio giving rise to very less or no background. By counting these localized signals, one can quantify and compare protein-protein interactions that occur in different cells and tissues, and that are the result of different treatments (*77*)
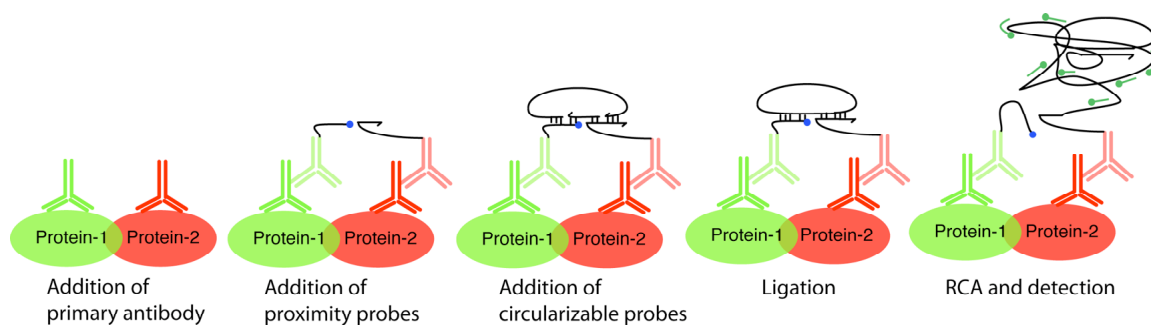
Figure 2.3 Steps Involved in detection of protein-protein interactions by proximity ligation assay.

A number of studies indicate that the signals generated by the performance of PLA are specific and quantitative. When PLA for inducible protein heterodimers was performed, signals were obtained only upon induction. Furthermore, when non-specific primary antibodies or interfering mutant proteins were transfected into the cells, no signals were generated (*78, 79*). PLA has been used to quantify the up-regulation of activated signaling proteins during the progression of various cancers and for drug screening in various tissues and cell lines (*78, 80-84*).

2.1.14 Experimental Outline to Address Proposed Questions

The questions raised in Section 2.1.8 will be addressed using a pathway in which c-Fos and c-Jun are expressed upon serum addition, resulting in the formation of heterodimers that turn on the expression of a pair of downstream genes, collagenase and cox-2 (Figure 2.1). This pathway provides a unique opportunity to study the questions raised. The first question, whether two mRNAs that are expressed at the population level in a coordinated manner are also expressed in a coordinated manner in individual cells, was addressed by counting the number of c-Fos and c-Jun mRNAs in HeLa cells as a function of time after

serum induction. With respect to the question of propagation of noise in the gene expression pathway, one would have counted the mRNAs for c-Fos, c-Jun, collagenase and cox-2, simultaneously; however, by the time collagenase and cox-2 are beginning to be expressed, the c-Fos and c-Jun mRNAs have disappeared. Therefore, the number of heterodimers formed by c-Fos and c-Jun proteins were counted with the aid of PLA, as well as the number of mRNAs of collagenase and cox-2 were counted with the aid of single-molecule FISH within the same cells. Finally, the use of this pathway also addressed the question of how variations in the number of mRNAs encoding the component proteins affect the variation of the multi-subunit proteins.

2.2   Results

2.2.1 Imaging Individual Molecules of c-Fos and c-Jun mRNAs

The expression of c-Fos and c-Jun in individual HeLa cells was studied using the single-molecule FISH procedure described above. A set of 48 probes each for c-Fos and c-Jun mRNA were synthesized and labeled with Alexa 594 and tetramethylrhodamine, respectively (Figure 2.4 A). HeLa cells were cultured on glass cover slips in the absence of serum for 48 hours, and 20 % serum containing 200 μM 12- O- tetradecanoylphorbol -13-acetate (TPA) was added to induce the expression of c-Fos and c-Jun mRNAs. The cells were fixed and *in situ* hybridization was performed using both sets of probes. 20 to 40 optical slices were acquired; each 0.2 μm apart, and they were imaged in each fluorescence channel with a one-second exposure. For each mRNA species, discrete spots corresponding to individual mRNAs were observed in each channel. These three-dimensional stacks of images were further merged into one composite image, and the

resulting image was color coded with green for c-Fos mRNAs and red for c-Jun mRNAs. These color-coded merged images were further merged into one RGB image, to create the composites shown in Figure 2.4 B.

In order to count the number of mRNA molecules for each species, the three-dimensional stacks of images for each mRNA species were analyzed using a custom image-processing program. This program identifies each spot in three dimensions, based on a user-provided threshold intensity, and counts the number of mRNA molecules in an area that corresponds to the cell boundary drawn by the user using a DIC image of the cell. The locations of the identified color-coded spots were plotted over the DIC image (Figure 2.4 B). Evidence supporting the sensitivity and specificity of mRNA detection by this method was presented previously (59, 62, 75).
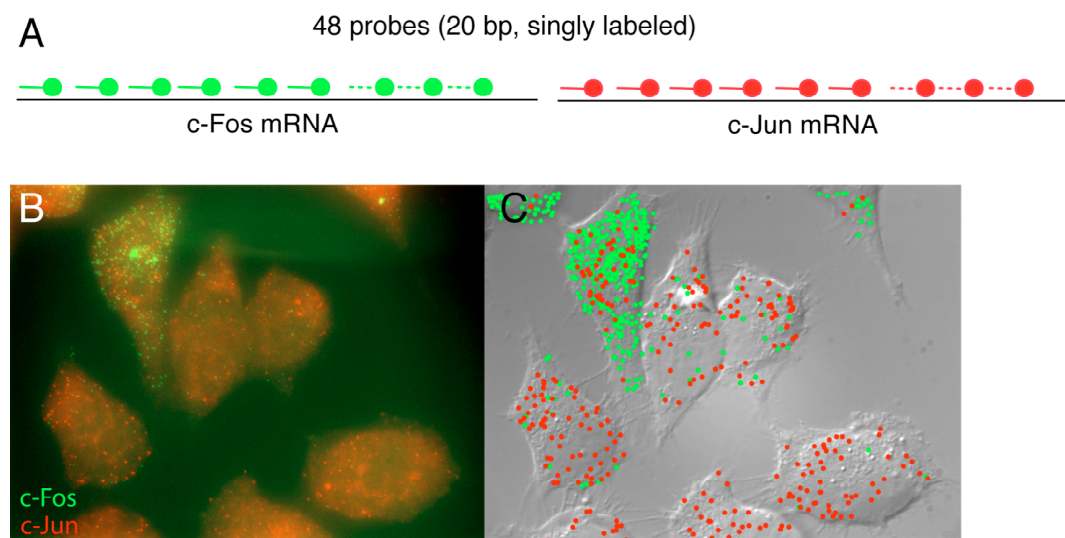
Figure 2.4 Single-molecule imaging of c-Fos and c-Jun mRNAs in individual cells. A. Scheme for imaging individual molecules of c-Fos and c-Jun mRNA molecules using labeled oligonucleotide probes. B. Three-dimensional stacks for individual colors were merged and compressed into one composite image of cells induced with 20 % serum for 30 minutes. Red represents c-Jun mRNAs, and green represents c-Fos mRNAs. C. Identification of mRNA species using our image-processing algorithm. Filled circles are drawn around each detected mRNA molecule.

2.2.2 c-Fos and c-Jun mRNAs are Expressed in Bursts and are Not Coordinated with Each Other in Individual Cells

The number of c-Fos and c-Jun mRNA molecules from 100 randomly selected cells was counted. The result of these measurements is presented in Figure 2.5 A. The expression level of each mRNAs was wide-ranging between individual cells. c-Fos expression varied from 0 to 700 mRNAs per cell with a mean of 218, where as c-Jun expression varied from 0 to 200 mRNAs per cell with a mean of 87.

In order to provide an understanding of the distribution of mRNAs in the population of cells, the data is presented as histograms in Figure 2.5 B. These histograms reveal extremely wide distributions of expression levels of each mRNAs in cell population. These distributions stem from a stochastic expression of mRNA as observed before for other genes (*8*). These distributions depart from Poisson distribution, which would be expected if the mRNAs were produced in a steady manner in each cell. The Poisson distribution is overlaid on the observed distributions (green dotted line) (Figure 2.5 B).
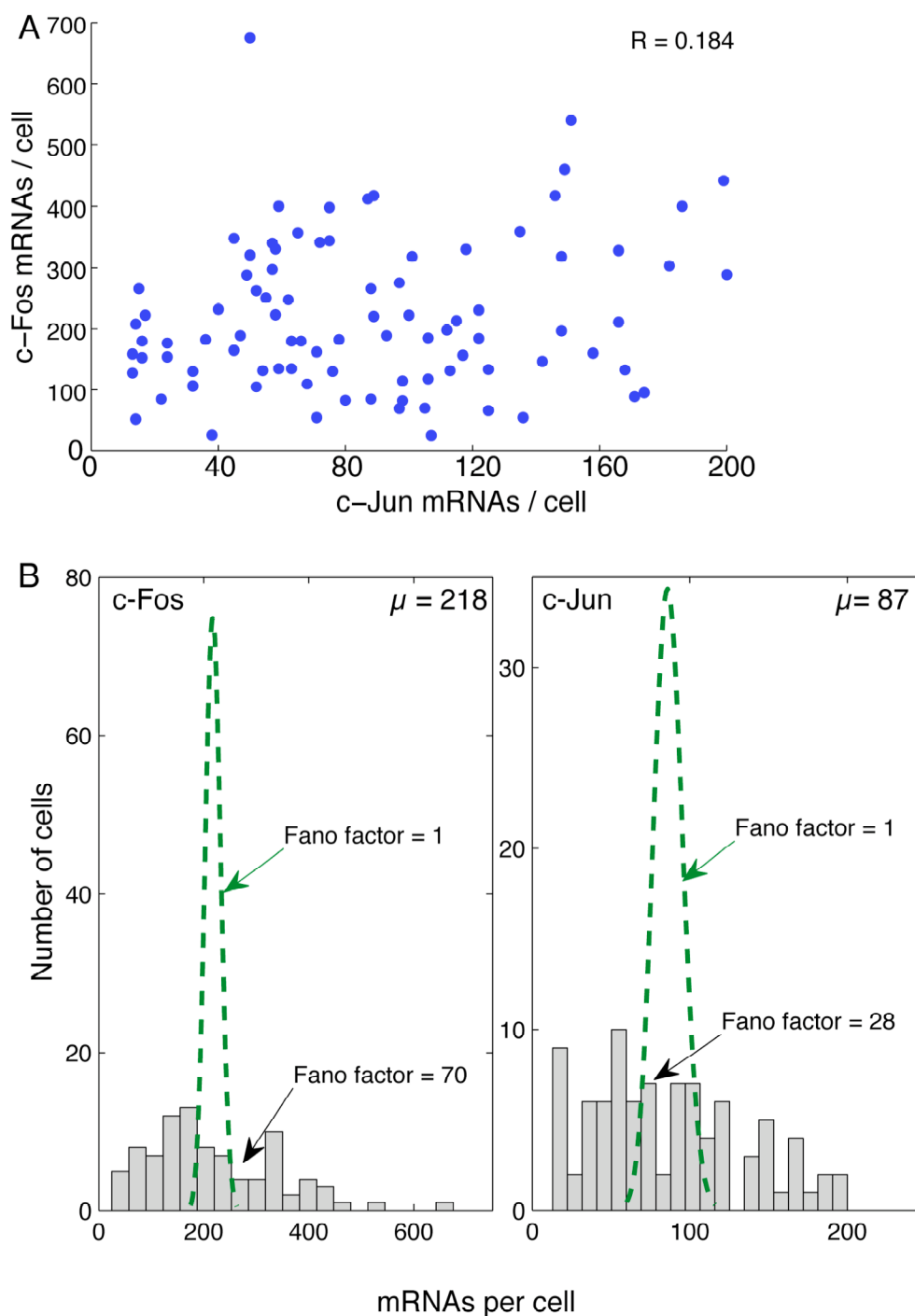
Figure 2.5 Expressions of c-Jun and c-Fos mRNAs occur through transcriptional bursts. A. Scatter plot of total c-Jun and c-Fos mRNAs in 100 individual cells after induction of cells for 30 minutes. B. Histograms showing observed distribution of c-Jun and c-Fos mRNA molecules per cell (grey bars) overlaid with their calculated Poisson distribution (green dotted line) and Fano factor values.

A measure of the departure from the Poisson distribution is obtained by calculating the Fano factor (*85*).

$$FanoFactor = \frac{Variance}{Mean} \qquad (2.1)$$

If mRNAs were produced with constant rate and followed a Poisson distribution, the mean number of mRNAs per cell will be equal to its variance, yielding a Fano factor of 1. Instead c-Fos and c-Jun yielded a Fano factor of 70 and 28 respectively. These large Fano factors signify that mRNAs are being produced in random bursts followed by steady decay over time (*8*). Most surprisingly, we found that numbers of c-Fos mRNA molecules were not correlated with the number of c-Jun mRNA molecules in the same cells. The correlation between these two measurements in the population was 0.184 (Figure 2.5 A). This indicates that the bursts of c-Fos mRNA synthesis are random in relation to the bursts of c-Jun mRNA synthesis. Although the results are consistent with previous results of Raj et al (*8*), it is extremely surprising given that these two genes need to be expressed in a coordinate manner in each cell as their proteins form heterodimers.

2.2.3 Expression of c-Fos and c-Jun mRNAs as a Function of Time

The expression of c-Fos and c-Jun mRNAs as a function of time after the addition of serum was studied in a similar manner. The mean values at each time-point for each mRNA are shown in Figure 2.6 A (upper panel), along with a measure of their dispersion (95 % confidence interval). Expression peaked at 30 minutes, and returned to pre-serum

levels 60 minutes after the addition of serum. This time course resembles what was observed in the population-based measurements (Figure 2.6 B).

In order to explore how c-Fos and c-Jun mRNA expressions are correlated in individual cells during the course of their expression, correlation coefficients (R-values) were calculated at each time point. These R-values are presented along with their 95 % confidence intervals in Figure 2.6 A (lower panel). At all the time-points in which two mRNAs were expressed, the correlation between them was poor, indicating the absence of coordinated expression in individual cells over time. Thus, although the expression of c-Fos and c-Jun is coordinated at the population level, it is uncorrelated in individual cells.
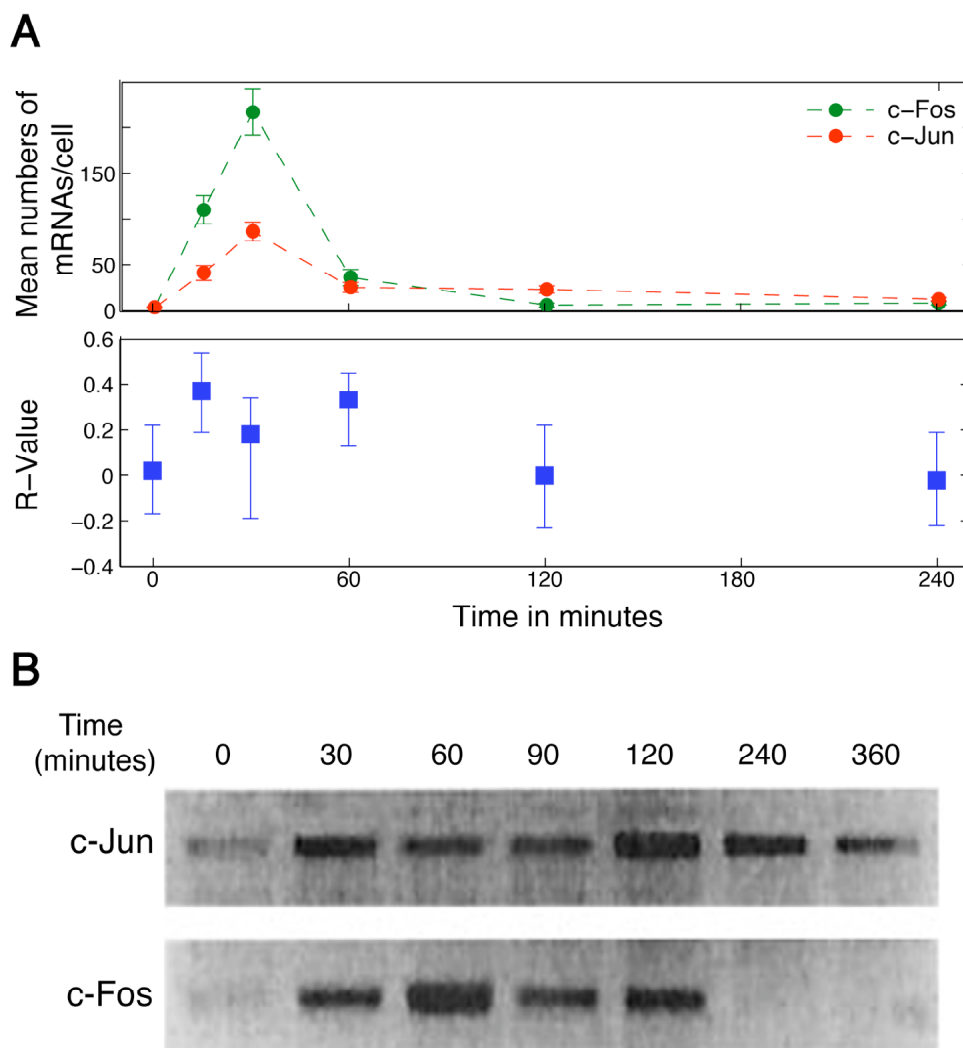
Figure 2.6 Expression of c-Jun and c-Fos mRNA is not correlated from cell to cell. A. Mean numbers of c-Jun and c-Fos mRNAs per cell at various times after induction with serum (top). Correlation coefficients (R-values) between c-Jun and c-Fos mRNAs at various time intervals (bottom). The error bar represents a 95 % confidence interval. Means and R-values were obtained after counting 100 cells in each category. B. Expression of c-Fos and c-Jun induction as assessed by RT-PCR at various times after treatment with serum (*55*).

2.2.4 Visualizing Heterodimers Formed by c-Fos and c-Jun Proteins in Individual Cells

The proteins encoded by c-Fos and c-Jun function by forming heterodimers, which

induce the expression of several secondary response genes involved in cell proliferation,

differentiation, inflammation, and apoptosis, in a cell-type and tissue-specific manner

(*57*). In order to study the propagation of noise from c-Fos and c-Jun mRNAs to their functional protein heterodimers, PLA was used to specifically detect c-Fos-c-Jun heterodimers with single-molecule resolution. As described in detail in the introduction, this method utilizes specific primary antibodies against c-Fos and c-Jun proteins. Secondary antibodies that mediate a rolling circle amplification reaction recognize these primary antibodies. Subsequent hybridization of complimentary oligonucleotide probes to the rolling circle amplification products generates strong fluorescence signals that remain localized at the site of each heterodimer.

In order to demonstrate that the resulting signals detect individual c-Fos-c-Jun heterodimers, HeLa cells were cultured in the absence of serum for 48-hours and were induced with serum for six hours. After fixing the cells, antibodies were added, rolling circle amplification was carried out, and signal detection was performed, as described in detail in the material and methods section below. About 100 bright fluorescence spots were observed in the nucleus of each cell (Figure 2.7 A). c-Fos and c-Jun proteins forms heterodimers in the cytoplasm. However, upon phosphorylation, the heterodimers migrate into the nucleus (*86*). Consistent with this, spots were infrequently found in the cytoplasm (Figures 2.7 B and C). In contrast to cells six hours after induction with serum, there were very few spots in cells that were not induced with serum (Figure 2.7 C). Further evidence of the specificity of detection emerged when we inhibited c-Fos expression with U0126 (a MEK 1/2 inhibitor) (*87*). In the presence of this inhibitor, no spots were detected, even after induction with serum. As a control to highlight the specificity of PLA, one of the primary antibodies or one of the secondary antibodies was

omitted from the mixture, which resulted in the complete absence of signals (Figure 2.7

C).
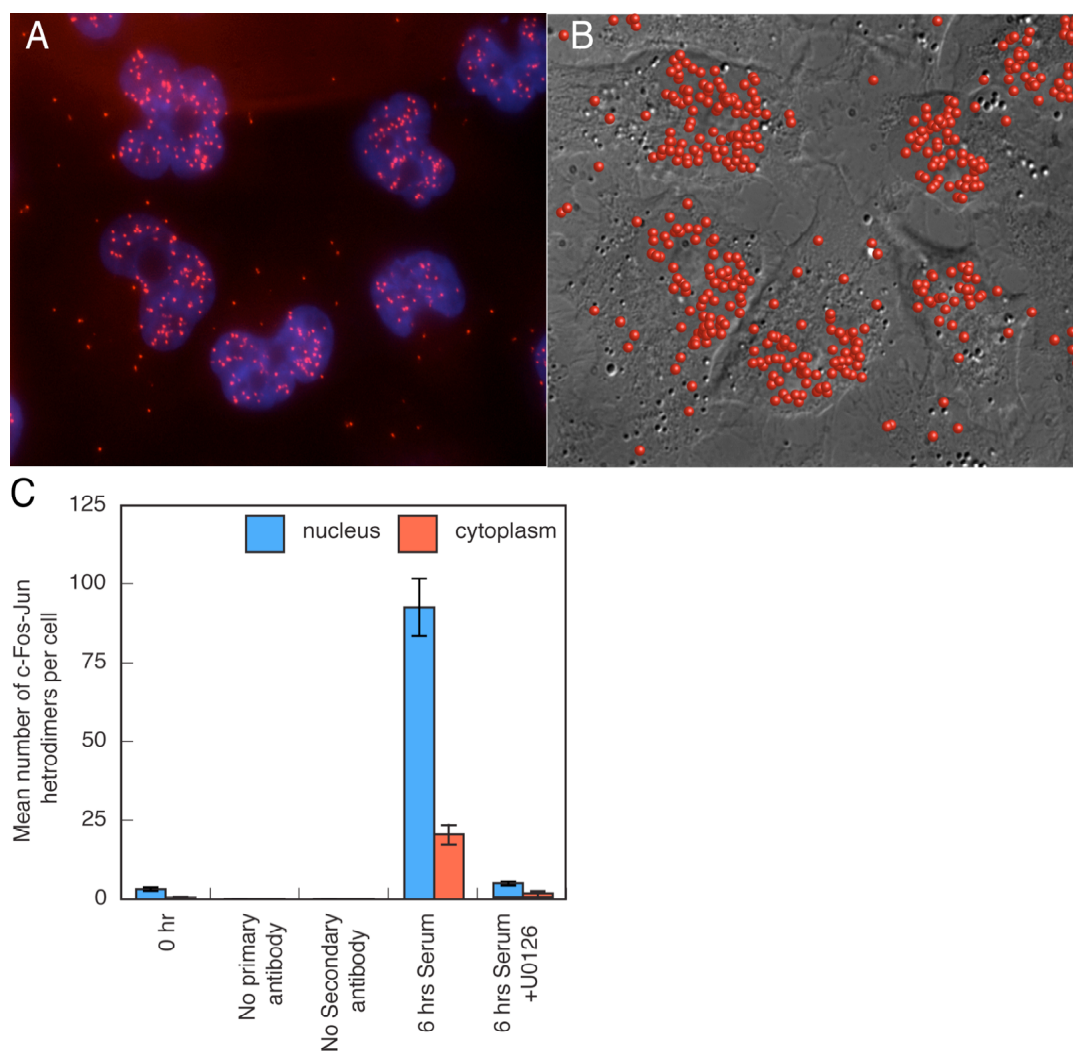
Figure 2.7 Detection of c-Fos-c-Jun heterodimers using the proximity ligation assay. HeLa cells were induced by the addition of serum and TPA for six hours. Proximity ligation assays were then performed, and the resulting amplification products were detected with fluorescein-labeled hybridization probes. A. Composite image showing a merged three-dimensional stack of images of several cells, where red spots represent individual c-Fos-c-Jun heterodimers, and blue DAPI staining highlights the nucleus of each cell. B. Identification and quantification of individual heterodimers using an image-processing algorithm. Filled red circles are drawn around each detected spot. C. Mean number of c-Fos-c-Jun heterodimers per cell under various conditions, in the nucleus and in the cytoplasm. The means were calculated by counting 100 cells in each category, and the error bars represent the 95 % confidence interval.

To further demonstrate that the signals are specific to the c-Fos-c-Jun heterodimers, a recombinant protein in which the coding sequence of c-Jun was fused with the coding sequences of c-Fos (with a spacer sequence in between them) was engineered (Figure 2.8 A). This recombinant construct was placed under the control of a doxycycline promoter, and it was integrated into the genome of a HeLa cell line that constitutively expressed the doxycycline-controlled transactivator (*88*). Even in the absence of serum induction, this cell line expresses the c-Fos-c-Jun fusion protein upon removal of doxycycline from the culture medium. As shown in Figure 2.8 B, the fluorescent spots were visible only when this hetrologous gene was turned on by the removal of doxycycline. Since the removal of doxycycline from the culture medium leads to the production of just one extra protein in the cell, the c-Fos-c-Jun fusion protein, these experiments unambiguously demonstrate that the spots produced by PLA represent authentic c-Fos-c-Jun heterodimers.
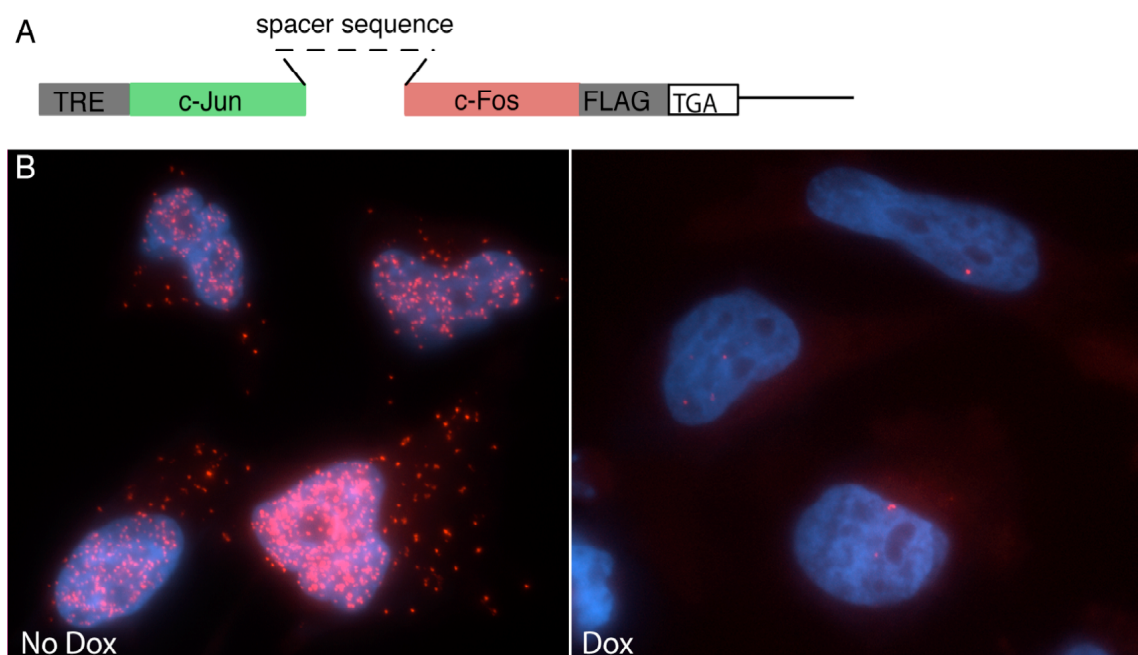
Figure 2.8 Specificity of the proximity ligation assay for the detection of c-Fos-c-Jun heterodimers using an engineered recombinant protein. A. Schematic representation of the engineered recombinant protein containing both the c-Fos and c-Jun sequences. B. Composite image showing merged three-dimensional stacks of images of several cells induced for four hours in the absence of doxycycline (left), and in the presence of doxycycline (right), with red representing individual c-Fos-c-Jun heterodimers, and blue DAPI staining highlights the nucleus of each cell.

2.2.5 c-Fos-c-Jun Heterodimers Show Less Cell-to-Cell Variation

In order to study variation in the number of c-Fos-Jun-heterodimers in each cell, the HeLa cells were imaged at various times after serum addition. PLA spots were counted using the same algorithm used to count mRNA spots. The distribution of the number of PLA spots per nucleus four hours after induction is shown in Figure 2.9. Compared to the distribution of c-Fos and c-Jun mRNAs, the distribution of c-Fos-c-Jun heterodimers was less variable. While the Fano factors of c-Fos and c-Jun mRNAs were 70 and 28, respectively, the Fano factor for the heterodimers was only 5, indicating that their distribution departs little from a Poisson distribution (Figure 2.9, green dotted line).

This indicates that the great variation observed in the mRNA distribution is diminished at the level of heterodimers. As discussed earlier, this is because gene expression noise is generally lower in the proteins, as compared to their parent mRNAs, due to the longer half-lives of proteins. Moreover, the observed reduced noise in the number of heterodimers reflects the half-life of the heterodimers, which is longer than that of their component proteins (*57*).
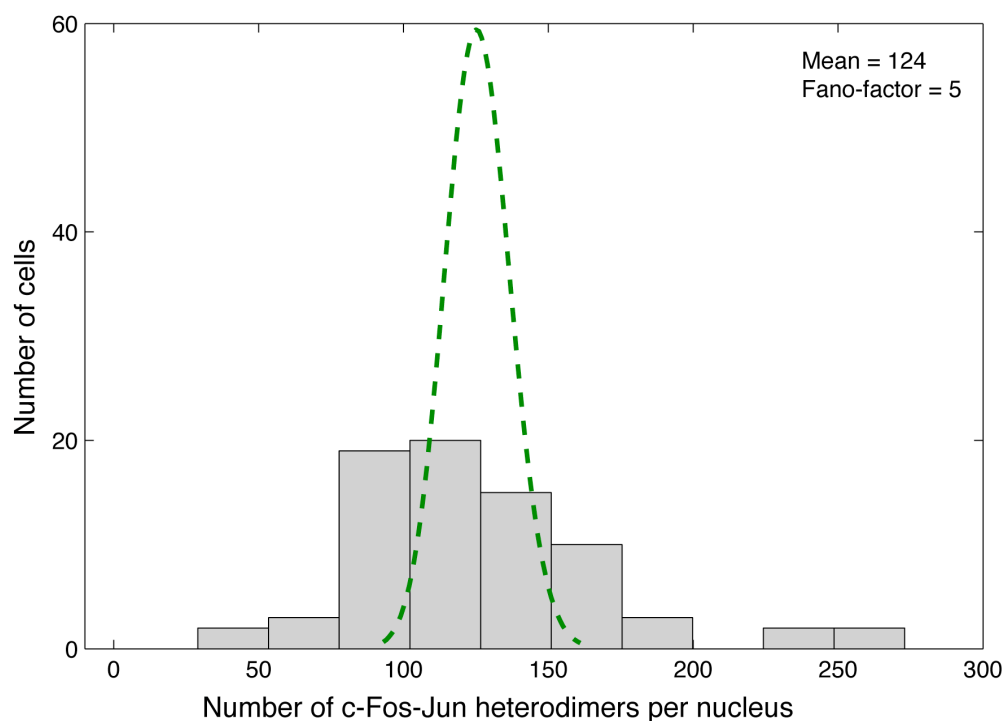
Figure 2.9 c-Fos-c-Jun heterodimers exhibit less cell-to-cell variation. Histogram showing distribution of c-Fos-c-Jun heterodimers per nucleus as determined by PLA after induction with serum + TPA for 4 hours.

The distribution of heterodimers in the cell population as a function of time after serum addition is presented in Figure 2.10. These results show that the heterodimers appear within 15 minutes of serum induction, and are present at a level of about 100 heterodimers per nucleus between four and ten hours, finally declining after 12 hours. The Fano factor varied from 3 to16 over this time course.
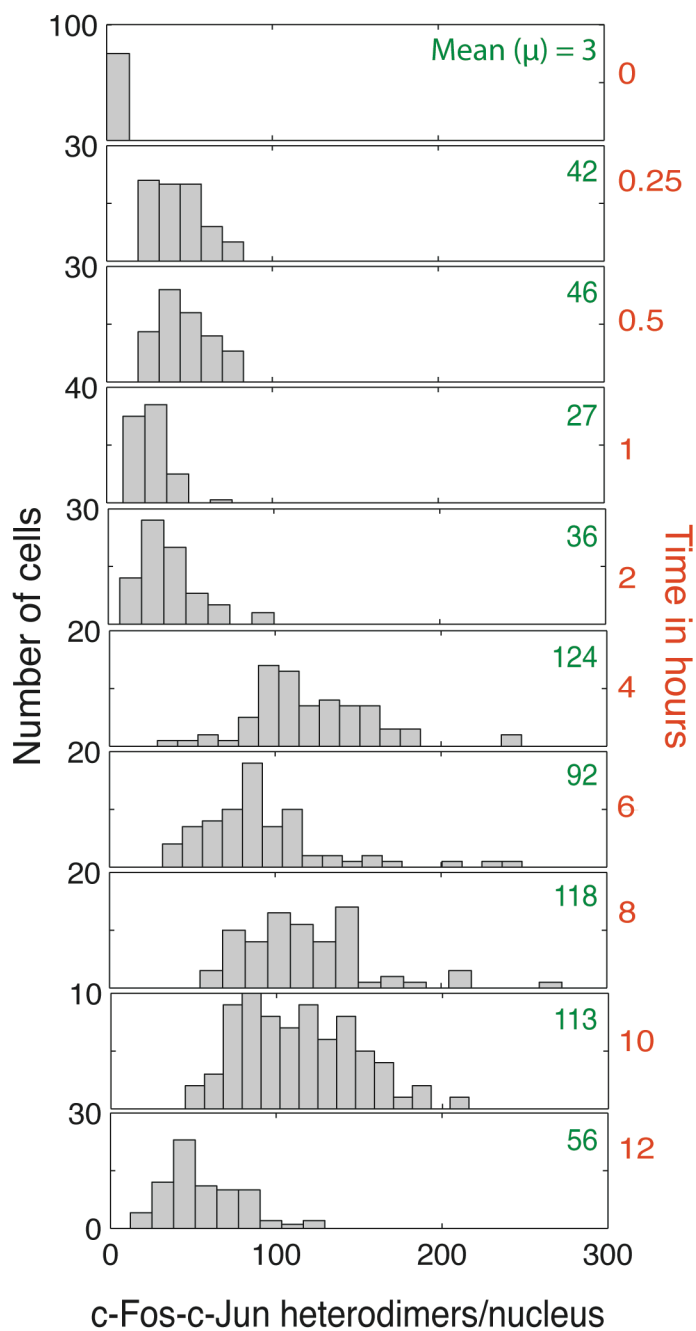
Figure 2.10 Cell-to-Cell variation in the number of c-Fos-c-Jun heterodimers as a function of time. Histograms showing the distribution of c-Fos-c-Jun heterodimers per nucleus (and their observed mean, shown in green letters) at various time interval after induction with 20 % serum and TPA.

2.2.6 The Expression of Collagenase and Cox-2 Genes is Stochastic and Not Correlated with Each Other

As discussed above, we found that cell-to-cell variation in c-Fos and c-Jun mRNAs is not propagated to the level of c-Fos-c-Jun heterodimers – their levels were relatively similar between cells. Since, these dimers are less variable from cell to cell, the level of variation in the expression of the downstream genes collagenase and cox-2 was explored. The levels of these downstream mRNAs in the same cells were measured as a function of time, using single-molecule FISH and two sets of probes, one specific for collagenase mRNA, and the other for cox-2 mRNAs, each set labeled in a different color. With a great surprise, the expression of collagenase and cox-2 was found to be highly variable (Figures 2.11 A and B). After six hours of expression, only a few cells expressed collagenase mRNA (4 out of 50 cells), however, when collagenase mRNA was expressed, there were 75 to 200 molecules in each cell. By contrast, cox-2 was expressed in a larger fraction of the cells, with expression levels ranging from 0 to 350 molecules per cell. A set of three fields that had at least one cell expressing each mRNA is shown in Figure 2.11 A. The distribution of each mRNA species is shown in the plot and in the histograms outside the plot in Figure 2.11 B.
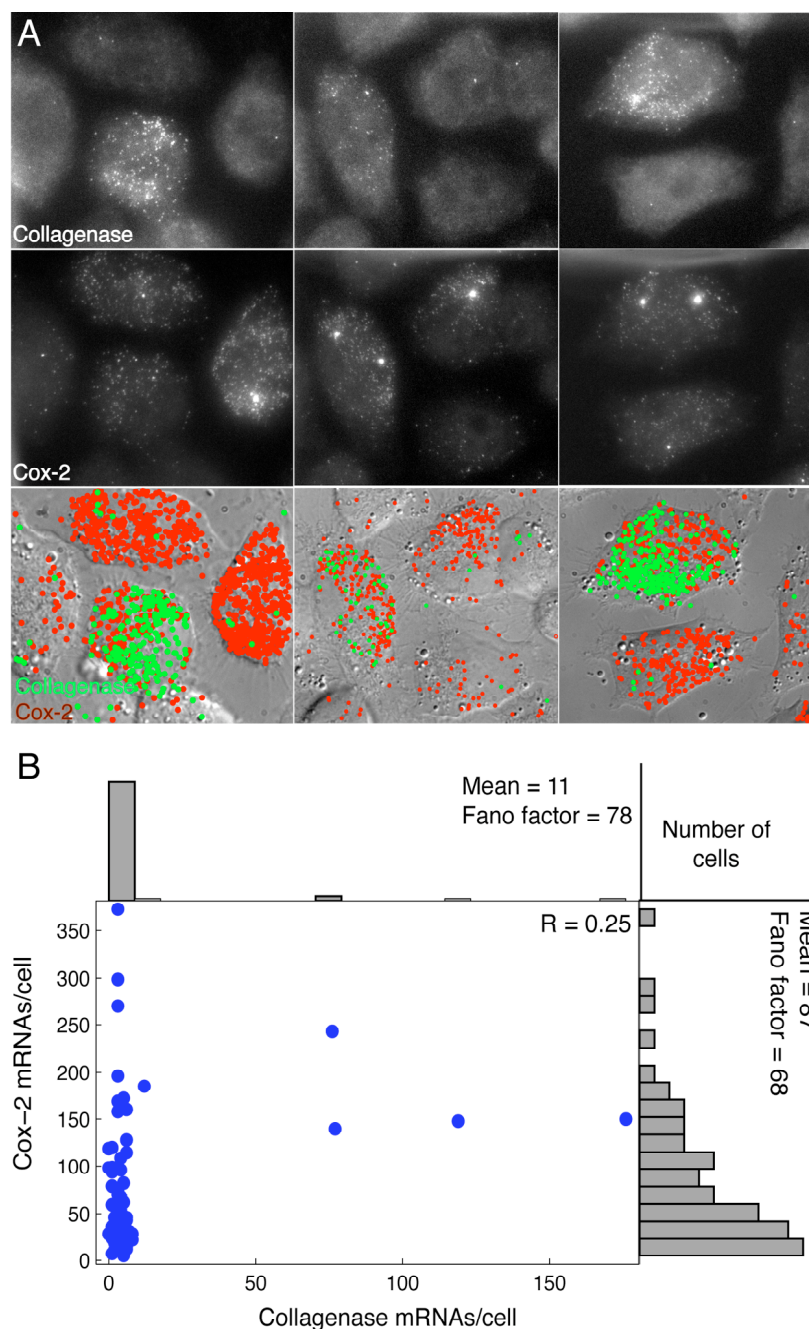
Figure 2.11 Expression of collagenase and cox-2 genes is stochastic. A. The upper and middle panel shows three-dimensional merged raw images of cells expressing collagenase and cox-2 mRNAs, respectively. The lower panels show molecules identified with an image processing algorithm (green and red representing signals from collagenase and cox-2 mRNAs, respectively) overlaid on DIC images. For these images, the fields were chosen with the criterion that there would be at least one cell expressing each mRNA (an infrequent occurrence). B. Scatter plot of total collagenase and cox-2 mRNAs after induction of cells with serum and TPA for six hours. Histograms outside the plot indicate the distribution of collagenase mRNA/cell (top) and cox-2 mRNA/cell (right).

In keeping with the stochasticity of the expression of c-Fos mRNA and c-Jun mRNA, we found that expression of these two mRNAs is not correlated with each other, and the pair gave a correlation coefficient of only 0.25. Furthermore, we measured how the level of their expression as a function of time. Figure 2.12 (top and middle panels) shows that for the first two hours after the addition of serum, neither of the two mRNAs can be detected. Their expression peaks between four to six hours, and ultimately declines. At no point during their expression do they show any significant correlation between each other (Figure 2.12, bottom panel).
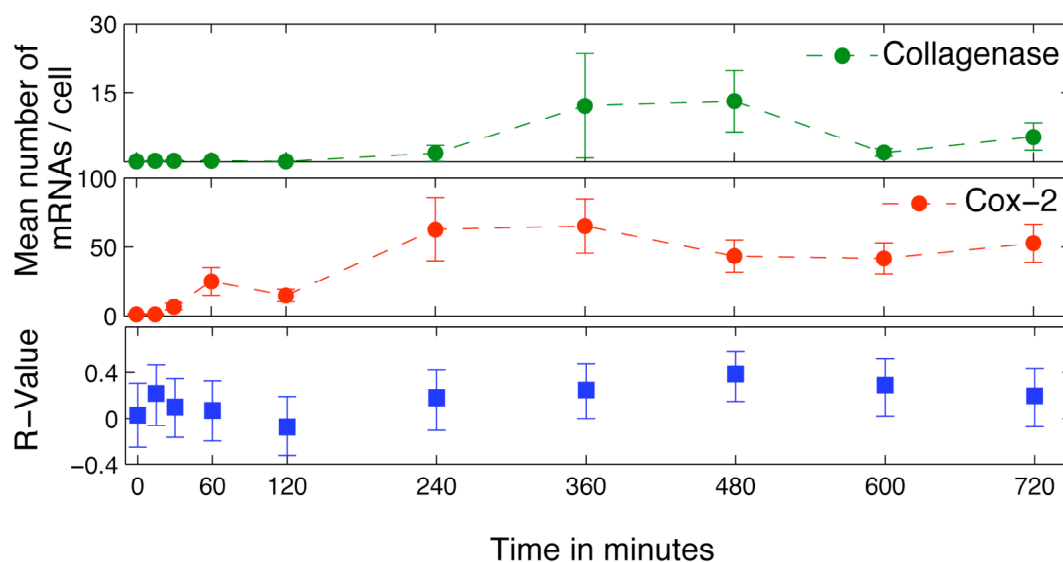
Figure 2.12 Expression of collagenase and cox-2 mRNAs are not correlated from cell to cell. Mean numbers of collagenase and cox-2 mRNAs per cell (top, middle) and the correlation coefficient (R-value) between collagenase mRNA and cox-2 mRNA (bottom) as a function of time. The error bar represents a 95 % confidence interval. Mean and R-values were obtained by counting 100 cells in each category.

These results indicate that the expression of collagenase mRNA and cox-2 mRNA in any given cell is independent of each other at any given time, and occurs in randomly initiated bursts that are uncorrelated. The evidence for bursts in synthesis was also seen by the clustering of many mRNAs at the gene locus in the case of each gene.

2.2.7 Expression of Collagenase and Cox-2 mRNAs are Poorly Correlated with the Transcription Factor Heterodimers that Induce Them

Since a large variation in the level of collagenase mRNA and cox-2 mRNA was observed, whether these variations correlate with the levels of c-Fos-c-Jun heterodimers in the same cell was explored. PLA for c-Fos-c-Jun heterodimers and single-molecule FISH for collagenase and cox-2 mRNAs within the same cells (that were induced with serum for six hours) was performed. In these three-label imaging experiments, collagenase was detected using cy5, cox-2 using Alexa 594, and PLA signals with fluorescein-labeled probes. The results are shown in Figures 2.13 A and B. Surprisingly, it was found that the cells contained a similar number of heterodimers show different amounts of collagenase and cox-2 mRNAs with many cells showing no expression of either of these two genes (Figures 2.13 A and B). The correlation between the number of heterodimers and the number of each of the two mRNAs was 0.47 and 0.46, respectively.
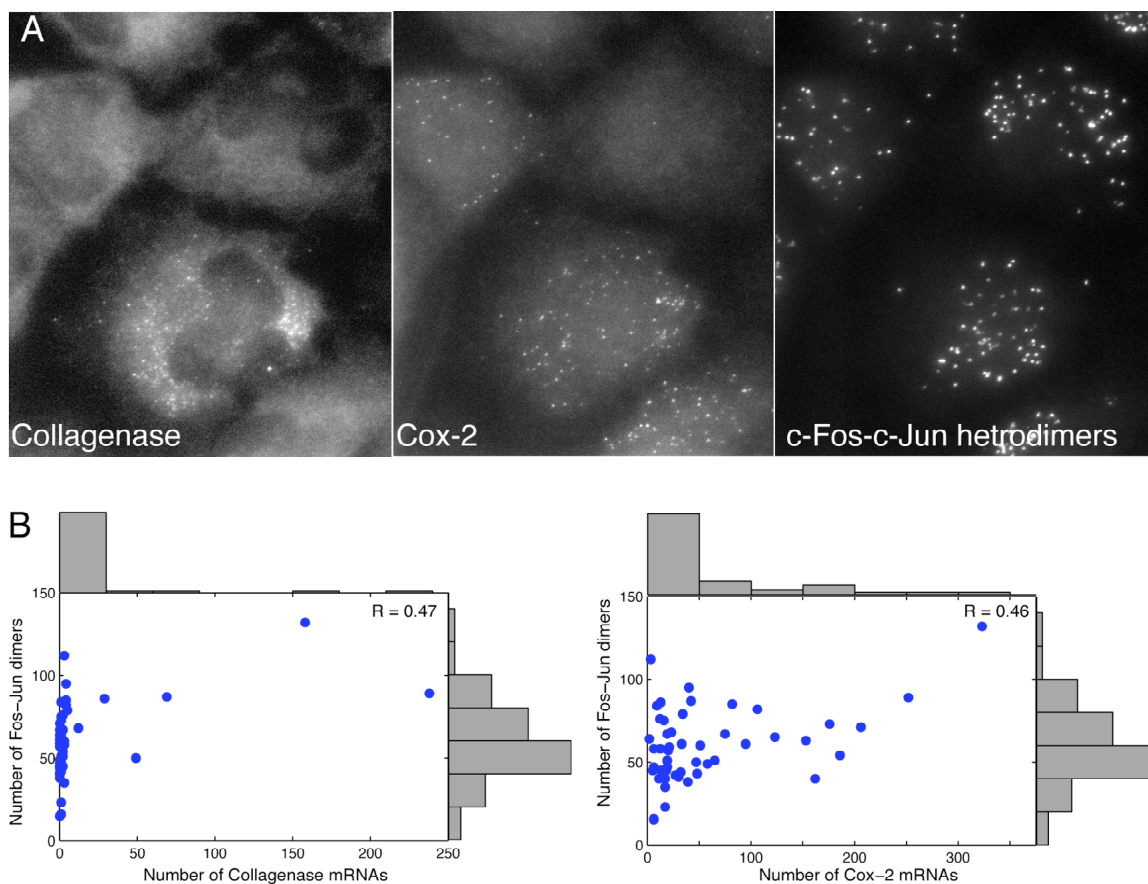
Figure 2.13 Expressions of collagenase and cox-2 genes are poorly correlated with the number of c-Fos-c-Jun heterodimers. A. Three-dimensional merged raw images of cells expressing collagenase mRNA (left), cox-2 mRNA (center) and individual c-Fos-c-Jun heterodimers (right) B. Scatter plot of total collagenase mRNAs and c-Fos-c-Jun heterodimers (left) and total cox-2 mRNAs and c-Fos-c-Jun heterodimers (right) in 50 individual cells after induction of cells with serum + TPA for 6 hours. Marginal histograms indicate the distribution of collagenase (top left), cox-2 (top right) mRNAs / cell and c-Fos-c-Jun heterodimers / nucleus (right).

The manner in which these correlations change as a function of time after the addition of serum is shown in Figure 2.14. These plots indicate that, although a certain minimum level of heterodimer (indicated by the blue line) is needed for the expression of each of the two mRNAs, the expression of neither mRNA correlates significantly with the number of heterodimers at any time after induction (Table 2.1).
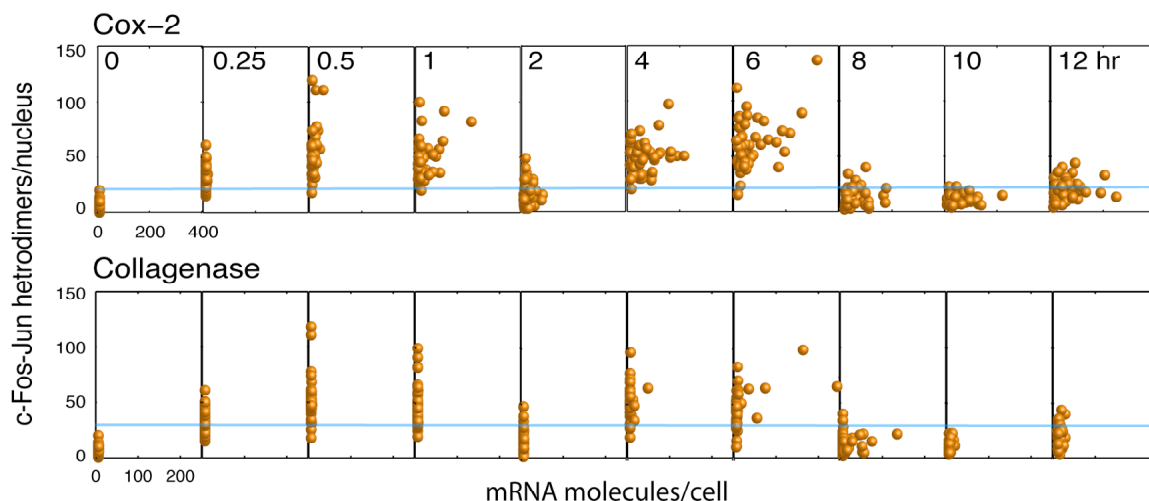
Figure 2.14 Correlation between c-Fos-c-Jun heterodimers and collagenase and cox-2 mRNAs as a function of time. Scatter plots showing the number of cox-2 mRNAs and the number of c-Fos-c-Jun heterodimers (top) and the number of collagenase mRNAs and the number of c-Fos-c-Jun heterodimers (bottom), at various times after induction.

Table 2.1 Correlation Coefficient Values: I (between c-Fos-c-Jun Heterodimers and Cox-2 mRNAs); and II (between c-Fos-c-Jun Heterodimers and Collagenase mRNAs) at Various Time Intervals

Time in Hours after serum addition	0	0.25	0.5	1	2	4	6	8	10	12
I	0.04	0.04	0.45	0.39	-0.05	0.29	0.43	0.11	0.10	0.13
II	0.12	-0.31	-0.10	0.08	0.04	0.16	0.46	0.08	0.38	0.21

These results demonstrate that, although the heterodimers are needed for the expression of collagenase mRNA and cox-2 mRNA, the presence of these heterodimers in the nucleus is not sufficient to produce these mRNAs at any given time. Instead, these two mRNAs are produced in randomly initiated bursts of expression, followed by periods of no expression, but steady mRNA decay. Therefore, at the time of fixation of the cells, only recently produced mRNAs are visible.

2.3   Discussion

The propagation of gene expression noise in an archetypical gene expression pathway was studied. The pathway that was chosen involves the production of c-Fos mRNA and c-Jun mRNA after serum induction, translation of these mRNAs into their respective proteins, formation of heterodimers between the two proteins, and induction of downstream genes (collagenase mRNA and Cox-2 mRNA) by the heterodimers (Figure 2.1). In this pathway, cell-to-cell variation in the number of c-Fos and c-Jun mRNAs, c-Fos-c-Jun protein heterodimers, and collagenase and cox-2 mRNAs, was measured. c-Fos mRNA and c-Jun mRNA was imaged in the same cells to determine if they are expressed in a coordinate manner in individual cells.   Furthermore, c-Fos-c-Jun heterodimers were simultaneously imaged, along with collagenase mRNA and Cox-2 mRNA in the same cells, in order to explore correlations between the three. All of these measurements were performed as a function of time after the addition of serum.

The expression of c-Fos, c-Jun, collagenase and cox-2 mRNAs occurs through transcriptional bursts. Several copies of mRNAs are transcribed during each burst. On an average 150-200 copies of mRNAs were made for each individual gene. Such pulsatile pattern of gene expression has also been observed previously in yeast, *Drosophila* and mammalian cells, suggesting that occurrence of transcriptional bursts in expression of endogenous genes is a common phenomenon (*4, 26, 27, 34*). However, recently in one of the study in *Drosophila* embryos, both stochastic and synchronous pattern of gene expression was observed for a set of developmental genes. This change in expression pattern was found to be due to presences or absences of RNA polymerase II at the gene before their induction. Hence the genes containing RNA polymerase

prepositioned on them before their induction showed synchronous pattern of gene expression whereas others that did not contain RNA polymerase II prepositioned on them before their induction showed stochastic pattern of gene expression (*9*). Furthermore, using fetal and neonatal pituitary tissues, it was observed that the pulsatile expression pattern of pituitary hormone gene became stabilized as the tissue develops (*37*). This two examples of developmental genes suggest that transcriptional bursts is a general phenomenon for gene expression, and the pattern of gene expression of certain genes might get changed during their developmental stages.

Although both c-Fos mRNA and c-Jun mRNA are expressed during the same time period (15 to 30 minutes), and they are expressed, on an average, in a coordinated manner after the addition of serum (*55*), their levels do not correlate with each other in individual cells. This lack of correlation is because each of the two mRNAs is expressed in randomly initiated bursts in different cells, followed by a period of no RNA synthesis during which there is steady RNA decay (*8*). The short half-life of these mRNAs (the half-life of c-Fos mRNA and c-Jun mRNA is 9 minutes and 11 minutes, respectively (*89, 90*)) contributes to the observed variation. Gandhi et al showed similar uncorrelated expression in yeasts. While measuring the level of coordination in the expression of functionally related genes within single *Saccaromyces cerevisiae* cells, transcription of these genes was found to be not coordinated in individual cells due to stochastic fluctuations in gene expression (*35*).

Raj et al. (*8*) showed that the high variation in the number of mRNAs is buffered for the protein products, because the half-life of proteins ranges from hours to a day, whereas mRNA half-life range from a few minutes to hours. The extent of buffering of

this noise depends on the actual half-life of the proteins. The longer the half-life, the greater is the buffering. For a multi-subunit complex, each subunit will have its own characteristic variation. Therefore, when a given cell assembles a multi-subunit complex, the number of molecules of complex that the cell can make, will be equal to the number of the type of subunit that is present in the lowest amount. The remaining excess subunits will remain, either unused or, in some cases, will be used for the assembly of other complexes.

Multi-subunit complexes can have longer half-life then the component proteins. This is the case for c-Fos-c-Jun heterodimers (*57*). Therefore, it is expected that the variation in the number of c-Fos-c-Jun heterodimers will be less compared to the variation in individual c-Fos and c-Jun proteins. This is likely to be the reason why there was low variation in the amounts of c-Fos-c-Jun heterodimers in each cell.

Given the relatively low level of variation in c-Fos-c-Jun heterodimers, a priori, one expects that the downstream genes will show less variation. However, the opposite was observed. Cell-to-cell variation in the expression of two downstream genes was very high, and their expression was not correlated with each other in individual cells. Furthermore, although the downstream genes were expressed in a noisier manner then the upstream genes, it was not due to the propagation of noise from the former to the latter. This is because only a limited variation in the number of heterodimers was observed. These observations suggest that the transcription machinery is inherently noisy, both for the upstream genes and for the downstream genes – and this occurs for mechanistic reasons.

The mechanism of inherently noisy transcription likely resides in the structure of chromatin in higher eukaryotes. In the interphase nucleus, chromatin exists in an extremely compact organization. Therefore, even if transcriptional activators are present in large numbers in the nucleus, they cannot access the promoter region of the gene. Random "breathing" events at the promoter sites in the chromatin permit the initial entry of the activator proteins. Once bound at the site, these proteins attract the chromatin decondensation apparatus and the RNA polymerization machinery, which lead to decondensation of the relatively large region of chromatin surrounding the gene locus (*91*). Transcription of all the genes located in the decondensed region can then take place, if their respective transcription factors are present. The c-Fos, c-Jun, collagenase, and cox-2 genes are located on chromosomes 1, 14, 11, and 9, respectively. This explains why c-Fos, c-Jun, cox-2, and collagenase expression were not correlated with each other in current study and observed co-regulated expression of two reporter genes that were present at the same genomic locus in Raj et al (*8*).

There are a few other mechanisms that may occur during transcription and that might be the source of these bursts of synthesis. The first one is variation in the availability and the retention of the transcriptional activators for their genes. Within a cell, these molecules exist in very low-copy number; hence their characteristic binding times and falling off times might result in pulsatile mRNA synthesis. Secondly, pre-initiation complex proteins are assembled sequentially near promoter regions, and similar to the activators and repressors, these components are also present in low copy number, and they are responsible for the transcription of most of the mRNA in a cell. Hence, competition for these proteins might result in the stochastic expression of any given gene.

In summary, the transcription patterns of c-Fos, c-Jun, collagenase and cox-2 mRNAs were studied as a function of time at a single molecule level in HeLa cells and evidence for transcriptional bursts have been provided. Furthermore, despite of their coordinated expression at the population levels, the expression of c-Fos and c-Jun; and cox-2 and collagenase mRNAs was found to be not correlated with each other in individual cells. Lastly, by imaging and counting individual c-Fos-c-Jun heterodimers and cox-2 and collagenase mRNAs within the same cells, it was shown that in a gene regulatory pathway, even though the variation in the expression of upstream gene mRNAs gets buffered at their protein levels, the expression of downstream genes are still stochastic due to their own inherent property. These data directly provided evidence that the expressions of functionally related genes are coordinated post-translationally.

2.4   Experimental Methods

2.4.1 Cell Culture

HeLa cells were cultured in modified Eagle's minimal essential medium (Sigma, St. Louis, MO) supplemented with Tet-system-approved 10 % fetal bovine serum (Clontech, Mountain view, CA). For the induction experiments, the cells were cultured on gelatin-coated glass cover slips and serum-starved for 48 hours. After starvation, expression from the c-Fos and c-Jun genes was induced by adding 20 % serum for 0, 0.25, 0.5, 1, 2, and 4 hours. At the end of induction, cells were fixed (along with serum-starved cells and cells growing in regular medium, as controls to count the basal level of c-Fos and c-Jun mRNA expression). For the expression of the downstream genes, collagenase and

cox-2, HeLa cells were induced with 20 % serum plus 200 µM TPA for 0, 0.25, 0.5, 1, 2, 4, 6, 8, 10, and 12 hours.

2.4.2 Cloning of c-Fos-c-Jun Fusion Protein

To tether c-Jun and c-Fos proteins together, c-Jun gene was amplified using specific primers from DNA obtained from HeLa cells and was cloned into pCR-4 TOPO cloning vectors. The primer sequences were: 5'-UTR_c-Jun forward primer, GTGTCCCCCGC TTGCCACAG; and 3'-UTR_c-Jun reverse primer, TCAGCCCCCGACGGTCTCTC. Utilizing site directed mutagenesis (Agilent Technologies, Santa Cruz, CA), BamHI and MluI restriction enzyme sites were created before the start codon in the c-Jun gene and at the end of the c-Jun gene. The BamHI-MluI-digested fragment containing the c-Jun coding sequence was inserted into a pTRE-c-Fos-Hygromycin vector (*75*) in front of the c-Fos coding sequence. A fragment containing a FLAG-tag sequence with a stop codon (5' GAC TAC AAG GAC GAC GAC GAC AAG TGA-3') was inserted at the end of the c-Fos coding sequence, followed by an EcoRI restriction enzyme recognition site, using site-directed mutagenesis. Furthermore, the c-Jun stop codon was removed, and a 24-nucleotide oligonucleotide encoding many glycine and serine residues was inserted between the last codon of the c-Jun coding sequence and the initiation codon of c-Fos, using site-directed mutagenesis. The primer sequences were as follows. Insert_24nt_forward: TAACGCAACAGTTGCAAACATTTAAGCTTGGGGGATCAG GCTCGAGCACGCGTGCCACG, Insert_24nt_reverse: CGTGGCACGCGTGCTCGAG CCTGATCCCCCAAGCTTAAATGTTTGCAACTGTTGCGTTA. The pTRE-Jun-linker -Fos-FLAG-hygromycin plasmid was linearized with FspI and was transfected into HeLa-tet-off cells (Clontech) using FuGene-6 (Roche, Applied Science, Indianapolis,

IN). After ten days of selection in a medium containing hygromycin (200 µg/ml) and doxycycline (10 ng/ml), individual clones were isolated and were confirmed by performing *in situ* hybridization.

2.4.3 Probe Sets and Antibodies

Sets of probes containing 48 labeled oligonucleotides were designed to hybridize to each target mRNA. The sequences of the individual probe sets can be found in Appendix A. The probes were 20 nucleotides in length, containing about 45 % GC and bind to mRNA target sequences that are at least two nucleotides apart from each other. Each probe set was purchased from Biosearch Technologies (Novato, CA) with a 3'-amino modification. Each probe set was pooled in equimolar amounts, coupled to either tetramethylrhodamine (TMR), Alexa 594, or Cy5 dyes, using their succinimidyl esters, and were purified by high pressure liquid chromatography. A detailed procedure for the attachment of labels and purification of the probes has been described previously (*66, 92*). Primary antibodies specific for c-Fos and raised in a mouse (Cat #SC-8047) and specific for c-Jun and raised in a rabbit (Cat #SC-1694) were obtained from Santa Cruz Biotechnology Inc. (Santa Cruz, CA). Fluorescein-conjugated goat anti-rabbit IgG secondary antibody (Cat #SC-2012) was purchased from Santa Cruz Biotechnology Inc. (Santa Cruz, CA), and Cascade blue conjugated goat anti-mouse IgG (Cat #C-962) was obtained from Life Technologies (Grand Island, NY).

2.4.4 Fluorescence *in situ* Hybridization

Cover slips containing HeLa cells were washed with 1X phosphate buffer saline (PBS), fixed in 4 % formaldehyde in 1X PBS for 15 minutes, washed with 1X PBS, and were permeabilized with 7 0% ethanol at 4 $^{\circ}$C for one to two hours. The cells were equilibrated with 10 % formamide in 2X SSC solution, and were then hybridized overnight with one or more probe sets. Hybridization was performed in a moist chamber maintained at 37 $^{\circ}$C, with the cover slips placed upside over the hybridization solution. The hybridization solution contained 10 % (w/v) dextran sulfate (Sigma), 1 µg/µl *E. coli* tRNA (Sigma), 2 mM ribonucleoside vanadyl complex (Sigma) to inhibit ribonucleases, 0.02 % RNase-free bovine serum albumin (Ambion), 10 % formamide (Ambion), and 10 ng/µl of each probe set. After hybridization, the cover slips were washed twice with a solution containing 10 % formamide in 2X SSC. The cover slips were mounted on glass slides, using deoxygenated mounting medium, and sealed with clear nail polish (*66*).

2.4.5 Proximity Ligation Assay

To image individual c-Fos and c-Jun heterodimers, cover slips containing HeLa cells induced with 20 % serum plus TPA were washed, fixed, permeabilized, blocked, and treated with both c-Fos and c-Jun primary antibodies, as described in Section 2.2.5. Anti-rabbit PLA plus and anti-mouse PLA minus probes were obtained from Olink Bioscience (Uppsala, Sweden), and were diluted 1:5 in blocking solution. The cells were incubated with the secondary antibody probe mixture, and the cells were then incubated with in a preheated humidified chamber for an hour at 37 $^{\circ}$C. The cells were washed twice in 1X PBS, and were then incubated with a ligation mixture containing connector oligonucleotides and ligase enzyme (diluted 1:5 with RNase-free water) for 30 minutes in

a preheated humidified chamber at 37 °C. The cells were once again washed twice with 1X PBS, and were then incubated with amplification solution (diluted 1:5 in RNase-free water) containing labeled oligonucleotides and polymerase enzyme in a preheated humidified chamber for two hours at 37 °C. Cells were washed twice with 2X SSC, and were mounted for imaging, as described in Section 1.4.4.

2.4.6 Proximity Ligation – Fluorescence *in situ* Hybridization Assay

To image c-Fos-c-Jun heterodimers and mRNAs for downstream target genes, a modified protocol was used. First, the entire PLA protocol up to the detection step was performed in the presence of RNase inhibitor (NEB, Ipswich, MA). Once the signals were detected and amplified, they were further fixed with 4 % formaldehyde in 1X PBS for 10 minutes, and then washed with 2X SSC. The cells were then hybridized overnight with a hybridization mixture containing labeled probe sets in a moist humidified chamber maintained at 37 °C. The rest of the protocol was followed as Described in Section 2.4.4.

2.4.7 Imaging

Wildfield epifluorescence microscope was used for imaging along with strong light source such as mercury lamp and cooled CCD camera. The resolution of a given microscope is proportional to the size of its objective lens used and is inversely proportional to the wavelength of light at which the samples are being observed. Hence, the ultimate limit to the resolution of a light microscope is set by the wavelength of visible light, which ranges from about 0.4 μm (for violet) to 0.7 μm (for far red). Furthermore, the limiting separation at which two objects can be still seen as distinct spots is called limit of resolution or diffraction limits of a given microscope. This

depends on both the wavelength of light (λ) and numerical aperture (NA) of the lens system used and can be obtained by following formula (*93*).

$$d = \frac{\lambda}{2 * NA} \qquad (2.2)$$

The NA is usually 1.4 for most of the 100X optical lens; hence the diffraction limit is roughly half of the wavelength of light used. For the epifluorescence microscope used in the current study, the diffraction limit was found to be 250 nm and was used as a limit in our image analysis computer program.

2.4.8 Image Analysis

For each image, 20 to 40 optical slices, 0.2 μm apart were acquired in each fluorescence channel with a one-second exposure using an Axiovert 200M inverted fluorescence microscope (Carl Zeiss, Thornwood, NY). The images were obtained using Openlab acquisition software (Perkin-Elmer, Sheffield, united Kingdom). Three-dimensional stacks of images were analyzed using custom computer programs written in Matlab (Mathworks, Natick, MA). These programs enhance the stack of images using a Laplacian filter optimized for the size of spots that we expect, permit users to select a threshold based on a three-dimensional display of spot intensity, segment the image based on the provided threshold, and produce a list of coordinates of the centers of all spots in three dimensions in each channel. Using the cell and nucleus boundaries and

location determined by DIC and DAPI images, the number of mRNAs and PLA signals were counted in each cell.

2.4.9 Statistical Analysis

75 to 100 cells were analyzed for each category of data reported in Figures 2.1 to 2.11. The data points represent the mean values, and the error bars represent a 95% confidence interval. The 95% confidence interval for calculating correlation coefficients (R-values) was obtained with a bootstrapping method in Matlab (Mathworks, Natick, MA).

CHAPTER 3

SINGLE-MOLECULE IMAGING OF TRANSCRIPTIONALLY COUPLED AND UNCOUPLED SPLICING

3.1   Introduction

As a eukaryotic gene is transcribed, the pre-mRNAs contain both coding sequences called exons, interrupted by non-coding sequences called introns. These introns need to be removed from the transcript before it is exported out of the nucleus to the cytoplasm, where it is translated into functional protein. The process of removal of the introns and joining of the exons occurs via two transesterification steps with the help of a large ribonucleoprotein complex called a spliceosome.

The spliceosome consists of five small nuclear ribonucleoproteins (snRNPs) and more then 100 other proteins. Each snRNP (U1, U2, U4, U5, and U6) is made up of a single uridine-rich, small nuclear RNA and multiple proteins.   As the introns are transcribed, the splicing machinery recognizes specific *cis*-acting sequences within the introns that defines their boundaries.   Once the introns are recognized, their excision occurs in two chemical steps: in the first step, the 5' splice site is cleaved and a lariat is formed. In the next step, the 3' splice site is cleaved and the exons adjacent to the splice sites are ligated to each other.   Once the exons are ligated, the spliceosome is disassembled, releasing its components for the *de novo* synthesis of other spliceosomes (*94*).

3.1.1 Co-transcriptional Splicing

Whether splicing is coupled to transcription, or whether it takes place after the pre-mRNA is released from the transcript, has been an unsolved question for several years. When Lamond et al. first discovered that splicing factors are concentrated in 20 to 50 sub-nuclear structures called speckles, it was suggested that they could be the sites where splicing occurs (*95*). However, later *in situ* hybridization studies indicated that, even though the speckles might be present next to the site where active transcription takes place, splicing does not occur in the speckles (*96, 97*). Instead, splicing occurs during the time when nascent transcripts are still tethered to their encoding gene via RNA polymerase II (*97-102*).

3.1.2 Proposed Models for Coupling of Transcription and Splicing

While it is nearly universally accepted that transcription and splicing are coupled, two views concerning the mechanism of coupling prevail: structural coupling and kinetic coupling. According to the structural coupling model, splicing factors are pre-positioned on the C-terminal domain of RNA polymerase II and attach to the introns as they emerge from the polymerase (*103-105*). The kinetic coupling model is based on evidence that links the rate of transcriptional elongation and splice-site selection. Owing to their high concentration and mobility (*106*), splicing factors directly assemble on the nascent introns into productive spliceosomes as fast as the RNA polymerase can synthesize them (*107, 108*). Hence, the rate-limiting step is not splicing, but rather, it is the completion of mRNA synthesis, 3'-end processing, and release. If there is a perturbation in transcription elongation, the co-transcriptional recruitment of splicing factors to splice sites is greatly affected (*109*). In addition, with the aid of a chromatin

immunoprecipitation assay, Listerman et al. showed that some splicing factors, such as U2AF65, U1, and U5 snRNP, associate with introns at the sites of genes undergoing transcription, supporting the co-transcriptional model of splicing (*110*). Further support for the kinetic coupling model comes from the finding that exon inclusion is promoted by an intrinsically slow RNA polymerization, or by nucleosomes that impede the progress of the polymerase (*111, 112*). Furthermore, there is evidence that the rates of the two processes are sometimes coordinated, ensuring that only fully spliced mRNAs are released (*36, 113-115*).

Although, the co-transcriptional removal of introns as they emerge from RNA polymerase provides an attractive explanation for the high fidelity of splicing in joining constitutively spliced exons in the proper sequential order, it is not ideal for explaining alternative splicing, wherein splicing must be slowed down until all of the splice sites involved in the choice have been synthesized (*30, 31, 37, 70, 72, 83, 116-120*). Is processing just delayed briefly until alternative splice sites are generated, or does alternative splicing result instead in the uncoupling of splicing from transcription, so that it is concluded post-transcriptionally? The former has been found to be the case for several alternatively spliced transcripts (*74, 101, 121*). However, the manner in which RNA-binding splicing regulators impact splicing-transcription coupling, in situations where strict tissue and developmental stage-specific alternative splicing patterns occur, remains to be explored. Moreover, when synthetic pre-mRNAs were injected into the nucleus, rapid splicing was observed, suggesting the possibility of post-transcriptional splicing (*122*).

3.1.3 Alternative Splicing

Alternative splicing leads to the joining of different 5' and 3' splice sites, allowing an individual gene to express multiple processed mRNAs that encode different proteins. This mechanism contributes to the generation of complex proteomes. This occurs frequently in metazoans, in the nematode, *Caenorhabditis elegans*, in the fruitfly, *Drosophila melanogaster*, and in humans. Genetic and biochemical approaches have identified *cis*-acting regulatory elements, such as enhancers and silencers, and *trans*-acting factors that control the alternative splicing of specific pre-mRNAs. Analyses of expressed sequence tags, and microarray data, estimated that 73% of human genes are alternatively spliced, making alternative splicing the rule, rather than the exception. Apart from constitutive splicing, alternative splicing also plays an important role in the quantitative control of gene regulation, by targeting RNAs for nonsense-mediated decay (*12, 13*). There are many examples of cell-line and tissue-specific proteins that bind to introns and cause alternative splicing (*119*).

3.1.4 Sxl Protein in *Drosophila*

One of the many proteins involved in alternative splicing is Sxl protein in the fruitfly, *Drosophila melanogaster*. In *Drosophila*, the expression of the binary switch gene Sex-lethal (Sxl), which controls somatic sexual development, is regulated at the level of alternative splicing. The X-chromosome to autosome ratio determines initial sexual identity. The activation of the Sxl gene depends on four X-encoded proteins: SISA, SCUTE, RUNT, and UNPAIRED.

In females (XX), when the expression levels of these X-linked proteins reach a threshold, the four X-encoded transcription factors, SISA, SCUTE, RUNT, and

UNPAIRED, stimulate the transcription of the Sex-lethal (Sxl) gene by binding to the Sxl establishment promoter (Sxl-Pe). When this occurs, the resulting mature mRNAs do not contain exon-2 and do not contain exon-3, joining exon-1 to exon-4, thus preventing the inclusion of exon-3, which contains a stop codon, thereby resulting in the formation of fully functional Sxl protein. In males (XY), on the other hand, the expression levels of the four X-linked proteins produced from the single X chromosome fail to reach the threshold concentration needed to activate the Sxl establishment promoter (Sxl-Pe), and the gene remains turned off. The establishment promoter remains active for only a short period of time, becoming inactive at the cellular blastoderm stage, about three hours after fertilization. Figure 3.1 shows the initial establishment of Sxl protein during early embryogenesis. To further maintain the expression of the Sxl gene during development in females, an autoregulatory mechanism is established, in which Sxl protein controls its own synthesis by promoting the female-specific splicing of Sxl pre-mRNAs transcribed from the Sxl maintenance promoter (Sxl-Pm) (*123*).
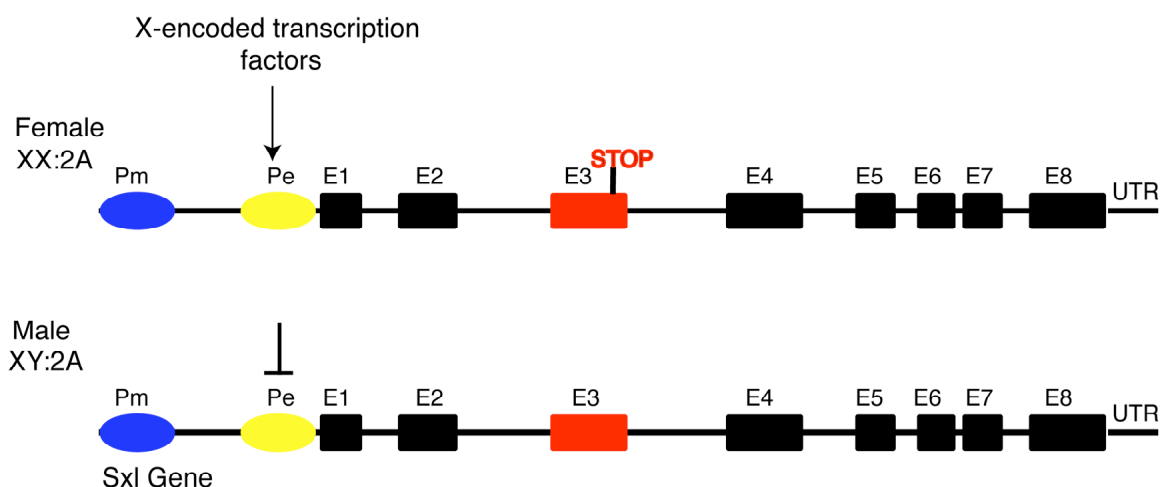
Figure 3.1 Initial establishment of Sxl protein during early embryogenesis.

3.1.5 Splicing of Sxl Gene

The Sxl gene is made up of seven introns and eight exons, within which exon-3 is the male-specific exon. In females, after the initial establishment, the constitutively active maintenance promoter drives the expression of the Sxl gene. The Sxl protein binds to multiple polypyrimidine tract sites on both the upstream and the downstream introns of male exon promoting female-specific splicing of Sxl pre-mRNA, skipping the male exon. The translation of the resulting mRNA ensures the maintenance of female identity by providing a continuous source of Sxl protein. In males, on the other hand, since the initial establishment of Sxl protein does not take place, exon-3 is incorporated by the default splicing machinery, which has an in-frame stop codon. Hence, a truncated, non-functional protein is made. (*124*).

By performing ribonuclease protection assays on poly (A) RNAs, the splicing of the regulated exons (exon 2, 3. and 4) was observed to be slower than the splicing of the unregulated exons. Also, Sxl protein requires poly (A) binding protein to promote female-specific splicing (*125*). Furthermore, while studying the mechanism of the default and the regulated splicing of the Sxl gene, Horabin et al. (*126*) found that multiple *cis*-acting elements, both upstream and downstream of the male exon, are required, and the 5' splice site of the male exon appears to be dominant in regulation, whereas the 3'-splice site plays a secondary role. However, the timing of the splicing of the regulated exon, as compared to the timing of the unregulated exons is not known. Moreover, since the Sxl gene requires poly (A) binding protein to function, it might be possible that the splicing of the regulated male exon occurs after the transcript is released from the gene locus.

3.1.6 Proposed Study

The question of whether splicing is coupled to or uncoupled from, transcription during certain cases of alternative splicing will be addressed using a single-molecule *in situ* hybridization technique. The intracellular distribution and dynamics of individual molecules of pre-mRNAs and their spliced products were imaged utilizing a set of reporter genes, and also utilizing the classically well-studied alternatively spliced gene: Sex-lethal (Sxl) in *Drosophila*.

3.2   Results

3.2.1 Imaging Individual Molecules of pre-mRNA, mRNA and Introns

The coupling of splicing to transcription was first examined utilizing a pair of green fluorescent protein (GFP) reporters that have a tandemly repeated sequence, array 3, inserted into their 3'-untranslated region (3'-UTR) and one of two tandem arrays (1 or 2) inserted into an artificial intron (with canonical splice sites) placed in the middle of the GFP-coding sequence (Figure 3.2 A). The tandem arrays of randomly selected sequences were used to achieve single-molecule sensitivity with just one oligonucleotide probe. Doxycycline-controlled versions of the two genes were stably integrated into the genome of Chinese hamster ovary (CHO) cell lines. Upon induction by removal of doxycycline from the culture medium, both cell lines produced appropriately spliced mRNAs and exhibited GFP fluorescence.

The pre-mRNAs and their spliced products were imaged by fixing the cells after six hours of induction, followed by *in situ* hybridization with fluorescently labeled probes against array 3 repeats, and against either array 1 or array 2 repeats. Three classes of diffraction-limited spots were observed. Two of these corresponded to single molecules

containing either the intron array or the 3'-UTR array (Figure 3.2 B). The third class consisted of unspliced molecules containing both the intron and the 3'-UTR arrays (Figure 3.2 B). When the center of a spot seen in one channel was located within 0.25 µm of the center of a spot seen in the other channel, it was considered to be co-localized and was represented as an unspliced pre-mRNA (Figures 3.2 C and D).

Transcripts expressed from the two reporter genes exhibited striking differences in how they coordinate transcription and splicing. In the case of array 1, high levels of pre-mRNA accumulated at the gene locus, while pre-mRNA was rarely seen elsewhere in the nucleoplasm. While splicing and transcription were tightly coupled for array 1 transcripts, this was not true of array 2 transcripts. Most array 2 pre-mRNA molecules were scattered throughout the nucleoplasm, with little retention at the transcription site. In addition, the spliced introns from array 1 and 2 transcripts degraded differently. Only a few spliced array 1 intron molecules diffused away from the gene locus, whereas, a large number of array 2 intron molecules were found scattered in the nucleoplasm. For both constructs, the spliced mRNAs were exported efficiently into the cytoplasm, while the pre-mRNAs and the introns were retained within the nucleus (Figures 3.2 B-D).
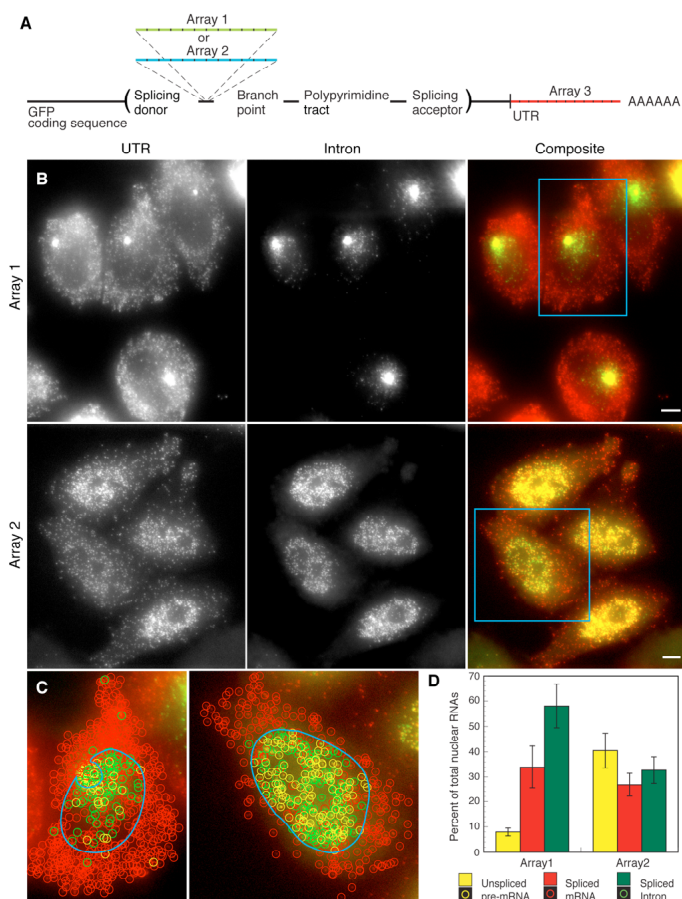
Figure 3.2 Imaging the intracellular distribution of single-molecules of pre-mRNAs and their spliced products expressing from a pair of reporter genes. A. Schematic depiction of two reporter genes. B. Images of cells from clone expressing the reporter gene in two fluorescence channels. The targets of the probes are indicated on the top of the panels, and the array within the intron is shown on the left. In the composite images, red represents the 3'-UTR and green represents the introns. C. Identification of RNA species using an image-processing program. Circles of different colors are drawn around each detected mRNA species. D. Percentage of three different species in the nuclei of individual cells from two cell lines. Examples of regions from which the counts were obtained are indicated by blue circles in C. Error bars represent 95% confidence intervals. The scale bar is 5 μm. This experiment was performed by Diana Vargas (*75*).

3.2.2 Pre-mRNA Molecules Dispersed in the Nucleus are Capable of being Spliced

To show that the array 2 pre-mRNAs that are dispersed into the nucleoplasm are substrates for splicing, the reporter was induced for a short period (two hours), and then, after turning off the reporter, the fate of the previously synthesized pre-mRNAs was monitored. In cells fixed immediately after two hours of induction, many pre-mRNA

molecules were seen scattered within the nucleoplasm, with little accumulation of spliced mRNA molecules in the cytoplasm (Figure 3.3 A). After the chase period, there was a decrease in the proportion of pre-mRNA molecules with a remarkable increase in spliced mRNA molecules in the cytoplasm (Figure 3.3 B), suggesting that the dispersed array 2 pre-mRNAs are splicing competent.
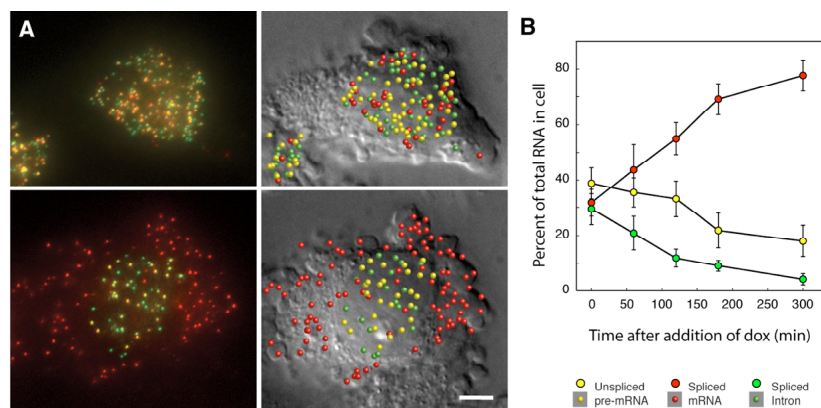
Figure 3.3 Demonstration that pre-mRNA molecules dispersed in the nucleus are capable of being spliced. A. Upper panels show composite images of cells in which the gene containing the array 2 as intron was induced for a brief period (two hours). Lower panels show images from the same batch of cells as above, but in which induction was followed by a period of suppression (two hour). Raw images are shown on the left, and overlays with colored circles identifying the RNA species are presented on the right. B. Percentage of three different RNA species in individual cells as a function of time after the addition of doxycycline. This experiment was performed by Diana Vargas (*75*).

3.2.3 The Intron with Array 1 is Removed Co-transcriptionally and the One with Array 2 is Removed Post-transcriptionally Irrespective of their Order in the Gene

In order to see if there will be any change in the splicing behavior of array 1 or array 2 when both arrays are included in the same pre-mRNAs, a pair of reporter genes, "array 1-array 2" and "array 2- array 1," in which the two arrays are present in the same pre—mRNA, but in different order (surrounded by the same splice sites as before), were constructed (Figure 3.4 A). The first intron was placed towards the 5' end, and the second intron was placed in the middle of the GFP-coding sequence. The detection of

spliced mRNA molecules was accomplished using 48 labeled oligonucleotides complementary to the GFP-coding sequence.

Cells expressing these reporters were imaged for the two intronic arrays and the GFP-coding sequence in three different fluorescence channels, and molecules corresponding to each of the seven possible permutations were computationally identified (Table 3.1). For both constructs, the partially spliced array 2-GFP pre-mRNA was one of the most abundant species, and was found scattered throughout the nucleus (Figure 3.4 B and Table 3.1). By contrast, the other partially spliced product, array 1-GFP, and the unspliced array 1-array 2-GFP or array 2-array 1-GFP pre-mRNAs were rarely detected, except at the gene locus. These observations indicate that, irrespective of their order in the transcript, array 1 is spliced co-transcriptionally and array 2 is spliced post-transcriptionally. Significantly, in the case of the array 2-array 1 transcript, array 2 was not spliced at the gene locus, even though the splicing apparatus assembled on the downstream array 1 intron and spliced it co-transcriptionally.
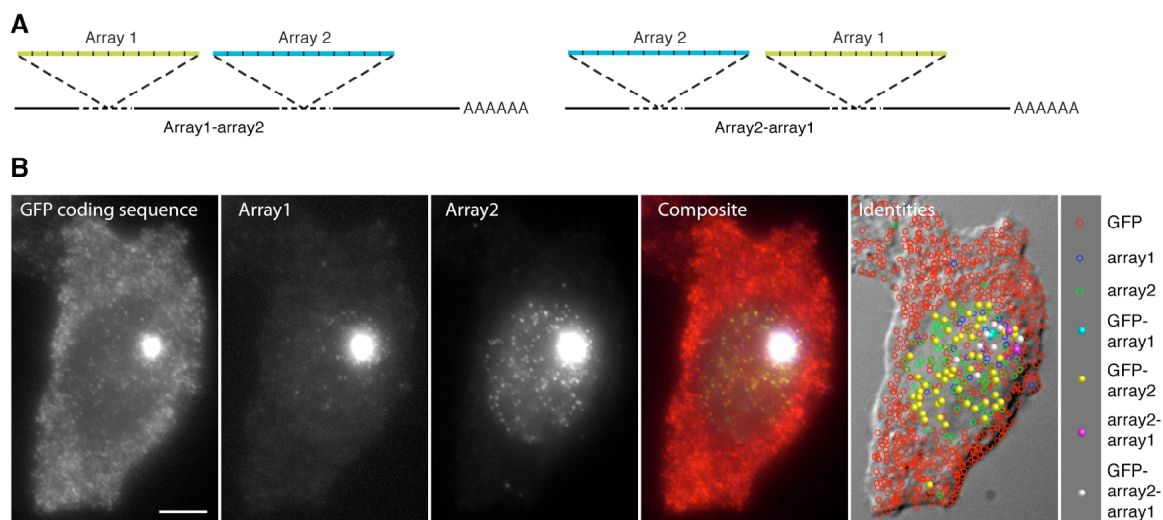
Figure 3.4 The splicing behavior of array 1 and array 2 remained the same irrespective of their order within the GFP coding sequence. A. Schematic representation of two genes that contain the two introns in a different order. B. Raw, composite, and interpreted images of a cell expressing the construct array 2-array 1 in the three different fluorescence channels that detect the GFP coding sequence, array 1, and array 2. The composite image shows the GFP-coding sequence in red, array 1 in blue, and array 2 in green. The color key on the right lists each of the seven combinations of spliced and unspliced RNA species that can occur.

Table 3.1 Percentage of pre-mRNA and Spliced Products in Individual Cells from Reporter Genes Containing Two Introns

Species	Spliced	Introns		Partially Spliced			Unspliced
	GFP	Array 1	Array 2	GFP-array 1	GFP-array2	Array1-array2	GFP-array1-array2
Array1-array2	31.4±6.0	21.04±6.8	15.5±6.1	6.6±2.7	17.6±4.7	3.7±1.5	4.0±1.6
Array2-array1	39.7±6.9	5.7±1.7	19.0±2.9	1.0±0.5	30.9±5.2	1.8±1.1	1.8±0.7

3.2.4 The splicing Behavior of Array 1 and Array 2 Remained the same, Irrespective of the Chromatin Context in which the Gene was Integrated

To demonstrate that the site of integration of the reporter genes within the genome does not influence the splicing behavior of the array 1 intron and the array 2 intron, these reporter genes were integrated at the same chromatin location using Flip-recombination.

Flip-recombinase target sites were integrated into different genomic sites in CHO cells and five different clones were isolated. The reporter genes containing array 1 or array 2 were separately inserted into these sites by targeted cloning. In each clone, the array 1 and array 2 genes are present at the same genomic locus, which varies between the clones. Introns with two arrays appear to be processed in their characteristic manner, irrespective of which clone they are in, indicating that chromatin context is not a significant determinant of their behavior, as seen in Figure 3.5.
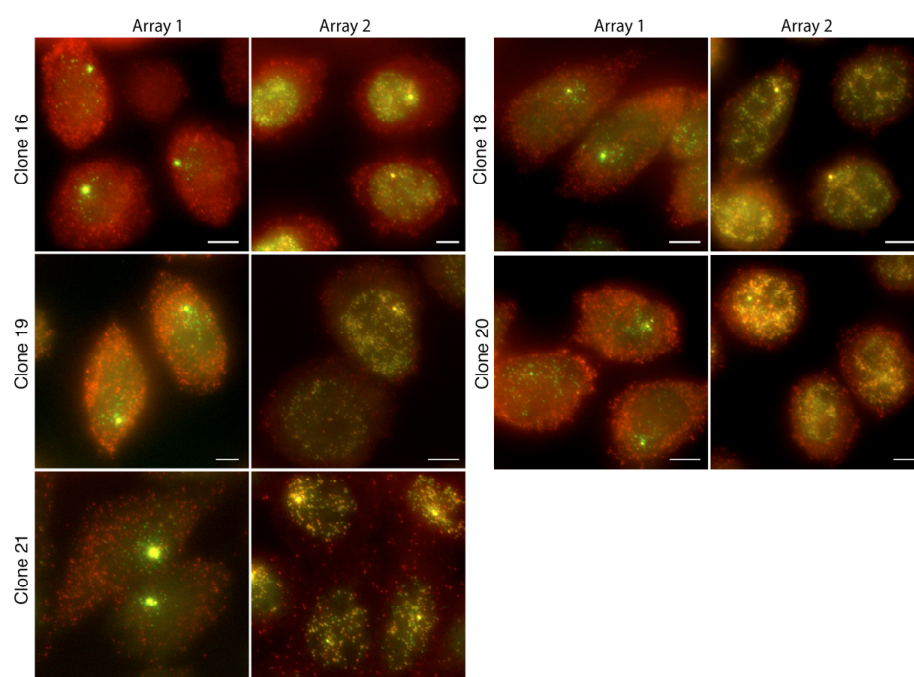
Figure 3.5 Demonstration that the location of the reporter genes within the genome of CHO cells does not influence their splicing behavior. Merged z-stacks in two colors are shown with green representing the intron signal, red representing the exon signal, and yellow representing the pre-mRNA signal.

3.2.5 Sequestration or Mutation of Intronic Polypyrimidine Tract Uncouple Splicing from Transcription

Post-transcriptional splicing of array 2 is either an intrinsic property of the array sequence, or it arises from interactions between array 2 and the splice sites in the GFP reporter construct. Secondary structure folding patterns of full-length GFP array 2

pre-mRNA predict that the polypyrimidine tract (a key intron recognition element that is situated towards the 3' end of introns) is sequestered in a double-stranded region. If sequestration of the polypyrimidine tract causes post-transcriptional splicing in transcripts containing array 2, then would sequestration of the polypyrimidine tract in the array 1 reporter cause its transcripts to behave in a similar manner?

To test this possibility, the array 1 intron sequence upstream of the polypyrimidine tract was modified so that it would be present within a strong double-stranded region (Figures 3.6 A and B). The cell line expressing this construct exhibited an increased number of unspliced pre-mRNAs in the nucleoplasm compared to the parent construct (Figure 3.6 C and D). The uncoupling of transcription and splicing likely arises because splicing factor U2AF has reduced or slower access to the polypyrimidine tract. This hypothesis suggests that other means of reducing the U2AF polypyrimidine interaction may produce the same effect. This was tested by weakening the polypyrimidine tract by converting two pyrimidine residues into purines (Figure 3.6 B). This perturbation resulted in the release of unspliced pre-mRNA molecules into the nucleoplasm (Figure 3.6 C and D).
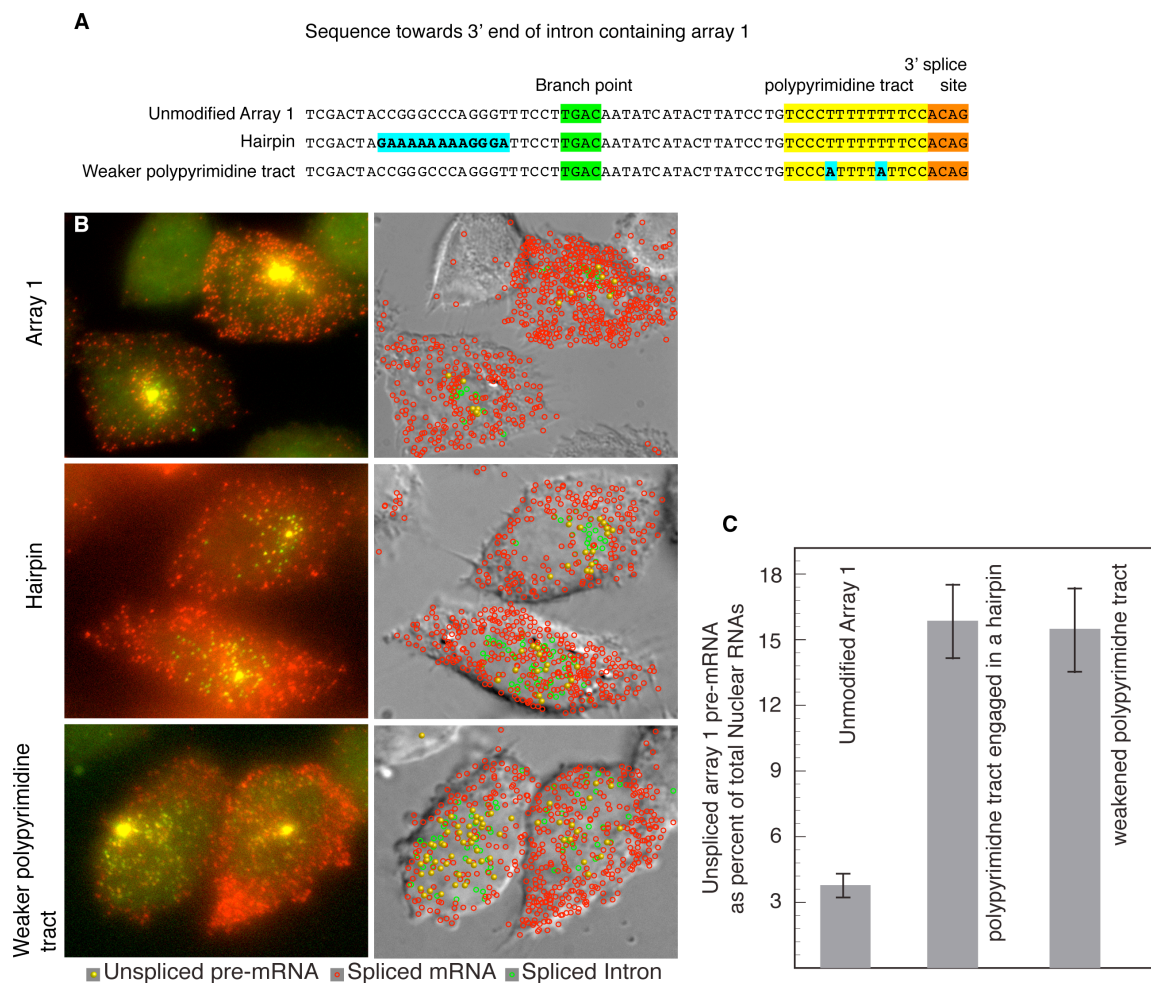
Figure 3.6 Sequestration of the polypyrimidine tract of array 1 leads to an increase in the number of unspliced pre-mRNAs in the nucleus. A. The sequence of the 3' region of the intron in the array 1 construct and the sequence modifications (highlighted in blue) that were introduced. B. Merged z-stacks in a composite image are shown, with green representing the intron signal, red representing the exon signal, and yellow representing the pre-mRNA molecules for each cell line. C. Percentage of pre-mRNAs in the nucleus for three different array 1 construct: unmodified array 1; array 1 in which an upstream sequence forms a secondary structure with the polypyrimidine tract; and array 1 in which the polypyrimidine tract is modified.

3.2.6 Regulated Splicing in *Sxl* pre-mRNAs Occurs Post-Transcriptionally

Drosophila gene *Sxl* controls sex determination by regulating the splicing of several pre-mRNAs, including its own. In males, where *Sxl* is off and there is no Sxl protein, pre-mRNAs are spliced in the default pattern to include a translation-terminating male-specific exon, exon-3 (Figure 3.7 A). In females, Sxl protein binds to multiple

polypyrimidine tracts in introns 2 and 3, forces the splicing machinery to skip exon-3, thereby linking exon 2 directly to exon 4 (*119, 125, 127*) (Figure 3.7 A). Translation of the resulting mRNAs into Sxl protein establishes a positive feedback loop that serves to maintain female identity. To examine the coordination between transcription and splicing, the *Drosophila* male and female cell lines were used that have been used to study the mechanisms of *Sxl*-dependent splicing regulation (*127, 128*).

Initially the coupling of transcription and splicing of the first *Sxl* intron was examined, using sets of distinctly labeled fluorescent probes for this intron and for downstream exon 8. This intron is spliced in the same pattern in both sexes. When the number of pre-mRNAs and spliced products were counted from 50 randomly selected cells and a histogram of their distribution was plotted, there was, on an average, about one molecule of pre-mRNA containing intron 1 and exon 8 in both male (1.12 ±0.09) and female (1.27±0.09) cells (Figure 3.7 C). This suggests that this constitutively spliced intron is generally processed co-transcriptionally in both sexes.

Next, the splicing pattern of the regulated intron 2- exon 3-intron 3 cassette in both male and female cells was examined using distinctly labeled probes specific to intron 2, intron 3, and exon 8. As observed for constitutively spliced intron 1, only about one pre-mRNA molecule per cell containing the intron 2, intron 3, and exon 8 sequences was detected in nuclei from the male cell line (Figure 3.7 B and C). Thus, in spite of the fact that the splice sites of the male exon are sub-optimal (*126-128*), the default splicing machinery joins the regulated cassette exons 2, 3, and 4 together co-transcriptionally. Strikingly, a quite different result is obtained with female cells. Pre-mRNAs containing intron 2, intron 3, and exon 8 sequences were dispersed throughout the nuclei of female cells (Figure 3.7 B). On average, there were about 3 pre-mRNA molecules containing

both of the introns and exon 8 in female nuclei, with many nuclei having 5-7 molecules

of these incompletely spliced pre-mRNAs (Figure 3.7 C). Thus, unlike the processing of

constitutively spliced intron 1, which is co-transcriptional, the splicing of the regulated

Sxl intron 2-exon 3-intron 3 cassette is uncoupled from transcription in female cells.
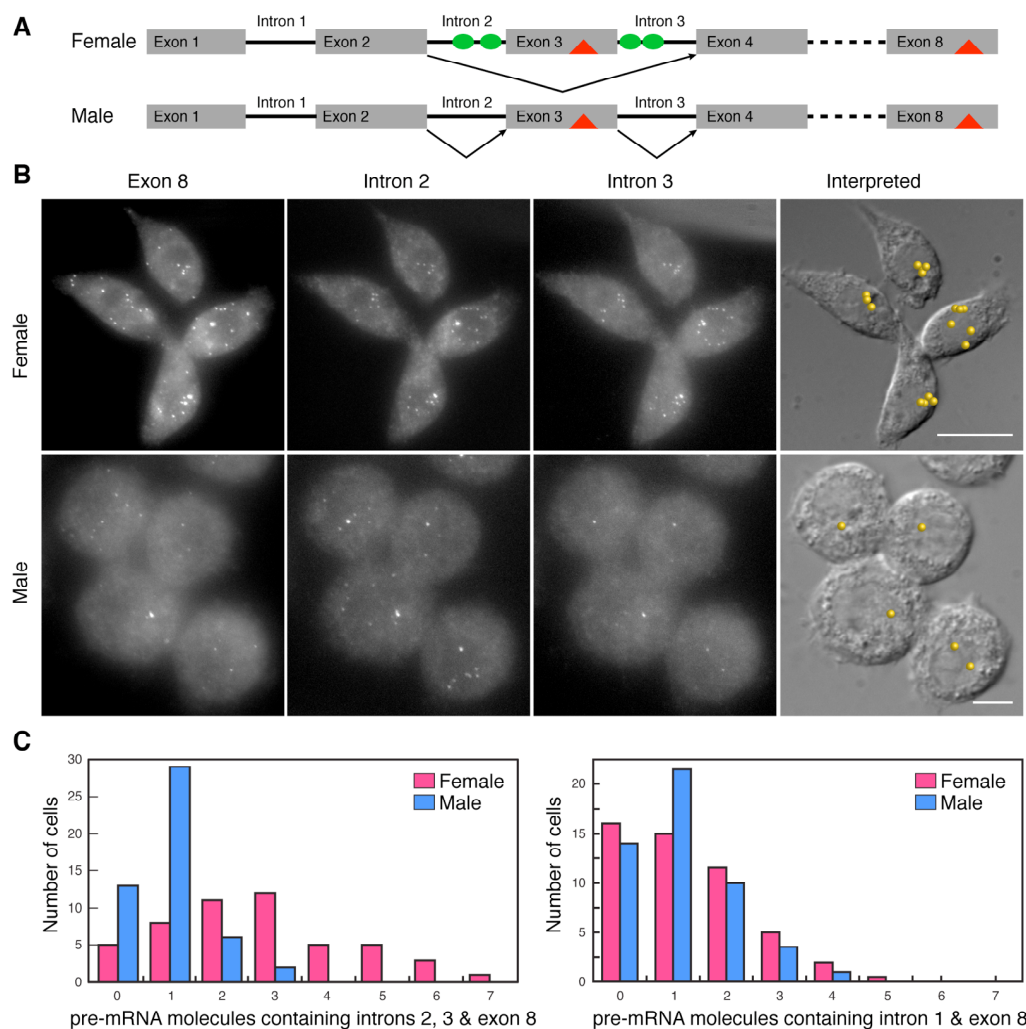
Figure 3.7 Alternative splicing that skips Exon-3 of *Drosophila* Sex-Lethal Pre-mRNA in female cells occurs post-transcriptionally A. Sxl protein (green ovals) binds to polypyrimidine tracts of introns 2 and 3, there by preventing inclusion of exon 3, which contains a stop codon (red triangle). Males do not have Sxl protein, and constitutive splicing yields a transcript containing exon 3, generating a truncated version of non-functional protein. B. Images of spots produced by probes against the indicated components of Sxl transcripts in female and male cells. The locations of the identified molecules are shown on DIC images. C. Histogram showing frequency distribution with which the given numbers of pre-mRNA molecules were found in a group of 50 cells**.**

Furthermore, to show that the pre-mRNAs that are dispersed into the nucleoplasm in female cells are substrates for splicing, female cells were incubated with actinomycin-D (a transcription inhibitor) for various time periods to monitor the fate of previously synthesized pre-mRNAs. As seen from the results shown in Table 3.2, the number of pre-mRNAs was reduced within five mins of exposure to actinomycin-D and the number of spliced mRNAs increased. However, due to rapid mRNA degradation and to there being no new mRNA synthesis, the number of mature mRNAs decreased after 30 minutes exposure to actinomycin-D.

Table 3.2 Mean Number of pre-mRNAs and Spliced Products with 95% Confidence Interval in Individual Female Cells after Treatment with Actinomycin-D

Female cells	Spliced	Introns		Unspliced
	Exon 8	Intron 2	Intron 3	Exon 8-Intron 2-Intron 3
No Actinomycin-D	14.68±2.88	0.32a±0.21	0.16±0.14	2.52±1.73
5mins AMD	14.48±3.20	0.32±0.24	0.08±0.10	0.24±0.17
10mins AMD	15.76±6.5	0	0.28±0.24	0
15mins AMD	9.18±4.07	0.47±0.18	0.28±0.47	0
30mins AMD	8.78±1.56	0.06±0.08	0.09±0.10	0

3.3   Discussion

Using single-molecule imaging, the splicing pattern of constitutively spliced introns of both artificial and natural genes was examined and was shown to complete prior to transcription termination. This supports the idea that co-transcriptional splicing is the default mechanism. One likely mechanism for coupling transcription and splicing is suggested by recent studies in yeast. These studies showed that RNA polymerase pauses

at regular interval during elongation at each 3' splice-site, which functions as a check point to ensure splicing is completed before polyadenylation (*36, 114, 115*). Similar checkpoints are likely to exist in higher eukaryotes (*27, 113*), and might account for the co-transcriptional splicing of constitutively spliced introns that we observed.

However, there might be certain situations wherein these checkpoints might have been escaped and the transcripts might get released from the gene locus before the splicing of certain introns takes place. One of these situations might be in artificial introns that have functionally impaired splice signals. For example, when polypytimidine tract of array 1 was sequestered in a secondary structure or was weakened, the splicing was delayed until after transcription is completed, and large number of unprocessed pre-mRNAs accumulated in the nucleoplasm. In this situation, the signals that normally trigger pausing might not be properly activated, and instead of pausing, the polymerase might transcribe through the termination signals and release incompletely processed transcripts. Once a functional complex is assembled on the defective 3' splice site, the remaining processing steps should proceed unimpeded. In these instances, functionally compromised splicing signals are, by themselves, sufficient to uncouple splicing from transcription.

The other circumstances in which splicing is uncoupled from transcription occur during the alternative splicing of *Sxl* pre-mRNAs. However, the uncoupling seen in this regulated event is different from that observed when the 3' splice site is functionally compromised. When the alternatively spliced Sxl cassette is processed in the default pattern, as occurs in male flies, splicing is co-transcriptional, just like the constitutively spliced introns in the same transcript. Thus, even though the splicing signals in the regulated Sxl cassette are suboptimal, this is not sufficient to uncouple default splicing

and transcription in males. A plausible explanation is that these sub-optimal sites differ from functionally compromised signals, in that they are capable of directing the association of the needed splicing factors on at least a subset of the regulated splice sites while the mRNA is being actively transcribed. This allows them to signal to the polymerase to pause until the default splicing of the regulated cassette is complete. On the other hand, in female flies, Sxl bound at the regulated cassette is somehow able to disrupt this signaling and cause the release of the partially processed transcript.

One of the key principles that emerged from this study is that when transcription and splicing are uncoupled, uncoupling is restricted to the affected intron, and the preceding and succeeding introns continue to be removed co-transcriptionally. We found that no matter how introns containing array 1 and array 2 sequences are arranged within our GFP splicing reporter, splicing of the array 1 intron is co-transcriptional, while splicing of the array 2 intron remains post-transcriptional. Likewise, the processing of intron 1 and the regulated cassette in *Sxl* pre-mRNA were independent of each other. This indicates that spliceosomes assemble at each intron independently of the surrounding introns, and they catalyze the splicing reaction of each intron with its own unique kinetics.

3.4   Experimental Methods

3.4.1 Cell Cultures and Gene Expression

1. Chinese hamster ovary tet-off cells: CHO tet-off Cells were cultured in modified Eagle's minimal essential medium (Sigma, St. Louis, MO) supplemented with Tet-system-approved 10 % fetal bovine serum (Clontech, Mountain View, CA) at 37 ^{0}C. The expression of tetracycline-controlled genes was turned off by the inclusion of 10

ng/ml doxycycline in the culture medium, and turned on by replacing this medium with a medium that did not contain doxycycline.

2. Drosophila cell-lines: Schnieder cells for male and KC167 cells for female were bought from the *Drosophila* genomic resource center (Bloomington, IN) and were grown in M3+BPYE media at 25 ^{0}C without carbon dioxide. The experiments with actinomycin-D were performed by treating the KC167 cells with media containing 100 ng/ml actinomycin-D.

3.4.2 Cloning

1) Sequences of the introns and the 3'-UTR in GFP reporters

Introns and 3'-UTR sequences were inserted into the coding sequence of GFP in plasmid pTRE-d2EGFP (Clontech). The sites chosen for the insertion of the introns were after the first codon of GFP (referred to as the "5' site", introduced into position 447 in the plasmid sequence (GenBank accession number CQ871827.1) and after the 95th codon (referred to as the "middle site," position 729 in the plasmid sequence).

The 5'-site intron sequence, along with the surrounding coding sequences, were:

ATGCAGGTAAGTGGTTAG(array)$_{32}$TCGACTACCGGGCCCAGGGTTTCCTTGAC AATATCATACTTATCCTGTCCCTTTTTTTTTCCACAGGTGAGCAAG. The middle site intron sequence, along with the surrounding coding sequences was: GTCCAGGTAAGTGGTTAG(array)$_{32\ or\ 96}$TCGACTACCGGGCCCAGGGTTTCCTTG ACAATATCATACTTATCCTGTCCCTTTTTTTTTCCACAGGAGCGC.

The identity of the highlighted sequences is as follows: (i) GFP-coding sequences, blue; (ii) intronic portions of the splice donor, branch point, and splice acceptor sites, green; (iii) putative pyrimidine-rich sequence, yellow; (iv) array of probe-

target repeats, purple; and (v) spacer containing restriction endonuclease recognition sites, clear.

The sequences of the repeats in the arrays are: array 1, TCGACCGATCGTGGCCTAAGGAGTTTATATGGAAACCCTTACCAGCCGC; array 2, TCGACAGGAGTTGTGTTTGTGGACGAAGAGCACCAGCCAGCTGATCGACC; and array 3 (in the 3'UTR), TCGACGCGGAGACCACGCTCGGCTTGTCTTTCGCGC CATGCGACGCACGCGGATAGTTAGCTGCGGCGACGAGGCACC.

2) Cloning of introns and 3' UTR with Repeats

Our overall strategy was to first insert a host sequence that contained the canonical splice sites, along with a set of unique restriction endonuclease recognition sites within the GFP-coding sequence, and then insert the tandemly repeated probe-target sequences cloned into separate plasmids within the host sequence, utilizing these restriction sites.

To construct the plasmids that served as sources for the tandem repeats, the method of Robinett et al. was followed (*129*). A synthetic double-stranded oligonucleotide containing the repeated sequence possessing phosphorylated sticky ends and restriction recognition sites for endonucleases SalI and XhoI, was cloned into plasmid pGEM-11Zf(+) (Promega, Madison, WI), which possesses a polylinker containing restriction sites for SalI, XhoI, and BamHI, in that order. First the SalI-BamHI fragment containing the inserted sequence was isolated from the recombinant plasmid. In a separate step, the BamHI-XhoI fragment containing the inserted sequence and the rest of the plasmid was isolated. Since the sticky ends created by SalI and XhoI are compatible, the two isolated fragments were ligated to each other, creating a plasmid containing two head-to-tail tandemly repeated 50-nucleotide inserts. Since the union of a SalI sticky end with an XhoI sticky end generates a sequence that cannot be cut by either

restriction endonuclease, the process of isolating a small fragment and a large fragment from the same plasmid and then ligating these fragments to each other to create a new plasmid containing twice the number of tandem repeats was carried out five times, eventually resulting in a plasmid containing the desired 32 tandem repeats. Array 1 was moved from pGEM-11Zf(+) into pTRE2Hyg-YFP (*8*), and the resulting plasmid served as the source of repeats (this step was omitted in the case of array 2). The repeats for array 3 were cloned into pSV2-DHFR (American Type Culture Collection, Manassas, VA), using the same strategy.

The first step in the construction of the introns was to insert a host sequence that contained sites for restriction enzymes Sal1 and Apa1, surrounded by the splicing donor and acceptor sequences at the two sites in the GFP-coding sequence (discussed in the previous section). The resulting plasmids were used to insert the repeated sequence at the Sal1 site. In order to introduce the host sequence at the 5' site in GFP, the GFP-coding sequence was amplified using tailed primers TCCCCGCGGATGCAGGTAAGTGG TTAGTCGACTACCGGGCCCAGGGTTTCCTTGACAATATCATACTTATCCTGTC CCTTTTTTTTCCACAGGTGAGCAAGGGCGAG and CGAGCTCGAATTCCTACAC ATTGAT.

The amplified DNA fragment contained sites for restriction enzymes SacII and EcoRI near its 5'- and 3'-termini. It was digested with SacII and EcoRI and used to replace GFP from pTRE-d2EGFP. Insertion of the host sequence in the middle position was accomplished by amplifying GFP in two parts with tailed overlapping primers, and then joining the two parts together. The primers used for the 5' fragment were: TCCCCGCGGATGGTGAGCAAG and CTGTGGAAAAAAAAGGGACAGGATAAG

TATGATATTGTCAAGGAAACCCTGGGCCCGGTAGTCGACTAACCACTTACCG GACGTAGCCTTCGGGCA, and the primers used for the 3' fragment were: CAGGTAAGTGGTTAGTCGACTACCGGGCCCAGGGTTTCCTTGACAATATCACT TATCCTGTCCCTTTTTTTTCCACAGGAGCGCACCATCTTC and CGAGCTCGAAT TCCTACACATTGAT.

After separately amplifying the two fragments, they were purified by gel electrophoresis, mixed together, and re-amplified using the outer primers. The joined amplified product was digested with SacII and EcoRI and used to replace GFP pTRE-d2EGFP. The source plasmids were digested with SalI and XhoI, and the fragment containing the repeat was inserted into host plasmids that were digested with SalI (within the intron sequence). This resulted in plasmids pTRE-GFP-array (1 or 2) at the 5' site and pTRE-GFP-array (1 or 2) at the middle site. In order to create plasmids with both introns, we digested them with restriction enzyme BtgZI that cleaves these plasmids in two places: once within GFP, between the two introns, and again within the ampicillin gene. The two fragments were then cross ligated, resulting in plasmids having both introns within the GFP pre-mRNA.

The addition of array 3 at the 3' end of the GFP-coding sequence was accomplished by first removing array 3 from pSV2DHFR-array 3 using SalI-BamHI, and then transferring the array into pTRE-d2EGFP digested with XhoI-BamHI. The resulting plasmid was then digested with ScaI and XmnI and cloned into pTRE-GFP-array (1 or 2) at the middle site digested with ScaI and EcoRI. The identities of the inserts in all of the plasmids were confirmed by restriction mapping and sequencing.

3. Cloning two introns within GFP coding region

Plasmids pTRE-GFP-array (1 or 2)-5' and pTRE-GFP-array (1 or 2)-middle were used in order to create plasmids with both introns within the GFP-coding region. They were digested with restriction enzyme BtgZI, which cleaves these plasmids in two places: once within GFP, between the two introns, and then within the ampicillin gene. The two fragments were then cross-ligated, resulting in plasmids having both introns within their GFP-coding region. The resultant plasmids were linearized with ScaI and were co-transfected with a pTRE-hygromycin vector. Selection for the integrants was performed by culturing cells in the presence of hygromycin for ten days. Pure individual clones were selected by performing FACS sorting.

4. Inserting the array 1 and array 2 constructs at the same genomic site within CHO cells using the FLIP recombinase system

The FLIP-In System (Invitrogen, Carlsbad, CA) was utilized to integrate the two reporter genes at the same genomic locations within Chinese hamster ovary (CHO) cells. First, the Flippase Recognition Target (FRT)-sites was inserted at different locations within the genome of CHO cells by transfection with a pFRT/lacZeo vector (Invitrogen) that was linearized by digestion with ScaI. Five independent Zeocin-resistant clones were selected by culturing the cells for ten days in the presence of 250 ng/ml Zeocin (Invitrogen). Fragments corresponding to GFP-array 1 and GFP-array 2, that lacked the TRE region and the 3'-UTR region, were excised from plasmids pTRE-GFP-array (1 and 2) (containing arrays at the middle site) and were separately inserted into a pcDNA5/FRT vector (Invitrogen), resulting in two plasmids pcDNA5/FRT/ GFP-array 1 and pcDNA5/FRT/GFP-array 2. Each of the five CHO cell clones harboring the FRT site was co-transfected with either the linearized array 1 construct or the linearized array 2 construct and pOG44 (Invitrogen). Selection for the integrants was performed by

culturing the cells in the presence of hygromycin for ten days. At least three pairs of sub-

clones for each FRT clone were isolated. Site-specific integration at the FRT site was

confirmed by checking their sensitivity to Zeocin.

    5.   Modification of the array 1 intron sequence to create a secondary structure and to
          alter the polypyrimidine tract

Alteration of the polypyrimidine tract of the array 1 intron was accomplished using the

pTRE-GFP-array 1 in the middle. The sequences were modified as indicated in Figure

5.5 B with the aid of a Quick Change site-directed mutagenesis kit (Stratagene, San

Diego, CA). Clones containing appropriate mutated sequences were confirmed by

sequencing. Furthermore, the clones were linearized by digestion with ScaI and were co-

transfected with a pTRE-hygromycin vector. Selection for the integrants was performed

by culturing cells in the presence of hygromycin for ten days. Pure individual clones

were selected by performing FACS sorting.

3.4.3 Probe Sets

The *in situ* hybridization probes for binding to array 1 and to array 2 contained multiple

fluorescent label moieties attached to internal thymidines (shown as R in the sequences

below). Array 1 : CGGCRGGTAAGGGRTTCCATARAAACTCCTRAGGCCACGA;

Array 2 : RCGAGGTCGARCAGCTGGCTGGRGCTCTTCGRCCACAAACA

      48 or more oligonucleotides, each labeled with a single fluorophore at their 3' end

via an amino group, were used for binding to the GFP-coding sequence and to all natural

intronic and exonic targets. The labels that were used were tetramethylrhodamine

(TMR), Alexa 594, or Cy5. Methods for the attachment of labels, purification of probes,

in situ hybridization conditions, and the preparation of deoxygenated mounting medium

have been described before (*66*)

3.4.4 Fluorescence *in situ* Hybridization

For *in situ* hybridization, cells were attached to thin gelatin-coated cover slips, which were fixed with 4 % formaldehyde, permeabilized with 70 % alcohol and hybridized overnight with the probe sets in 2X SSC supplemented with 10 % formamide. The cover slips were washed and mounted in a special deoxygenated medium that limits photo bleaching, and then imaged in a wide-field microscope.

3.4.5 Imaging

For each image, 10 to 30 optical slices, with 0.2 μm separation between them, were acquired in each fluorescence channel with a 1-second exposure. These z-stacks were analyzed using custom computer programs written in the Matlab programming environment. These programs enhance the stack of images using a Laplacian filter optimized for the size of spots that we expect, permitting users to select a threshold based on a three-dimensional display of intensity in spots, to segment the image based on the provided threshold, and to produce a list of coordinates of the centers of all spots in three dimensions in each channel. The programs can also determine the distances between spots in two or three fluorescence channels, identify co-localized spots based on provided distance limits, draw circles to produce overlays on the raw images, and count the number of spots in a user-defined region.

3.4.6 Statistical analysis

Ten to twenty-five cells for each category of data reported in Figures 5.1 to 5.3, 5.5, and

Table 5.1; and 50 cells for each category of data reported in Figure 5.6 and Table 5.2,

were analyzed. The data bars represent mean values, and the error bars represent a 95 %

confidence interval (CI).

$$CI = t_{(n-1)} \bullet \frac{SD}{\sqrt{n}} \qquad (3.1)$$

Where, t is a value obtained from a table of t statistics corresponding to a 95% confidence

interval, and n is the number of cells or nuclei, and SD is the standard deviation (*130*).

Probabilities (P-values) described in the legend of Figure 5.1 are the probabilities of

obtaining average percentages as large as reported for the pre-mRNAs by random chance

from the total data set. These probabilities were calculated by bootstrapping, using a

custom Matlab program.

A list of sequences that were designed to image indicated mRNA species are as follows.

1. c-Jun

```
TCGTTTCCATCTTTGCAGTC   CGTTGAGGGCATCGTCATAG
CGCTCTCGGACGGGAGGAAC   TGGGGTTACTGTAGCCATAA
GGGTCATGCTCTGTTTCAGG   TCCCCACTGGGTCGGCCAGG
TGGCGCGGAGGTGCGGCTTC   GCGAGGTGAGGAGGTCCGAG
CCAGCTTGAGCAGCCCCACG   TCAGGCGCTCCAGCTCGGGC
TGTGCCCGTTGCTGGACTGG   GGGTGGGGGTCGGCGTGGTG
TCACGTTCTTGGGGCACAGG   CGGCGAAGCCCTCCTGCTCA
CGGCCAGGGCGCGCACGAAG   GCAGCGTGTTCTGGCTGTGC
GCTGCGCCGCCGACGTGACG   CCACCATGCCTGCCCCGTTG
CTGCCACCGAGGCTACCGCG   TGAAGCCGCCGCTGCCGCTG
GCGGCTCGCTGTGCAGGCTG   AGTTGCTGAGGTTTGCGTAG
CGCTGCTCAGCGCGCCTGGG   CGCCGTAGGAGGGCGCCCCG
GCGCGGGAAAGGCCAGGCCG   GCGGCTGCTGCTGCTGCTGG
GCATCTGCTGGGGCAGGTGG   CCTGCAGCCGCGGGTGCTGC
CTGTCTGAGGCTCCTCCTTC   GTGTCTCGCCGGGCATCTCG
CCATGTCGATGGGGGACAGG   CCGCCTTGATCCGCTCCTGG
TGCGGTTCCTCATGCGCTTC   TTTTTCGGCACTTGGAGGCA
GCCGGGCGATTCTCTCCAGC   TCAAGGTTTTCACTTTTTCC
ACGCCAGCTCCGAGTTCTGA   GTTCCCTGAGCATGTTGGCC
CTTTCTGTTTAAGCTGTGCC   ACCCACTGTTAACGTGGTTC
ACTGCTGCGTTAGCATGAGT
```

2. c-Fos

```
TCTTCTAGTTGGTCTGTCTC   TCTGCAAAGCAGACTTCTCA
TTCAGCAGGTTGGCAATCTC   ACTCTAGTTTTTCCTTCTCC
TCGGTGAGCTGCCAGGATGA   AGGTCATCAGGGATCTTGCA
AGACATCTCTTCTGGGAAGC   AGTCAGATCAAGGGAAGCCA
TGAAGGCCTCCTCAGACTCC   AGGGTCATTGAGGAGAGGCA
ACAGGTTCCACTGAGGGCTT   TCCATGCTGCTGATGCTCTT
ATCAAAGGGCTCGGTCTTCA   TGATGCTGGGAACAGGAAGT
TCAGAGCCACTGGGCCTGGA   ATAGGTCCATGTCTGGCACG
TGCTGCATAGAAGGACCCAG   ACTGTGCAGAGGCTCCCAGT
TCTGTGGCCATGGGCCCCAT   TACAGGTGACCACCGGAGTG
TGTAAGCAGTGCAGCTGGGA   TAGGTGAAGACGAAGGAAGA
ACAGCTGGGGAAGGAGTCAG   TCATTGCTGCTGCTGCCCTT
AGCTGAGCGAGTCAGAGGAA   AGTGGCACTTGTGGGTGCCG
TGTAATGCACCAGCTCGGGC   GAAGATGTGTTTCTCCTCTC
GTCTACAGGAACCCTCTAGG   ACAGATAAGGTCCTCCCTAG
ACAGCCTGGTGTGTTTCACG   CTTTCAAGTCCTTGAGGCCC
```

CTTGAGTCCACACATGGATG ATCTCCGGAAGAGGTAAGGA
CACTCCATGCGTTTTGCTAC GTGTCACTGGGAACAATACA
CTAACTACCAGCTCTCTGAA AGGCCTGGCTCAACATGCTA
AGAGAAAAGAGACACAGACC TATGAGAAGACTAAGGAGAA
CCCAATAGATTAGTTAATGC CCAGGTTAATTCCAATAATG
CAATTTGAAAATATCCAGCA GTTAAAATCAGCTGCACTAG
CCAGGAACACAGTAGTTATT CTAATCAGAACACACTATTG
CTTAGTATAATATTGGTCAT CCAGAAAATAAAGTCGTATC

3. Collagenase

TGCAAGGTAAGTGATGGCTT TGGCCTTTGTCTTCTTTCTC
TTGTGTTTCTAGAGTCGCTG TGGACTAAGTCCACATCTTG
CCCATCATTCTTCAGGTTGT TTCTCCGCTTTTCAACTTGC
ATTTTTCAACCACTGGGCCA GCCCAAAGAATTCCTGCATT
TCTGGTTTCCCAGTCACTTT TCATCACCTTCAGGGTTTCA
TAGGTCAGATGTGTTTGCTC GGCAAATCTGGCGTGTAATT
TTTCTCAATGGCATGGTCCA CATTACTCCAGAGTTGGAAG
TTGGTGAATGTCAGAGGTGT ATCATGATGTCTGCTTGACC
ATGATCTCCCTGACAAAG CATCAAAAGGAGAGTTGTCC
ATGAGCAAGATTTCCTCCAG AATGAGCATCCCCTCCAATA
GTCCACCTTTCATCTTCATC CTGCAACACGATGTAAGTTG
AAGAGAATGGCCGAGTTCAT CGATATCAGTAGAATGGGAG
TAGCTAGGGTACATCAAAGC AGCTGAACATCACCACTGAA
TGCCATCAATGTCATCCTGA GAACGTCCATATATGGCTTG
TTACTGTCACACGCTTTTGG CCCGAATCGTAGTTATAGCA
GCGCATGTAGAATCTGTCTT AAATTGAGCTCAACTTCCGG
GTTGTGGCCAGAAAACAGAA AAGCAGCTTCAAGCCCATTT
TCATCTCTGTCGGCAAATTC CCCAGTACTTATTCCCTTTG
TAGCACATTCTGTCCCTGAA TGTAGATGTCCTTGGGGTAT
ACAGTTCTAGGGAAGCCAAA AAGAGCAGCATCGATATGCT
CCTGGATCCATAGATCGTTT CAGGAAAGTCATGTGCTATC
GCATCAACTTTGTGGCCAAT GCTATTAGCTTTCTGGAGAG
CTCACACCATGTGTTTTCCA CAGTTCTTCAGGAAAACACC
GTATCAGTGACTCTAGAGGT CCCACCATTTGTGGAACTAA

4. Cyclooxygenase-2

GAACAGCAAGGATTTGCTGT CTCATACATACACCTCGGTT
GTGTTGAGCAGTTTTCTCCA GTTTGGAGTGGGTTTCAGAA
ATCCCTTGAAGTGGGTAAGT TTGCATTTCGAAGGAAGGGA
GTGATCTGGATGTCAACACA TGTAAGTTGGTGGACTGTCA
GAGGAAGGGCTCTAGTATAA CAGGAAGCTGCTTTTTACCT
GCCAGAGTTTCACCGTAAAT ATCCTTGAAAAGGCGCAGTT
GAGGATACATCTCTCCATCA TCATCTCTGCCTGAGTATCT
TCCTGTTTAAGCACATCGCA CCACTCAAGTGTTGCACATA
GGGTGTTAAATTCAGCAGCA GGTAATTCCATGTTCCAGCA
GCCTGGTGAATGATTCAACA CAATGGAAGCCTGTGATACT
GCGTTTGCGGTACTCATTAA CAACTCTGCAGACATTTCCT

CTTCTACCATGGTTTCACCA CAGTAGGCAGGAGAACATAT
AAAACCCACTTCTCCACCAA CTTGCATTGATGGTGACTGT
CAGTTCAGTCGAACGTTCTT ATCGCACTTATACTGGTCAA
CTTCCCAGCTTTTGTAGCCA TCAGGGATGAACTTTCTTCT
AGTGCTGGGCAAAGAATGCA CGCTTATGATCTGTCTTGAA
AACAACTGCTCATCACCCCA GTTCTGGGTCAAATTTCAGT
TTTCTACCAGAAGGGCAGGA CATAAGTCCTTTCAAGGAGA
ATTGCAGATGAGAGACTGAA CTGGAACACTGAATGAAGTA
GCAATTTTTCCACAATCTCA ATAGTCTCCTATCAGTAT
GTACTGGAATTGTTTGTTGA GTTGTATTTCTGGTCATGAA
CCTGTAAGTTCTTCAAATGA CCACAGCATCGATGTCACCA
TCAGACCAGGCACCAGACCA GTTGTGTTCCCGCAGCCAGA
ATGGCCCAGCCCGTTGGTGA GCAGCAGGGCGCGGGCGAGC

5. Green fluorescent protein

TCGCCCTTGCTCACCAT ACCCCGGTGAACAGCTC
TCGACCAGGATGGGCAC TTACGTCGCCGTCCAGC
GCTGAACTTGTGGCCGT TCGCCCTCGCCGGACAC
CGTAGGTGGCATCGCCC CTTCAGGGTCAGCTTGC
CCGGTGGTGCAGATGAA AGGGCACGGGCAGCTTG
AGGGTGGTCACGAGGGT ACTGCACGCCGTAGGTC
TCGGGGTAGCGGCTGAA CGTGCTGCTTCATGTGG
GGCGGACTTGAAGAAGT ACGTAGCCTTCGGGCAT
AGATGGTGCGCTCCTGG GCCGTCGTCCTTGAAGA
GCGCGGGTCTTGTAGTT CCTCGAACTTCACCTCG
GTTCACCAGGGTGTCGC CCCTTCAGCTCGATGCG
CCTCCTTGAAGTCGATG CCCCAGGATGTTGCCGT
TTGTACTCCAGCTTGTG CGTTGTGGCTGTTGTAG
GTCGGCCATGATATAGA ATGCCGTTCTTCTGCTT
TCTTGAAGTTCACCTTG CTCGATGTTGTGGCGGA
AGCTGCACGCTGCCGTC TGCTGGTAGTGGTCGGC
TCGCCGATGGGGGTGTT TTGTCGGGCAGCAGCAC
GGGTGCTCAGGTAGTGG TTTGCTCAGGGCGGACT
CGCTTCTCGTTGGGGTC GCAGGACCATGTGATCG
GGCGGTCACGAACTCCA ATGCCGAGAGTGATCCC
TCTTGTACAGCTCGTCC GAAGCCATGGCTAAGCT
TCCTGCTCCTCCACCTC TGGGCAGCGTGCCATCA
CTCCTGGGCACAAGACA TGACGGTCCATCCCGCT
AGAAGCACAGGCTGCAG TACACATTGATCCTAGC

6. Sex-lethal intron-1

GTGTTGTTGTCTTTTTCGCC ATTTTCGGCCCTTCACAACT
CACTCAGGTAAAGCGAAATC TGAAAAGGGACACGCGATA
TTTGCAGCGGAACTAAAGGA GCATAGTTGGTGAAACAGGA
ATGCGCAGCAAGAAACACAT GTCGACCACCAAAAAATAGG
ATGATAACAGCTGGCAATGC AGTGTACGCCAAAACACGTT
AACCAGGGACACATGTGTAA GCTCTGGGAAACTCCAAAAT

GTGTCGCTAATGAGGATCTT TCTTCAATGGAAAAGGGAGG
TGGAAATATCCAGGATCCCT ACGATAGAGACAACAACGGA
AAAAAATACGGCAGCTGAGC AGCCAGCTGCAACTTGAAAA
AATAAATGCAGCTGCCCTCT AGCGGAGATGAAGAGAAAAG
AACAGCAAGAAGGGGAATGA TTTAAGCGAAGGGAACGATC
CCGCCATTTCCAATATTTCC TTAAGAGTGTTGGTGGAGGT
GACCTCAATGGAATGCAGAA TAACGTGCAATTTGCGCAAG
AGGTTTAAAACGGGGTGCAA CCACTGTTGGGAAGCTTTTA
CGACGCTAAGTAGTTGCATT CATTGCGGAACCGAAACATA
CCAAATGCGCATGTATGTAC CGTCCGCTTGGATTGTTATT
ATCCATCGAACATTCCAGCT GCATTGCCAGACACAATCAA
CGGTCTGGAAACAAAATCAC ATGTACCCGCATGCATGAAT
AATCGCTCAGCTGGGAAAAT CATGTACACATCTGTGCATC
GCCCATTTAAGTACATGAGC GCTTGAGTGAATGTGTGTTC
GCTAATTTGGAGCACAGTTG GCCAGCGAATTGTAAAGCAA
GAGGGAAAGAGATGGAGAAA CAGACATTTTTCGGCAGTGT
GCATATCACAATGCGTGGTA CGGACAAAACTTTTGGGGAA
GGGTGAAAGCTAAATCACCA CCCATAGACTTTCCATATGC

7. Sex lethal intron-2

CGAAACGTGAGAACTCAAGT GGATTGTCATTCGAATGGGT
AAGGAATGGGTGGAAAGGAA CAAAGAGGTATGGGTAGCAA
GGTTGGTCTTTTGTTCGTTC GACTGCAACTATTTTCCAGC
GCCGACAAGTTTTGCTAGTA CTGCACAAATACCAATGCAC
GACTGTCAATCGACATCCTT GACATCGACAGCAAATCGAT
GGGGAGAACATCACTCAATA CTCTGATAATATCAGAGCGC
CTGTGGCTGTTCTTGTTGTT GCTGCGGTTATTGTTATCCA
CACCATCTAGCACATGTAAC CTGCTTTTGGTTCGCTATCA
CCCCTTAAAACGATTGACTG AATTTGCAATTGGGAGCGGA
GTGGATTGCTTGGAACGAAA GGGGATTCCTATGTAATGCA
CGTAACAACTAGATCGAACG GCGCATGAATGTATGCTATG
TGTGCGACTGTGGGATAAAT CGATCAGCTGATTCAAAGTG
GATTACGAAAACACGCAGAC GTTGTTTCACGTTGATCGGT
CGTTCGGTCTTTGTAAATGC ATCAAAACGCTTCTGATGGC
ATCAAGTTTCGATTCCCAGC CCTTGTTCTACGCGTGTATA
CCCAGATACGGATACATGTA GGCCGTTTCTATAATCCTCT
TTTAGGTTGCACAATCCGCA TTACCGGTACATTGTGAGAC
GCAGTAAGTGTGGAAGAACA CGAGGCGAGAAAGAAAAAC
GGCAAAACTTAAGGCTTACG CTTGACCCCGCTTTTCATTT
CCTCCATTTGTTTTATGCCC GAATGACGCTGCTTTTGCTT
GGAAAGAAGAAGCAGCTCAT GCCATAAATTATACGACCCC
GGCTGCTCTGTTTACATTTC ACTAAATTTGGAGAGGGTGC
TTTTTCGCACAGCACGTGAT TAAGGCTACTGCATGTCAAG
GGAAAAAGGACAACCATACG GTAAAATCTCCCAACACGTG

8. Sex-lethal intron-3

```
CTTAGAGTCTTGTTACTTAC GTTTTTGCCTTTGGCTTTAG
CATCATCATATTATTATATA ACGAGCTGCTTCCCCAATAT
GTTGATTTTTATAGTATTTT TTTAGGTCTAAGTTAGATCT
GCATATCATATTCGGTTCAT TCTTCCCACGTCGAATTTTG
TTATAGTTTCGGACATCGCC ACAAAAAGGATTTGGGGACT
TTGGCACTTTTTCATCACAT AAAATCAAAAAAATAATCAC
CTGGAGCACATTTTCACTTT ACCCATATCGGACACTTTGT
TAATCATGGACTATACTAG CACCGAAAAAAATAAAAAA
TTTTTTTTTTGTACTTTCGA CTTAAGAAAGCATGATGTA
GGTTTCACTTTAAATATTGG TTGCCTTAAGGTGAAAACAA
GTGAAATTCTGCAAACCTC GATCCCCAGTTATATTCAA
AACTTAGACTGACCCTCAAA TTTGCAGTTTCTCGACGAAT
TGAGAAATATTGATGTGACG CCATTAGTGGATTTTGAAAA
AGGCTTAATTTAAATGAGAA TGTGCAAATTAGCTTAAGAC
CAGAGTACAGTAAACTGTCT AATGAGTTTTGAAAACTTGC
CTACTAACTTAAACTTATGA TGATTTAACAATACTTTTCT
TAAGAAATGTTTTGGCGCTG TTGAAAATACTTTAAAAATG
TTACTTATATTATTTAGCCA AATTAAATGTATAAAGCGCA
AAATACTATTTAAAATTATC TTGTCGGTTATTGGTTAAAG
AGATATCATAGAAATGATTG TGGGGAAGAGAAAATATGAA
CGATGAATCGATTCCATTTC
```

9. Sex-lethal Exon-8+UTR

```
CCTCCTCACGCTTGTTGTAC CCTCCTCACGCTTGTTGTAC
TTGCCATGCTCCTCAGCCAA ATCTGCGACATAAAGTGGGC
TGGTACATTTGCTGGAACCA TATGTGCTGGCGGTTGTGGT
TGTGCATCATGTTGAATGCG GGAAGCGTTGCTGTGATTTA
TCGAAATAGGGATGCGAGTT CATCTTGTATTCGTTTCTGG
TTTGTGTGTGGCTTGTGTGT CAGCAGATCGCTCTTGATTA
CACTCTCAAGATTAATTGGT GAATCAAGGGTTTCAGATGT
GTTGTTGTTTGTGAAATACC TTTCGTAGCCTGCTGTTTTG
GTTTGTTGTTGGTTGGATTG TGTTGTTGGTTGTTAGTGTT
CGAATTACCGAATTAAGAGC GCACTGCGAAATTGCAACAA
CGTGATATTTCCCTTTCTCT CAAAGTTGTGAAAGAAAAGC
GTGTGTAGTTGTGAGTTCTG GCCTGGTTTTGTGTATGTGT
TGAAGTGCTAGTTTACCGTT ACTAACACTTTTACGCGAAC
CGCTTTGGAAAACACACACA GTGCTGTATCCGGGAAAAAA
TGCAGCTTTACGTTTTGTTA CTCATTATGTGCATTTCGGT
GTTGGTTTATCTCTTTGTTG GTTAGGTATCGATTCAGTCC
GTGGGTTATGCATTTCTTGC GCACTCTTTAGCTTCGTTCA
GTCTTTTGTGTAAACGGGAA AATCTTCTGTTGTCTCGCAC
CTGATTGCTTGATTGCAATG CTCGGCAACGCTTTTGTAAT
GGATCTAAAAGGGTTACAGG CTGAAGATGCTGAAGAGGTA
GGCGTGATTCGATTTCAATA GAGATGGGCAAATTAAAACC
ATTGTAGTTGTTGTCGCTGG CATGTTGTTTATCTTGCAGT
GAATAGGTTCCATATAGTGA GCTTTTAACGCACTACAAAC
CACACGTTCGGGATTTTGTT GTTTGTTTTGCTTTTCGCC
```

APPENDIX B

MATLAB CODE FOR IMAGE ANALYSIS

All the image analysis in the current study was performed using following Matlab code.

1. Image analysis program to count individual mRNA particle from two different channels

```matlab
%This program counts individual mRNA molecules from two different
channels
XXX = input('give full name of image file 1    ', 's');
YYY = input('give full name of image file 2    ', 's');

ims = readfile(XXX);
imsd = medianfilter(double(ims));
lapims = laplace(imsd);
lapims = lapims/max(lapims(:));
im1 = max(lapims,[],3);
fprintf('Draw rectangle with mouse for thresholding              %g\n');

figure(1);
imshow(im1);
R = getrect;
test = imcrop(im1,R);

figure(2);
surf(test);
threshx = input('Threshold? ');
L = sliceall(lapims/10,threshx/10);

ims = readfile(YYY);
imsd = medianfilter(double(ims));
lapims = laplace(imsd);
lapims = lapims/max(lapims(:));
im2 = max(lapims,[],3);
fprintf('Draw rectangle with mouse for thresholding              %g\n');

figure(3);
imshow(im2);
R = getrect;
test = imcrop(im2,R);
figure(4);
surf(test);
threshy = input('Threshold? ');
K = sliceall(lapims/10,threshy/10);
clear imsd
clear lapims

[lab1,n] = bwlabeln(L);
s = regionprops(lab1,'Area');
areas = [s.Area];
```

```
bw1 = ismember(lab1,find(areas > 10));
[lab1,n] = bwlabeln(bw1);
s1 = regionprops(lab1,'Centroid');
centers1 = cat(1,s1.Centroid);

[lab2,n] = bwlabeln(K);
s = regionprops(lab2,'Area');
areas = [s.Area];


bw2 = ismember(lab2,find(areas > 10));
[lab2,n] = bwlabeln(bw2);
s1 = regionprops(lab2,'Centroid');
centers2 = cat(1,s1.Centroid);

im1=imadjust(im1,[0.02 .1]);
im2=imadjust(im2,[0.02 .1]);

red = im1;
green = im2;
blue = zeros(size(im1));
RGB= cat(3,red,green,blue);

figure(7);
hold off;
imshow(RGB)
hold on;
plot(centers1(:,1),centers1(:,2),'ro','markersize',10);
plot(centers2(:,1),centers2(:,2),'go','markersize',10);
fprintf('Choose the cell...\n');

[nuclearpoints,x,y] = selectdata('selectionmode','Lasso');
particles_cell1=nuclearpoints{2};
particles_cell2=nuclearpoints{1};
```

2. Image analysis program to count individual protein hetrodimers (both from nucleus and cytoplasm) and down stream RNA particles

```
%This program is used to count PLA signals from nucleus and cytoplasm
along with counting mRNAs of two genes. The in put file should
containing z-stacks from all three channels in tiff format. The three
channels should be in following order. Channel 1: PLA, channel 2:RNA-1,
channel 3: RNA-2.

XXX = readfile('file_name.tiff');
YYY = input('give full name of DIC image file 3    ', 's');
im4=imread(YYY);
ZZZ = input('give full name of DAPI image file 3    ', 's');
im5=imread(ZZZ);

[m,n,p]=size(XXX);
j=1;
for i=1:3:p
ims(:,:,j)=XXX(:,:,i);
```

```
j=j+1;
end
imsd = medianfilter(double(ims));
clear ims
lapims = laplace(imsd);
clear imsd
lapims = lapims/max(lapims(:));
im1 = max(lapims,[],3);
fprintf('Draw rectangle with mouse for thresholding        %g\n');
figure(1);
imshow(im1);
R = getrect;
test = imcrop(im1,R);
figure(2);
surf(test);
threshx = input('Threshold? ');
L = sliceall(lapims/10,threshx/10);

[m,n,p]=size(XXX);
j=1;
for i=2:3:p
ims(:,:,j)=XXX(:,:,i);
j=j+1;
end
imsd = medianfilter(double(ims));
clear ims
lapims = laplace(imsd);
clear imsd
lapims = lapims/max(lapims(:));
im2 = max(lapims,[],3);
fprintf('Draw rectangle with mouse for thresholding        %g\n');
figure(1);
imshow(im2);
R = getrect;
test = imcrop(im2,R);
figure(2);
surf(test);
threshy = input('Threshold? ');
M = sliceall(lapims/10,threshy/10);

[m,n,p]=size(XXX);
j=1;
for i=3:3:p
ims(:,:,j)=XXX(:,:,i);
j=j+1;
end
imsd = medianfilter(double(ims));
lapims = laplace(imsd);
lapims = lapims/max(lapims(:));
im3 = max(lapims,[],3);
fprintf('Draw rectangle with mouse for thresholding        %g\n');
figure(1);
imshow(im3);
R = getrect;
test = imcrop(im3,R);
figure(2);
surf(test);
```

```matlab
threshz = input('Threshold? ');
N = sliceall(lapims/10,threshz/10);
clear lapims

[lab1,n] = bwlabeln(L);
s = regionprops(lab1,'Area');
areas = [s.Area];
bw1 = ismember(lab1,find(areas > 10));
[lab1,n] = bwlabeln(bw1);
s1 = regionprops(lab1,'Centroid');
centers1 = cat(1,s1.Centroid);

[lab2,n] = bwlabeln(M);
s = regionprops(lab2,'Area');
areas = [s.Area];
bw2 = ismember(lab2,find(areas > 10));
[lab2,n] = bwlabeln(bw2);
s2 = regionprops(lab2,'Centroid');
centers2 = cat(1,s2.Centroid);

[lab3,n] = bwlabeln(N);
s = regionprops(lab3,'Area');
areas = [s.Area];
bw3 = ismember(lab3,find(areas > 10));
[lab3,n] = bwlabeln(bw3);
s3 = regionprops(lab3,'Centroid');
centers3 = cat(1,s3.Centroid);

im1=imadjust(im1,[0.02 .1]);
im2=imadjust(im2,[0.02 .1]);
im3=imadjust(im3,[0.02 .1]);

red = im1;
green = im2;
blue = im3;
RGB= cat(3,red,green,blue);

figure(7);
hold off;
imshow(im4)
hold on;
plot(centers1(:,1),centers1(:,2),'ro','markersize',8);
plot(centers2(:,1),centers2(:,2),'go','markersize',8);
plot(centers3(:,1),centers3(:,2),'bo','markersize',8);

fprintf('Choose the cell...\n');
[nuclearpoints,x,y] = selectdata('selectionmode','Lasso');
PLA=nuclearpoints{3};RNA2=nuclearpoints{2};RNA1=nuclearpoints{1};

for n=1:3
    a=size(nuclearpoints{n});
    list(n)=a(1);
end
list=cat(2,list,threshx,threshy);

figure(8);
hold off;
```

```
imshow(im5)
hold on;
plot(centers1(:,1),centers1(:,2),'ro','markersize',8);
fprintf('Choose the nucleus...\n');
[cellpoints] = selectdata;
particles_nucleus=cellpoints{1};
particles_cytoplasm=size(PLA,1)-size(particles_nucleus,1);
```

3. Image analysis program to find co-localized and non-colocalized particles between three individual mRNAs

```
XXX = input('give full name of image file 1    ', 's');
YYY = input('give full name of image file 2    ', 's');
ZZZ = input('give full name of image file 3    ', 's');

ims = readfile(XXX);
imsd = medianfilter(double(ims));
lapims = laplace(imsd);
lapims = lapims/max(lapims(:));
im1 = max(lapims,[],3);
fprintf('Draw rectangle with mouse for thresholding           %g\n');
figure(1);
imshow(im1);
R = getrect;
test = imcrop(im1,R);
figure(2);
surf(test);
threshx = input('Threshold? ');
L = sliceall(lapims/10,threshx/10);

ims = readfile(YYY);
imsd = medianfilter(double(ims));
lapims = laplace(imsd);
lapims = lapims/max(lapims(:));
im2 = max(lapims,[],3);
fprintf('Draw rectangle with mouse for thresholding           %g\n');
figure(3);
imshow(im2);
R = getrect;
test = imcrop(im2,R);
figure(4);
surf(test);
threshy = input('Threshold? ');
K = sliceall(lapims/10,threshx/10);

ims = readfile(ZZZ);
imsd = medianfilter(double(ims));
lapims = laplace(imsd);
lapims = lapims/max(lapims(:));
im3 = max(lapims,[],3);
fprintf('Draw rectangle with mouse for thresholding           %g\n');
figure(5);
imshow(im3);
R = getrect;
test = imcrop(im3,R);
figure(6);
```

```
surf(test);
threshz = input('Threshold? ');
M = sliceall(lapims/10,threshz/10);
clear ims
clear imsd

[lab1,n] = bwlabeln(L);
s = regionprops(lab1,'Area');
areas = [s.Area];
bw1 = ismember(lab1,find(areas > 10));
[lab1,n] = bwlabeln(bw1);
s1 = regionprops(lab1,'Centroid');
centers1 = cat(1,s1.Centroid);

[lab2,n] = bwlabeln(K);
s = regionprops(lab2,'Area');
areas = [s.Area];
bw2 = ismember(lab2,find(areas > 10));
[lab2,n] = bwlabeln(bw2);
s1 = regionprops(lab2,'Centroid');
centers2 = cat(1,s1.Centroid);

[lab3,n] = bwlabeln(M);
s = regionprops(lab3,'Area');
areas = [s.Area];
bw3 = ismember(lab3,find(areas > 10));
[lab3,n] = bwlabeln(bw3);
s1 = regionprops(lab3,'Centroid');
centers3 = cat(1,s1.Centroid);

biggest=max(cat(2,size(centers1),size(centers2),size(centers3)));
centers=zeros(biggest,3);
colocab_b=zeros(biggest,3);
colocab_a=zeros(biggest,3);
colocac_c=zeros(biggest,3);
colocac_a=zeros(biggest,3);
colocbc_c=zeros(biggest,3);
colocbc_b=zeros(biggest,3);
colocabc_c=zeros(biggest,3);
colocabc_a=zeros(biggest,3);
colocabc_b=zeros(biggest,3);
colocabc_aba=zeros(biggest,3);
colocabc_abb=zeros(biggest,3);
colocabc=zeros(biggest,3);

[t r]=size(centers);
[q w]=size(centers1);
centersx=zeros((t-q),3);
centers1=cat(1,centers1,centersx);

[t r]=size(centers);
[q w]=size(centers2);
centersx=zeros((t-q),3);
centers2=cat(1,centers2,centersx);

[t r]=size(centers);
[q w]=size(centers3);
```

```matlab
centersx=zeros((t-q),3);
centers3=cat(1,centers3,centersx);

%Between a and b
i=1;
j=1;
for i=1:size(centers1,1);
for j=1:size(centers2,1);
    dist(i,j)=sqrt(((centers2(j,1)-
centers1(i,1))*.065)^2+((centers2(j,2)-
centers1(i,2))*.065)^2+((centers2(j,3)-centers1(i,3))*.2)^2);
if     dist(i,j) <0.5;
    colocab_b(j,:)=centers2(j,:);colocab_a(i,:)=centers1(i,:);
end
end
end

%Between a and c
i=1;
j=1;
for i=1:size(centers1,1);
for j=1:size(centers3,1);
    dist(i,j)=sqrt(((centers3(j,1)-
centers1(i,1))*.065)^2+((centers3(j,2)-
centers1(i,2))*.065)^2+((centers3(j,3)-centers1(i,3))*.2)^2);
if     dist(i,j) <0.5;
    colocac_c(j,:)=centers3(j,:);colocac_a(i,:)=centers1(i,:);
end
end
end

%Between b and c
i=1;
j=1;
for i=1:size(centers2,1);
for j=1:size(centers3,1);
    dist(i,j)=sqrt(((centers3(j,1)-
centers2(i,1))*.065)^2+((centers3(j,2)-
centers2(i,2))*.065)^2+((centers3(j,3)-centers2(i,3))*.2)^2);
if     dist(i,j) < 0.5;
    colocbc_c(j,:)=centers3(j,:);colocbc_b(i,:)=centers2(i,:);
end
end
end

%Between a,b and c with respect to c and ab_a
i=1;
j=1;
for i=1:size(colocab_a,1);
for j=1:size(centers3,1);
    dist(i,j)=sqrt(((centers3(j,1)-
colocab_a(i,1))*.065)^2+((centers3(j,2)-
colocab_a(i,2))*.065)^2+((centers3(j,3)-colocab_a(i,3))*.2)^2);
if     dist(i,j) < 0.5;
    colocabc_aba(j,:)=centers3(j,:);%colocabc_a(i,:)=centers1(i,:);
end
end
```

```matlab
end

%Between a,b and c with respect to c and ab_b
i=1;
j=1;
for i=1:size(colocab_b,1);
for j=1:size(centers3,1);
    dist(i,j)=sqrt(((centers3(j,1)-
colocab_b(i,1))*.065)^2+((centers3(j,2)-
colocab_b(i,2))*.065)^2+((centers3(j,3)-colocab_b(i,3))*.2)^2);
if      dist(i,j) < 0.5;
    colocabc_abb(j,:)=centers3(j,:);%colocabc_b(i,:)=centers2(i,:);
end
end
end

%between colocabc_aba and colocabc_abb-real triple localized
i=1;
j=1;
for i=1:size(colocabc_abb,1);
    for j=1:size(colocabc_aba,1);
        dist(i,j)=sqrt(((colocabc_aba(j,1)-
colocabc_abb(i,1))*.065)^2+((colocabc_aba(j,2)-
colocabc_abb(i,2))*.065)^2+((colocabc_aba(j,3)-
colocabc_abb(i,3))*.2)^2);
if      dist(i,j) < 0.5;
    colocabc(j,:)=colocabc_aba(j,:);%colocabc_b(i,:)=centers2(i,:);
end
end
end

a_all=centers1-colocab_a-colocac_a;
[a b]=size(a_all);
for i=1:a
for j=1:b
    if a_all(i,j)<0
        a_all(i,j)=0;
    end
end
end

b_all=centers2-colocab_b-colocbc_b;
[a b]=size(b_all);
for i=1:a
for j=1:b
    if b_all(i,j)<0
        b_all(i,j)=0;
    end
end
end

c_all=centers3-colocac_c-colocbc_c;
[a b]=size(c_all);
for i=1:a
for j=1:b
    if c_all(i,j)<0
        c_all(i,j)=0;
```

```matlab
    end
end
end

im1=imadjust(im1,[.02 .50]);
im2=imadjust(im2,[.03,.36]);
im3=imadjust(im3, [.01 .59]);

red = im1;
blue = im2;
green = im3;
RGB= cat(3,red,green,blue);

%colocab only
colocab_only=colocab_b;
i=1;
j=1;
for i=1:size(colocabc,1);
for j=1:size(colocab_b,1);
    dist(i,j)=sqrt(((colocab_b(j,1)-
colocabc(i,1))*.065)^2+((colocab_b(j,2)-
colocabc(i,2))*.065)^2+((colocab_b(j,3)-colocabc(i,3))*.2)^2);
if     dist(i,j) < 0.5;
    colocab_only(j,:)=0;%colocabc_b(i,:)=centers2(i,:);
end
end
end

%colocac only
colocc_only=colocac_c;
i=1;
j=1;
for i=1:size(colocabc,1);
for j=1:size(colocac_c,1);
    dist(i,j)=sqrt(((colocac_c(j,1)-
colocabc(i,1))*.065)^2+((colocac_c(j,2)-
colocabc(i,2))*.065)^2+((colocac_c(j,3)-colocabc(i,3))*.2)^2);
if     dist(i,j) < 0.5;
    colocac_only(j,:)=0;
end
end
end

%colocbc only
colocbc_only=colocbc_c;
i=1;
j=1;
for i=1:size(colocabc,1);
for j=1:size(colocbc_c,1);
    dist(i,j)=sqrt(((colocbc_c(j,1)-
colocabc(i,1))*.065)^2+((colocbc_c(j,2)-
colocabc(i,2))*.065)^2+((colocbc_c(j,3)-colocabc(i,3))*.2)^2);
if     dist(i,j) < 0.5;
    colocbc_only(j,:)=0;
end
end
end
```

```matlab
figure(10);
imshow(RGB)
hold on;
plot(colocab_only(:,1),colocab_only(:,2),'mo','markersize',5,'markerfac
ecolor','m');
plot(colocac_only(:,1),colocac_only(:,2),'yo','markersize',5,'markerfac
ecolor','y');
plot(colocbc_only(:,1),colocbc_only(:,2),'co','markersize',5,'markerfac
ecolor','c');
plot(colocabc(:,1),colocabc(:,2),'wo','markersize',5,'markerfacecolor',
'w');
plot(a_all(:,1),a_all(:,2),'ro','markersize',5);
plot(b_all(:,1),b_all(:,2),'bo','markersize',5);
plot(c_all(:,1),c_all(:,2),'go','markersize',5);
print('-depsc', cat(2,ZZZ,'all','.eps'));
fprintf('Choose the nucleus...\n');
[nuclearpoints,x,y] = selectdata('selectionmode','Lasso');

for n=1:7
    a=size(nuclearpoints(56));
    list(n)=a(1);
end
%list appears in reverse order of plotting commands, ie, c, b, a, all,
bc, ac, ba.
```

4. To run above programs, following functions are required to be present in your Matlab directory

```matlab
function theimages = readfile(filename)
   i = 1;
 reading = 1;
  while reading == 1
    try
       %theimages(:,:,i) = imread(filename,i);
       trialim = imread(filename,i);
    catch
       i;
       reading = 0;
    end;
    if ndims(trialim) == 2 & reading == 1
       theimages(:,:,i) = trialim;
    end;

    i = i+1;
 end;

 function outims = medianfilter(images)

 sz = size(images);
 outims = zeros(sz);

 for i = 1:sz(3)
   outims(:,:,i) = medfilt2(images(:,:,i),[3 3]);
 end;
```

REFERENCES

1.   T. Tsukamoto, N. Hashiguchi, S. M. Janicki, T. Tumbar, A. S. Belmont, D. L. Spector, Visualization of gene activity in living cells. *Nat Cell Biol* **2**, 871 (Dec, 2000).

2.   W. J. Blake, K. A. M, C. R. Cantor, J. J. Collins, Noise in eukaryotic gene expression. *Nature* **422**, 633 (Apr 10, 2003).

3.   L. Cai, N. Friedman, X. S. Xie, Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358 (Mar 16, 2006).

4.   J. R. Chubb, T. Trcek, S. M. Shenoy, R. H. Singer, Transcriptional pulsing of a developmental gene. *Curr Biol* **16**, 1018 (May 23, 2006).

5.   I. Golding, J. Paulsson, S. M. Zawilski, E. C. Cox, Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025 (Dec 16, 2005).

6.   M. B. Elowitz, A. J. Levine, E. D. Siggia, P. S. Swain, Stochastic gene expression in a single cell. *Science* **297**, 1183 (Aug 16, 2002).

7.   A. Raj, A. van Oudenaarden, Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216 (Oct 17, 2008).

8.   A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, S. Tyagi, Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**, e309 (Oct, 2006).

9.   A. N. Boettiger, M. Levine, Synchronous and stochastic patterns of gene activation in the Drosophila embryo. *Science* **325**, 471 (Jul 24, 2009).

10.   A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, N. Barkai, Noise in protein expression scales with natural protein abundance. *Nat Genet* **38**, 636 (Jun, 2006).

11.   Z. Waks, A. M. Klein, P. A. Silver, Cell-to-cell variability of alternative RNA splicing. *Mol Syst Biol* **7**, 506 (2011).

12.   N. A. Faustino, T. A. Cooper, Pre-mRNA splicing and human disease. *Genes Dev* **17**, 419 (Feb 15, 2003).

13.   A. J. Lopez, Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* **32**, 279 (1998).

14.   S. Ghaemmaghami, W. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, J. S. Weissman, Global analysis of protein expression in yeast. *Nature* **425**, 737 (Oct 16, 2003).

15.   P. Guptasarma, Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of Escherichia coli? *Bioessays* **17**, 987 (Nov, 1995).

16.   N. Fedoroff, W. Fontana, Genetic networks. Small numbers of big molecules. *Science* **297**, 1129 (Aug 16, 2002).

17.	N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, Gene regulation at the single-cell level. *Science* **307**, 1962 (Mar 25, 2005).

18.	M. S. Ko, H. Nakauchi, N. Takahashi, The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J* **9**, 2835 (Sep, 1990).

19.	J. Yu, J. Xiao, X. Ren, K. Lao, X. S. Xie, Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600 (Mar 17, 2006).

20.	Y. Taniguchi, P. J. Choi, G. W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X. S. Xie, Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533 (Jul 30, 2010).

21.	M. Kaern, T. C. Elston, W. J. Blake, J. J. Collins, Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* **6**, 451 (Jun, 2005).

22.	T. B. Kepler, T. C. Elston, Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys J* **81**, 3116 (Dec, 2001).

23.	J. M. Pedraza, J. Paulsson, Effects of molecular memory and bursting on fluctuations in gene expression. *Science* **319**, 339 (Jan 18, 2008).

24.	A. Becskei, B. B. Kaufmann, A. van Oudenaarden, Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.* **37**, 937 (Sep, 2005).

25.	B. Y. J. Peccoud, Markovian Modeling of Gene-Product Synthesis. *Theoretical Population Biology* **48**, 222 (October 1995).

26.	D. M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, F. Naef, Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**, 472 (Apr 22, 2011).

27.	A. Pare, D. Lemons, D. Kosman, W. Beaver, Y. Freund, W. McGinnis, Visualization of individual Scr mRNAs during Drosophila embryogenesis yields evidence for transcriptional bursting. *Current biology : CB* **19**, 2037 (Dec 15, 2009).

28.	S. Newlands, L. K. Levitt, C. S. Robinson, A. B. Karpf, V. R. Hodgson, R. P. Wade, E. C. Hardeman, Transcription occurs in pulses in muscle fibers. *Genes & development* **12**, 2748 (Sep 1, 1998).

29.	I. L. Ross, C. M. Browne, D. A. Hume, Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol Cell Biol* **72**, 177 (Apr, 1994).

30.	B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Global quantification of mammalian gene expression control. *Nature* **473**, 337 (May 19, 2011).

31.   E. Eden, N. Geva-Zatorsky, I. Issaeva, A. Cohen, E. Dekel, T. Danon, L. Cohen, A. Mayo, U. Alon, Proteome half-life dynamics in living human cells. *Science* **331**, 764 (Feb 11, 2011).

32.   P. Lu, C. Vogel, R. Wang, X. Yao, E. M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25**, 117 (Jan, 2007).

33.   J. R. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, J. S. Weissman, Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840 (Jun 15, 2006).

34.   D. Zenklusen, D. R. Larson, R. H. Singer, Single-RNA counting reveals alternative modes of gene expression in yeast. *Nature structural & molecular biology* **15**, 1263 (Dec, 2008).

35.   S. J. Gandhi, D. Zenklusen, T. Lionnet, R. H. Singer, Transcription of functionally related constitutive genes is not coordinated. *Nat. Struct. Mol. Biol.* **18**, 27 (Jan, 2011).

36.   A. Raj, S. A. Rifkin, E. Andersen, A. van Oudenaarden, Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913 (Feb 18, 2010).

37.   K. Featherstone, C. V. Harper, A. McNamara, S. Semprini, D. G. Spiller, J. McNeilly, A. S. McNeilly, J. J. Mullins *et al.*, Pulsatile patterns of pituitary hormone gene expression change during development. *J Cell Sci* **124**, 3484 (Oct 15, 2011).

38.   J. W. Tullai, M. E. Schaffer, S. Mullenbrock, G. Sholder, S. Kasif, G. M. Cooper, Immediate-early and delayed primary response genes are distinct in function and genomic architecture. *J Biol Chem* **282**, 23981 (Aug 17, 2007).

39.   V. R. Iyer, M. B. Eisen, D. T. Ross, G. Schuler, T. Moore, J. C. Lee, J. M. Trent, L. M. Staudt *et al.*, The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83 (Jan 1, 1999).

40.   S. B. McMahon, J. G. Monroe, Role of primary response genes in generating cellular responses to growth factors. *FASEB J* **6**, 2707 (Jun, 1992).

41.   A. B. Harvey Lodish, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell, *Mol. Cell. Biol.* (W. H. Freeman and Company, New York, ed. 4th, 2000).

42.   K. Kovary, R. Bravo, The jun and fos protein families are both required for cell cycle progression in fibroblasts. *Mol Cell Biol* **11**, 4466 (Sep, 1991).

43.   M. E. Greenberg, A. L. Hermanowski, E. B. Ziff, Effect of protein synthesis inhibitors on growth factor activation of c-fos, c-myc, and actin gene transcription. *Mol. Cell. Biol.* **6**, 1050 (Apr, 1986).

44. W. Jochum, E. Passegue, E. F. Wagner, AP-1 in mouse development and tumorigenesis. *Oncogene* **20**, 2401 (Apr 30, 2001).

45. P. W. Vesely, P. B. Staber, G. Hoefler, L. Kenner, Translational regulation mechanisms of AP-1 proteins. *Mutat Res* **682**, 7 (Jul-Aug, 2009).

46. M. Karin, The regulation of AP-1 activity by mitogen-activated protein kinases. *J Biol Chem* **270**, 16483 (Jul 14, 1995).

47. Y. Chinenov, T. K. Kerppola, Close encounters of many kinds: Fos-Jun interactions that mediate transcription regulatory specificity. *Oncogene* **20**, 2438 (Apr 30, 2001).

48. K. McBride, M. Nemer, The C-terminal domain of c-fos is required for activation of an AP-1 site specific for jun-fos heterodimers. *Mol. Cell. Biol*. **18**, 5073 (Sep, 1998).

49. L. Bakiri, K. Matsuo, M. Wisniewska, E. F. Wagner, M. Yaniv, Promoter specificity and biological activity of tethered AP-1 dimers. *Mol. Cell. Biol*. **22**, 4952 (Jul, 2002).

50. L. O. Murphy, J. P. MacKeigan, J. Blenis, A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration. *Mol. Cell. Biol*. **24**, 144 (Jan, 2004).

51. D. Bokemeyer, A. Sorokin, M. J. Dunn, Multiple intracellular MAP kinase signaling cascades. *Kidney Int* **49**, 1187 (May, 1996).

52. T. Deng, M. Karin, c-Fos transcriptional activity stimulated by H-Ras-activated protein kinase distinct from JNK and ERK. *Nature* **371**, 171 (Sep 8, 1994).

53. C. K. Yeh, I. S. Ambudkar, E. Kousvelari, Differential expression of early response genes, c-jun, c-fos, and jun B, in A5 cells. *Am J Physiol* **263**, G934 (Dec, 1992).

54. W. W. Lamph, P. Wamsley, P. Sassone-Corsi, I. M. Verma, Induction of proto-oncogene JUN/AP-1 by serum and TPA. *Nature* **334**, 629 (Aug 18, 1988).

55. J. H. Martens, M. Verlaan, E. Kalkhoven, A. Zantema, Cascade of distinct histone modifications during collagenase gene activation. *Mol. Cell. Biol*. **23**, 1808 (Mar, 2003).

56. C. Abate, D. Luk, E. Gagne, R. G. Roeder, T. Curran, Fos and jun cooperate in transcriptional regulation via heterologous activation domains. *Mol. Cell. Biol*. **10**, 5532 (Oct, 1990).

57. K. Kovary, R. Bravo, Expression of different Jun and Fos proteins during the G0-to-G1 transition in mouse fibroblasts: in vitro and in vivo associations. *Mol Cell Biol* **11**, 2451 (May, 1991).

58. R. Chiu, W. J. Boyle, J. Meek, T. Smeal, T. Hunter, M. Karin, The c-Fos protein interacts with c-Jun/AP-1 to stimulate transcription of AP-1 responsive genes. *Cell* **54**, 541 (Aug 12, 1988).

59. H. A. Ferguson, J. A. Goodrich, Expression and purification of recombinant human c-Fos/c-Jun that is highly active in DNA binding and transcriptional activation in vitro. *Nucleic Acids Res* **29**, E98 (Oct 15, 2001).

60. T. Shiozawa, T. Miyamoto, H. Kashima, K. Nakayama, T. Nikaido, I. Konishi, Estrogen-induced proliferation of normal endometrial glandular cells is initiated by transcriptional activation of cyclin D1 via binding of c-Jun to an AP-1 sequence. *Oncogene* **23**, 8603 (Nov 11, 2004).

61. D. Zhang, J. Li, L. Song, W. Ouyang, J. Gao, C. Huang, A JNK1/AP-1-dependent, COX-2 induction is implicated in 12-O-tetradecanoylphorbol-13-acetate-induced cell transformation through regulating cell cycle progression. *Mol Cancer Res* **6**, 165 (Jan, 2008).

62. A. Schonthal, P. Herrlich, H. J. Rahmsdorf, H. Ponta, Requirement for fos gene expression in the transcriptional activation of collagenase by other oncogenes and phorbol esters. *Cell* **54**, 325 (Jul 29, 1988).

63. B. M. Hagmeyer, H. Konig, I. Herr, R. Offringa, A. Zantema, A. van der Eb, P. Herrlich, P. Angel, Adenovirus E1A negatively and positively modulates transcription of AP-1 dependent genes by dimer-specific regulation of the DNA binding and transactivation activities of Jun. *The EMBO journal* **12**, 3559 (Sep, 1993).

64. S. Gack, R. Vallon, J. Schaper, U. Ruther, P. Angel, Phenotypic alterations in fos-transgenic mice correlate with changes in Fos/Jun-dependent collagenase type I expression. Regulation of mouse metalloproteinases by carcinogens, tumor promoters, cAMP, and Fos oncoprotein. *J Biol Chem* **269**, 10363 (Apr 8, 1994).

65. L. C. Chen, B. K. Chen, J. M. Chang, W. C. Chang, Essential role of c-Jun induction and coactivator p300 in epidermal growth factor-induced gene expression of cyclooxygenase-2 in human epidermoid carcinoma A431 cells. *Biochim Biophys Acta* **1683**, 38 (Jul 5, 2004).

66. A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, S. Tyagi, Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5**, 877 (Oct, 2008).

67. A. K. Raap, M. P. van de Corput, R. A. Vervenne, R. P. van Gijlswijk, H. J. Tanke, J. Wiegant, Ultra-sensitive FISH using peroxidase-mediated deposition of biotin- or fluorochrome tyramides. *Hum Mol Genet* **4**, 529 (Apr, 1995).

68. D. Tautz, C. Pfeifle, A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* **98**, 81 (Aug, 1989).

69. A. M. Femino, F. S. Fay, K. Fogarty, R. H. Singer, Visualization of single RNA transcripts in situ. *Science* **280**, 585 (Apr 24, 1998).

70. K. Anwar, I. Voloshyna, M. J. Littlefield, S. E. Carsons, P. A. Wirkowski, N. L. Jaber, A. Sohn, S. Eapen, A. B. Reiss, COX-2 inhibition and inhibition of cytosolic phospholipase A2 increase CD36 expression and foam cell formation in THP-1 cells. *Lipids* **46**, 131 (Feb, 2011).

71. I. Topalidou, A. van Oudenaarden, M. Chalfie, Caenorhabditis elegans aristaless/Arx gene alr-1 restricts variable gene expression. *Proc Natl Acad Sci U S A* **108**, 4063 (Mar 8, 2011).

72. L. H. So, A. Ghosh, C. Zong, L. A. Sepulveda, R. Segev, I. Golding, General properties of transcriptional time series in Escherichia coli. *Nature genetics* **43**, 554 (Jun, 2011).

73. D. Y. Vargas, A. Raj, S. A. Marras, F. R. Kramer, S. Tyagi, Mechanism of mRNA transport in the nucleus. *Proc Natl Acad Sci U S A* **102**, 17008 (Nov 22, 2005).

74. Z. Waks, A. M. Klein, P. A. Silver, Cell-to-cell variability of alternative RNA splicing. *Mol Syst Biol* **7**, 506 (2011).

75. D. Y. Vargas, K. Shah, M. Batish, M. Levandoski, S. Sinha, S. A. Marras, P. Schedl, S. Tyagi, Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147**, 1054 (Nov 23, 2011).

76. A. K. Kenworthy, Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy. *Methods* **24**, 289 (Jul, 2001).

77. S. Fredriksson, M. Gullberg, J. Jarvius, C. Olsson, K. Pietras, S. M. Gustafsdottir, A. Ostman, U. Landegren, Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol* **20**, 473 (May, 2002).

78. B. Baan, E. Pardali, P. ten Dijke, H. van Dam, In situ proximity ligation detection of c-Jun/AP-1 dimers reveals increased levels of c-Jun/Fra1 complexes in aggressive breast cancer cell lines in vitro and in vivo. *Mol Cell Proteomics* **9**, 1982 (Sep, 2010).

79. O. Soderberg, M. Gullberg, M. Jarvius, K. Ridderstrale, K. J. Leuchowius, J. Jarvius, K. Wester, P. Hydbring *et al.*, Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat Methods* **3**, 995 (Dec, 2006).

80. K. J. Leuchowius, M. Jarvius, M. Wickstrom, L. Rickardson, U. Landegren, R. Larsson, O. Soderberg, M. Fryknas, J. Jarvius, High content screening for inhibitors of protein interactions and post-translational modifications in primary cells by proximity ligation. *Mol Cell Proteomics* **9**, 178 (Jan, 2010).

81. L. C. Murphy, G. P. Skliris, B. G. Rowan, M. Al-Dhaheri, C. Williams, C. Penner, S. Troup, S. Begic *et al.*, The relevance of phosphorylated forms of estrogen receptor in human breast cancer in vivo. *J Steroid Biochem Mol Biol* **114**, 90 (Mar, 2009).

82. C. Desmedt, J. Sperinde, F. Piette, W. Huang, X. Jin, Y. Tan, V. Durbecq, D. Larsimont *et al.*, Quantitation of HER2 expression or HER2:HER2 dimers and differential survival in a cohort of metastatic breast cancer patients carefully selected for trastuzumab treatment primarily by FISH. *Diagn Mol Pathol* **18**, 22 (Mar, 2009).

83. Y. Shi, W. Huang, Y. Tan, X. Jin, R. Dua, E. Penuel, A. Mukherjee, J. Sperinde *et al.*, A novel proximity assay for the detection of proteins and protein complexes: quantitation of HER1 and HER2 total protein expression and homodimerization in formalin-fixed, paraffin-embedded cell lines and breast cancer tissue. *Diagn Mol Pathol* **18**, 11 (Mar, 2009).

84. S. Fredriksson, J. Horecka, O. T. Brustugun, J. Schlingemann, A. C. Koong, R. Tibshirani, R. W. Davis, Multiplexed proximity ligation assays to profile putative plasma biomarkers relevant to pancreatic and ovarian cancer. *Clin Chem* **54**, 582 (Mar, 2008).

85. A. Raj, A. van Oudenaarden, Single-molecule approaches to stochastic gene expression. *Annu Rev Biophys* **38**, 255 (2009).

86. K. Chida, S. Nagamori, T. Kuroki, Nuclear translocation of Fos is stimulated by interaction with Jun through the leucine zipper. *Cell Mol Life Sci* **55**, 297 (Feb, 1999).

87. M. F. Favata, K. Y. Horiuchi, E. J. Manos, A. J. Daulerio, D. A. Stradley, W. S. Feeser, D. E. Van Dyk, W. J. Pitts *et al.*, Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J Biol Chem* **273**, 18623 (Jul 17, 1998).

88. M. Gossen, H. Bujard, Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**, 5547 (Jun 15, 1992).

89. H. J. Rahmsdorf, A. Schonthal, P. Angel, M. Litfin, U. Ruther, P. Herrlich, Posttranscriptional regulation of c-fos mRNA expression. *Nucleic Acids Res* **15**, 1643 (Feb 25, 1987).

90. I. Boldogh, S. AbuBakar, C. Z. Deng, T. Albrecht, Transcriptional activation of cellular oncogenes fos, jun, and myc by human cytomegalovirus. *J Virol* **65**, 1568 (Mar, 1991).

91. T. Tumbar, G. Sudlow, A. S. Belmont, Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *J Cell Biol* **145**, 1341 (Jun 28, 1999).

92. A. Raj, S. Tyagi, Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Methods Enzymol* **472**, 365 (2010).

93. Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, P. Walter, *Molecular Biology of the Cell*. (ed. fourth, 2002).

94. M. S. Jurica, M. J. Moore, Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12**, 5 (Jul, 2003).

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

96.   Y. Xing, C. V. Johnson, P. T. Moen, Jr., J. A. McNeil, J. Lawrence, Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J Cell Biol* **131**, 1635 (Dec, 1995).

97.   G. Zhang, K. L. Taneja, R. H. Singer, M. R. Green, Localization of pre-mRNA splicing in mammalian nuclei. *Nature* **372**, 809 (Dec 22-29, 1994).

98.   G. Bauren, L. Wieslander, Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76**, 183 (Jan 14, 1994).

99.   A. L. Beyer, Y. N. Osheim, Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & development* **2**, 754 (Jun, 1988).

100.   I. Listerman, A. K. Sapra, K. M. Neugebauer, Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.* **13**, 815 (Sep, 2006).

101.   A. Pandya-Jones, D. L. Black, Co-transcriptional splicing of constitutive and alternative exons. *Rna* **15**, 1896 (Oct, 2009).

102.   J. Singh, R. A. Padgett, Rates of in situ transcription and splicing in large human genes. *Struct. Mol. Biol.* **16**, 1128 (Nov, 2009).

103.   A. Yuryev, M. Patturajan, Y. Litingtung, R. V. Joshi, C. Gentile, M. Gebara, J. L. Corden, The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A* **93**, 6975 (Jul 9, 1996).

104.   N. Fong, D. L. Bentley, Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev* **15**, 1783 (Jul 15, 2001).

105.   R. Das, J. Yu, Z. Zhang, M. P. Gygi, A. R. Krainer, S. P. Gygi, R. Reed, SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell* **26**, 867 (Jun 22, 2007).

106.   R. D. Phair, T. Misteli, High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604 (Apr 6, 2000).

107.   K. M. Neugebauer, On the importance of being co-transcriptional. *J Cell Sci* **115**, 3865 (Oct 15, 2002).

108.   F. C. Oesterreich, N. Bieberstein, K. M. Neugebauer, Pause locally, splice globally. *Trends Cell Biol* **21**, 328 (2011).

109.   M. de la Mata, A. R. Kornblihtt, RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol* **13**, 973 (Nov, 2006).

110. I. Listerman, A. K. Sapra, K. M. Neugebauer, Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat Struct Mol Biol* **13**, 815 (Sep, 2006).

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

112. E. Batsche, M. Yaniv, C. Muchardt, The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol* **13**, 22 (Jan, 2006).

113. N. Custodio, M. Carmo-Fonseca, F. Geraghty, H. S. Pereira, F. Grosveld, M. Antoniou, Inefficient processing impairs release of RNA from the site of transcription. *Embo J* **18**, 2855 (May 17, 1999).

114. F. Carrillo Oesterreich, S. Preibisch, K. M. Neugebauer, Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40**, 571 (Nov 24, 2010).

115. R. D. Alexander, S. A. Innocente, J. D. Barrass, J. D. Beggs, Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* **40**, 582 (Nov 24, 2010).

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

117. S. C. Little, G. Tkacik, T. B. Kneeland, E. F. Wieschaus, T. Gregor, The formation of the Bicoid morphogen gradient requires protein movement from anteriorly localized mRNA. *PLoS biology* **9**, e1000596 (Mar, 2011).

118. S. C. Reshmi, J. L. Miller, D. Deplewski, C. Close, L. J. Henderson, E. Littlejohn, S. Schwartz, D. J. Waggoner, Evidence of a mechanism for isodicentric chromosome Y formation in a 45,X/46,X,idic(Y)(p11.31)/46,X,del(Y)(p11.31) mosaic karyotype. *Eur J Med Genet* **54**, 161 (Mar-Apr, 2011).

119. D. L. Black, Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72**, 291 (2003).

120. M. Chen, J. L. Manley, Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**, 741 (Nov, 2009).

121. M. Dutertre, G. Sanchez, M. C. De Cian, J. Barbier, E. Dardenne, L. Gratadou, G. Dujardin, C. Le Jossic-Corcos *et al.*, Cotranscriptional exon skipping in the genotoxic stress response. *Nat Struct Mol Biol* **17**, 1358 (Nov, 2010).

122. Y. Ishihama, H. Tadakuma, T. Tani, T. Funatsu, The dynamics of pre-mRNAs and poly(A)+ RNA at speckles in living cells revealed by iFRAP studies. *Exp Cell Res* **314**, 748 (Feb 15, 2008).

123. H. K. Salz, Male or female? The answer depends on when you ask. *PLoS Biol* **5**, e335 (Dec, 2007).

124. L. N. Keyes, T. W. Cline, P. Schedl, The primary sex determination signal of Drosophila acts at the level of transcription. *Cell* **68**, 933 (Mar 6, 1992).

125. M. E. Samuels, P. Schedl, T. W. Cline, The complex set of late transcripts from the Drosophila sex determination gene sex-lethal encodes multiple related polypeptides. *Mol Cell Biol* **11**, 3584 (Jul, 1991).

126. J. I. Horabin, P. Schedl, Sex-lethal autoregulation requires multiple cis-acting elements upstream and downstream of the male exon and appears to depend largely on controlling the use of the male exon 5' splice site. *Mol Cell Biol* **13**, 7734 (Dec, 1993).

127. H. Sakamoto, K. Inoue, I. Higuchi, Y. Ono, Y. Shimura, Control of Drosophila Sex-lethal pre-mRNA splicing by its own female-specific product. *Nucleic Acids Res* **20**, 5533 (Nov 11, 1992).

128. L. O. Penalva, M. J. Lallena, J. Valcarcel, Switch in 3' splice site recognition between exon definition and splicing catalysis is important for sex-lethal autoregulation. *Mol Cell Biol* **21**, 1986 (Mar, 2001).

129. C. C. Robinett, A. Straight, G. Li, C. Willhelm, G. Sudlow, A. Murray, A. S. Belmont, In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol* **135**, 1685 (Dec, 1996).

130. G. Cumming, F. Fidler, D. L. Vaux, Error bars in experimental biology. *J Cell Biol* **177**, 7 (Apr 9, 2007).