ABSTRACT

Title of Dissertation: POST-TRANSCRIPTIONAL REGULATION OF SPERMATOGENESIS THROUGH INTRON RETENTION IN THE FERN *MARSILEA VESTITA*

Thomas Christopher Clark Boothby, Doctor of Philosophy, 2013

Dissertation directed by: Dr. Stephen M. Wolniak
Department of Cell Biology and Molecular Genetics

Many rapidly developing systems rely on the use of stored transcripts to carry out their developmental program. The microspore of *M. vestita* transcribes and stores RNA during a requisite period of desiccation. Rehydration of the microspore triggers spermatogenesis to begin, a process that is mediated by the utilization of these stored RNAs. Here I investigate mechanisms controlling the spatial and temporal utilization of these stored transcripts. Next generation Solexa based RNAseq was conducted using poly(A)+ RNA isolates from specific time ranges during spermatogenesis. A reference transcriptome as well as temporally specific transcriptomes were assembled *de novo* and analyzed for gene ontology enrichments. This analysis revealed an overrepresentation of catalytic splicing and nuclear speckle factors early in development suggesting that some transcripts are not fully mature. An in house Visual Basic for Applications program was used to identify potential intron retaining transcripts (IRTs) within our transcriptomes. A large subset of IRTs was identified and *in silico* and molecular biological approaches

demonstrated that these IRTs are matured in a spliceosome dependent fashion at different times during development. Intron retention appears to confer a translational block to IRTs and splicing of retained introns alleviates this block. IRTs appear to be associated with splicing machinery organized in nuclear speckles. These subnuclear domains aggregate during desiccation and upon rehydration are proportioned asymmetrically to spermatogeneous cells. It appears that intron retention mediates both the association and asymmetric distribution of IRTs with nuclear speckles as well as their temporal utilization through post-transcriptional splicing.

Post-transcriptional Regulation of Spermatogenesis Through Intron Retention in the Fern *Marsilea vestita*

By

Thomas Christopher Clark Boothby

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Committee:

Dr. Todd Cooke

Dr. Philip DeShong, Dean's Representative

Dr. Stephen Mount

Dr. Leslie Pick

Dr. Stephen Wolniak, Chair

Preface

The hereditary information of all living organisms is encoded by their DNA in units called genes. Genes are analogous to blue prints and are used to instruct living cells how to make cellular machinery that is essential for life. There are several different molecular forms that this information flows through in order for genes to be expressed. Information encoded in DNA molecules is first transcribed into a different type of molecule called RNA. Cells use machinery to decipher the RNA code and translate this code into a different molecular form, animo acids, which combine to make proteins.

Proteins carryout many essential processes in cells but we do not fully understand the nuaincess of their regulated production. Understanding how genes and their products are regulated is important, as this regulation is the basis of life.

For a long time it has been assumed that the rate-limiting step in gene expression occurs as information contained within DNA is copied to RNA. While it is true that cells can control the flow of genetic information at this step, researchers are finding with increasing frequency that regulating the transfer of information from RNA into proteins is a major mechanism for controlling gene expression.

The project presented here investigates how cells from a fern, *Marsilea vestita*, are able to make and store RNA in a state where proteins will initially not be produced. These plant cells are able to tolerate extreme drying and use stored RNA molecules to recover from this stress as well as to continue developing in an effort to generate offspring. We found that these cells make RNA in a precursor state, which cannot be used in the production of proteins. When needed by the cells, these precursor RNA molecules are modified to their final functional form, which can then be used by the cell. These percursors also interact with other molecules that deliver them only to certain types of cells. In this way generating precursor RNA molecules allows cells to regulate their use both spatially (in specific types of cells) and temporally (only when they are needed).

Although these studies were conducted in a plant, regulation of RNA is seen in many different organisms, including humans. The production of precursor RNAs is a widespread event in many types of organisms, but until now whether these precursors had a function in plant cells was unknown. The mechanisms identified here have increased our understanding of how organisms are able to regulate the production of cellular machinery through the processing of intermidate RNA molecules.

Acknowledgments

I acknowledge both past and present members of the Wolniak lab.

I thank Dr. Minoru Yoshida (RIKEN) for providing Spliceostatin A. I appreciate the assistance and advice from Drs. Najib El-Sayed and Jungmin Choi (University of Maryland, College Park) for their help in performing RNAseq and transcriptome analysis. I am grateful to Dr. Joseph Gall (Carnegie Institution of Washington, Baltimore, MD) for his insights and for generously providing us with 4G3 and 72B9 antibodies. I would like to thank Dr. Stephen Mount (University of Maryland, College Park) for providing insight, encouragement, and technical knowledge that has been essential for this project. I thank Suqei Zhao and Kongyi Jiang (IBBR sequencing core, University of Maryland, College Park) as well as AbdulShakur Abdullah (Cosmos ID Inc.) for aid in assessing the quality of cDNA library preps and performing RNAseq runs.

I would like to acknowledge and thank all the members of my dissertation committee; Drs. Todd Cooke, Philip DeShong, Stephen Mount, and Leslie Pick for sharing with me their time, insight, and encouragement.

Finally, I would like to offer my most heartfelt appreciation to my mentor Dr. Stephen M. Wolniak.

TABLE OF CONTENTS

List of Abbreviations

AS – alternative splicing

BLAST – basic local alignment search tool

cDNA – complementary DNA

Cen1 – Centrin 1

DAPI – 4',6-diamidino-2-phenylindole

DNA - Deoxyribonucleic acid

EJC – exon junction complex

ESS – Exonic splice silencers

FISH – fluorescence *in situ* hybridization

FPKM – fragments mapped per kilobases of exon per million reads

G.O. – gene ontology

ICG – interchromatin granule

IR – intron retention

IRT – intron retaining transcript

ISH – *in situ* hybridization

ISS – Intronic splice silencers

jk – jacket cell

MIG – mitotic interchromatin granule

miRNA – micro RNA

mRNA – messenger RNA

NMD – nonsense mediated decay

NTC – nineteen complex

NudCD2 – NudC containing-domain protein 2

Poly(A)+ RNA – polyadenylated RNA

PY – pyronin Y

qc-mRNA - Quiescent cytoplasmic messenger RNA

RACE-PAT – Rapid amplification of cDNA ends polyadenylation test

RNA - Ribonucleic acid

RNAi – RNA interference

RNAseq – RNA sequencing

RT-PCR – reverse transcription polymerase chain reaction

rRNA – ribosomal RNA

RUST – regulated unproductive splicing and translation

sc – sporocarp

siRNA – small interfering RNA

snRNP – small nuclear ribonucleoprotein

sp – spermatogenous cell

SPD – spermidine

SPDS – spermidine synthase

TBO – Toluidine blue O

UTR – untranslated region

Chapter I – An introduction to spermatogenesis in *Marsilea vestita* and post-transcriptional regulation

Introduction to *M. vestita*

Marsilea vestita is a semi-aquatic, heterosporous water fern whose sporophyte superficially resembles a four-leaf clover (Figure I-1A). Like all plants, *M. vestita* undergoes alternation of generations, cycling between a diploid sporophyte and a haploid gametophyte. Being heterosporous, *M. vestita* produces both mega and microspores through meiosis, which will develop into female and male gametophytes, respectively (Figure I-1B). Spores develop within sporangia, which, in turn, are clustered in structures known as sori that are attached to a modified leaf packaged within a protective sporocarp (Figure I-1C). The sporocarp undergoes desiccation and can remain in this state up to 130 years (Moran, 2004). Upon hydration, the sporocarp opens and the modified leaf structure expands exposing the sori (Figure I-1C). Uptake of water by spores triggers the start of gametophyte development. Microspores develop into the male gametophytes, which will divide to form spermatids that ultimately develop into the male gametes or spermatozoids through a process known as spermatogenesis (Figure I-2).

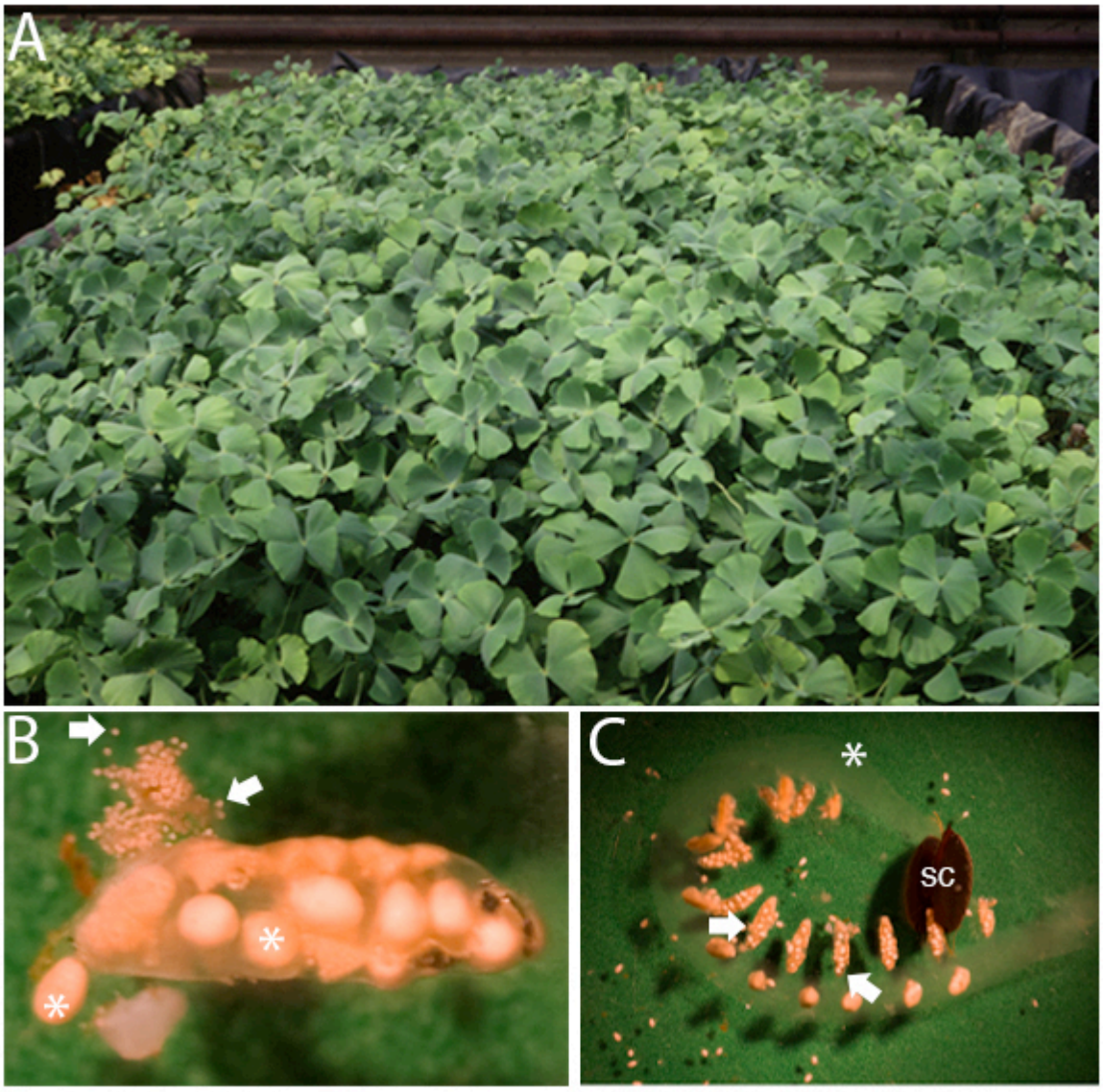


Figure I-1. *Marsilea vestita is a semi-aquatic fern*. (**A**), *M. vestita* sporophytes growing in artificial ponds at the University of Maryland greenhouse. (**B**), Microspores (arrows) and megaspores (*) of *M. vestita*. (**C**), A rehydrated sporocarp (sc) after ejection of its sori (arrows) containing modified leaf structure (*). Images by S.M. Wolniak.

Spermatogenesis in *Marsilea vestita*

Spermatogenesis (Figure I-2) is initiated when the spore is immersed or imbibed in water. The duration of development is temperature dependent with fully motile spermatozoids emerging after 11 hours at 20ºC, while culturing of gametophytes at 30ºC

results in spermatozoid release in about 6 h. At 20ºC development of the male gametophyte occurs in two distinct phases, the division phase and the differentiation phase.

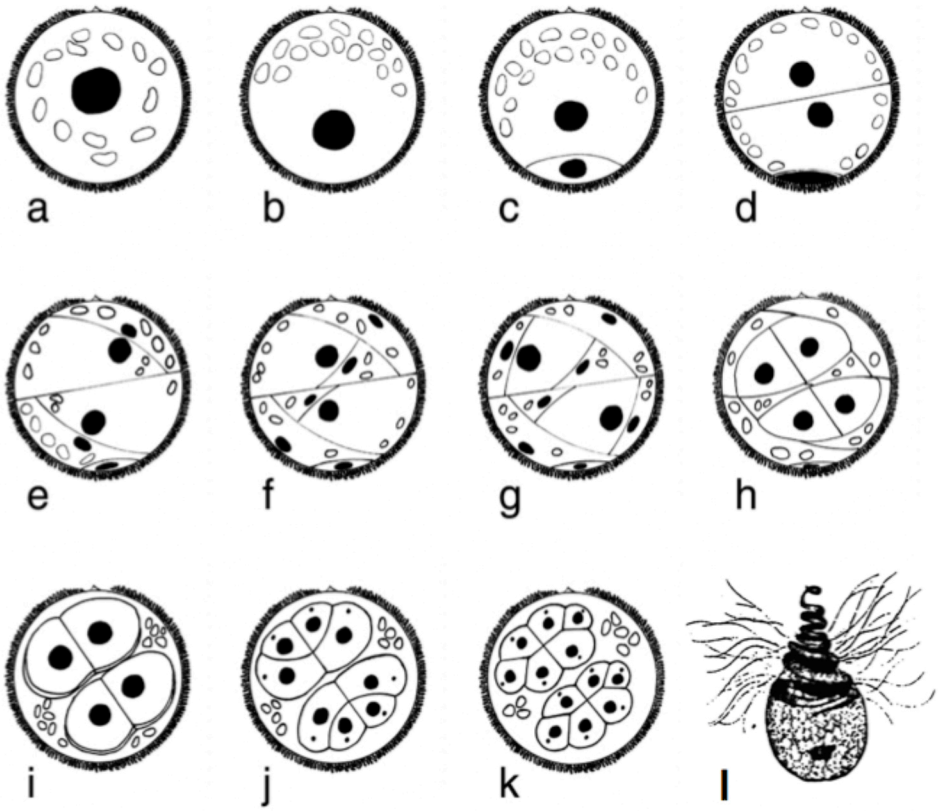


Figure I-2. *Spermatogenesis in M. vestita*. (**A**), The desiccated microspore contains a single meiotic cell with a centrally located nucleus (black) and many starch containing plastids (small white spheres). (**B**), Upon rehydration the nucleus and plastids displace to opposite poles. (**C**), The first asymmetric division occurs generating the antheridial initial (bigger cell) and a sterile prothallial (smaller) cell. (**D**), The first asymmetric division divides the antheridial initial in half. (**E-G**), 3 rounds of asymmetric divisions give rise to 6 sterile "jacket cells." (**H-K**), The remaining 4 rounds of symmetric divisions generate a total of 32 spermatogenous cells. (**I**), Spermatogenous cells differentiate into fully motile chemotactic sperm, each possessing ~140 cilia. Adapted from drawings by S.M. Wolniak; Sharp, 1914.

Before hydration, the microspore houses a single meiotic cell with a large centrally located nucleus surrounded by plastids. Upon hydration, which prompts the beginning of the division phase of spermatogenesis, a dramatic localization process begins resulting in the aggregation of plastids and the nucleus to opposite poles of the cell. By 45 minutes post-imbibition, the first asymmetric division has begun and generates a germ cell and prothallial cell. Approximately 2 hours after hydration the germ cell undergoes a symmetric division establishing two antheridial initials. Each of these antheridial initials undergoes 3 asymmetric divisions generating 6 sterile jacket cells and 2 spermatogenous initials. These initials continue to divide symmetrically an additional 4 times. This series of divisions ultimately gives rise to 32 spermatids and 7 sterile cells (6 jacket and 1 prothallial cell). The division phase ends with these last 4 divisions, and the spermatids enter their differentiation phase during which they will synchronously develop into fully motile spermatozoids. During this second phase, each spermatid will elongate and coil its nucleus and other organelles, produce ~140 cilia, and upon emergence from the microspore wall will have become a fully differentiated motile and chemotaxtic spermatozoid. Each of these 2 phases takes approximately 5.5 hours to complete, so that within 11 hours a single quiescent cell divides and gives rise to 32 spermatozoids, which are morphologically distinct from their initial progenitors (Figure I-2). This developmental program is extremely fast and synchronous. Later in this chapter, preliminary evidence, suggesting that this rapid burst of development is mediated post-transcriptionally will be presented. The remaining chapters of this dissertation will present novel evidence that rapid development of spermatozoids of the fern *M. vestita* is mediated through the use of stored transcripts and that post-transcriptional splicing of

retained introns and association with nuclear speckles regulates both temporal and spatial use of these stored transcripts.

The choice and use of *M. vestita* as an experimental system

Why use *Marsilea* as a model for biological studies?

The experiments conducted in this project were carried out using the male gametophyte of the fern *Marsilea vestita*. Many aspects of *M. vestita* male gametophyte development make it a biologically intriguing and logistically excellent system to study. As a system for studying development, the gametophyte, which is housed within the microspore is ideal because of the speed, precision, and synchrony with which profound developmental changes occur. As a system for studying morphological events, *Marsilea* is amazing, as within 5.5 hours of their initial formation spermatids transition from typical plant cells fixed in place to elongated, coiled, chemotactic cells capable of swimming in response to chemical cues through the formation of ~140 motile cilia. For studying cell fate determination, *Marsilea* offers several advantages. Firstly, there are only two cell types formed within the microspore, sterile cells and spermatogenous cells. Secondly, only 32 spermatogenous cells and 7 sterile cells are generated and their divisions are temporally and spatially precise. The fate map of microspore development is simple and obvious. Thirdly, the types of divisions (asymmetric for sterile, and symmetric for spermatogenous) giving rise to and the markers identifying different cell types have been established. The diagnosis and identification of components involved in aspects of cell fate determination is greatly increased because of these properties.

The microspore of *M. vestita* is an excellent system to study post-transcriptional regulation, since *de novo* synthesis of RNA is not required (and does not occur). This means that experiments addressing the regulated use of stored RNA are not confounded by new input material, essentially giving studies on the microspore the simplicity of *in vitro* studies while being conducted *in vivo*. Numerous other qualities of the microspores make them logistically amenable to working with in the lab. Growing plants from which to harvest microspores is easy and once having matured one literally neglects the plant, allowing it to dry up and desiccate. Once desiccated, within a day or two enough sporocarps containing microspores can be harvested to conduct experiments with for several years. The dry microspore is capable of being stored for decades (the longest estimated case is ~130 years; Moran, 2004) while retaining their developmental potential. RNA and proteins can easily be isolated from the spores, making them amenable for conducting most molecular biological assays. Microspores are easily fixed and their exines (outer spore wall) removed for fluorescence confocal microscopy, or embedded for sectioning. Traditional cell labeling methods, such as tissue staining and *in situ* hybridization, as well as more contemporary methods such as fluorescence *in situ* hybridization and fluorescent antibody labeling are easily performed on fixed whole mounts or sections making most cell visualization assays possible. Reverse genetics studies via RNAi are easily performed, as microspores will take up dsRNA suspended in water at the time of imbibition. Likewise, many small drugs are capable of being introduced to microspores at the time of imbibition through approximately 3 hours of development. Taken together, the ease and convenience of working with microspores,

coupled with the profound nature of their biology makes them a worthy system to work on.

How do the experiments performed with *Marsilea* expand on what is known in other biological systems?

While many aspects of spermatogenesis in *M. vestita* seem novel at first glance, further investigation reveals that as a developmental system the microspore highlights certain phenomena and mechanisms that exist else where in the biological world though they may be obscured or otherwise not as patent as in the microspore. Work from our laboratory has informed and been informed by various fields of study. The discovery that the exon junction complex protein Mago nashi influences the symmetry of cell divisions in *Marsilea* (van der Weele *et al*., 2007) informed later studies in more mainstream model systems (Silver *et al*., 2010). Our investigations (Deeb *et al*., 2010) into polyamines as morphogenic agents have subsequently been confirmed by, and have served as a basis for other studies (Vuohelainen *et al*., 2010; Zhang *et al*., 2011; Valdes-Santiago *et al*., 2012). Investigations into Mv-Cen1 in *Marsilea* (Klink and Wolniak, 2001) was the first time a function had been assigned to Centrin (which is now well characterized and widely studied). The list goes on, but it should be apparent by these examples that research in the somewhat obscure fern *Marsilea vestita* has enlightened many investigators working in diverse fields and systems from mammals to fungi.

The research reported in this dissertation mainly relates to novel functions of nuclear speckles and the regulated post-transcriptional splicing of intron retaining

transcripts, which together drive development and differentiation of spermatids. The discovery of nuclear speckles in the microspore of *M. vestita* was accomplished through a combination of speckle marker labeling as well as observation of speckle dynamics. Mammalian antigens were used to produce antibodies against all speckle markers used in this study. Speckles in *Marsilea* also contain a subset of poly(A)+ RNA and many of these transcripts contain retained introns that are subjected to post-transcriptional splicing. These molecular components and the role of speckles in their processing have been reported in other systems. Nuclear speckle dynamics in the microspore mimic those previously described in other systems, aggregating during transcriptional/splicing inhibition and entering the cytosol during mitosis. The fact that nuclear speckles in *Marsilea* contain the same molecular constituents and that their dynamics in control and experimental cells are identical to those reported in other species, points to the fact that functions and mechanisms of developmental control carried out in nuclear speckles are well conserved. Therefore, it is likely that novel functions for nuclear speckles discovered in *Marsilea* are not unique but will be found to extend to other organisms as well.

Intron retention (IR) is widespread in all eukaryotic kingdoms, and is the predominant form of AS in plants (Campbell *et al*., 2006; Kim and Gladyshev, 2006; Ner-Gaon *et al*., 2007; Syed *et al.,* 2012). It has been established that IR serves a functional role in *metazoan* cell biology, however such functions are much less well defined in plants. Here I present evidence that a subset of transcripts sorted in the microspore of *M. vestita* contain retained introns and that the majority of these posses conserved U2 splice signals. The degree to which these splice signals display conservation between species and even kingdoms is not surprising, considering RNA is

an ancient molecule and splicing an ancient phenomenon. Likewise, splicing factors show strong conservation between *Marsilea*, other plants, and humans with many splicing factors being identified via reciprocal best match searches using *Arabidopsis* and human databases. Inhibition of splicing with an inhibitor (Spliceostatin A) originally tested in animals is capable of inhibiting splicing in *Marsilea*. The combination of highly conserved intronic sequences, splicing machinery, organization of splicing machinery (in nuclear speckles), and maturation of the spliceosome all points to the fact that splicing in *Marsilea* is carried out in a similar fashion to other model organisms. We therefore suspect that novel aspects of the regulation of splicing and the regulation of development through splicing are not unique to *Marsilea* but rather have the potential to be extremely widespread and ancient phenomena.

Post-transcriptional regulation: a brief history

Many rapidly developing systems depend on little or no transcription for the proper completion of specific phases of development (Raff *et al.,* 1972; Capco and Jeffery, 1979; Davidson, 1986; Rosenthal, *et al.,* 1993; Curtis *et al.,* 1995). The preceding statement may seem counter-intuitive since the control of gene expression has historically been viewed as the control of transcription; that is, the production of ribonucleic acid (RNA) from a deoxyribonucleic acid (DNA) template. In the traditional view, tightly packed heterochromatin is converted to open euchromatin that can be readily transcribed. Beyond this "open versus closed" view, transcription factors able to modulate gene activity add a further layer of complication and precision to the regulation of gene expression. In this overly simplistic view, during or just prior to transcription, pre-mRNA (in eukaryotes) is spliced, 5' capped, and polyadenylated and then quickly exits the

nucleus. Upon entry into the cytoplasm, mRNA becomes associated with translational machinery and immediately translated. By this view, regulation at the level of transcription appears the obvious and predominant mechanism for controlling gene expression. However, today our understanding of cell and RNA biology has led us to the inescapable conclusion that post-transcriptional regulation of RNA is extensive, precise, and potentially more important than transcriptional regulation in establishing both complexity and order in living things (Bevilacqua *et al*., 2003; Filichkin *et al*., 2008; Kornbihtt *et al*., 2013).

In the summer of 1936 Ethel Browne Harvey published on her work with eggs from the sea urchin *Arbacia punctulata.* Dr. Harvey had devised a method of centrifuging unfertilized eggs that allowed her to generate egg fragments lacking a maternal nucleus reproducibly, which, when artificially stimulated to develop, were capable of undergoing cellular divisions. The finding that parts of these cells lacking both maternal and paternal genetic material could generate embryos of approximately 500 cells and could live up to a month, led Harvey to conclude that the maternal cytoplasm contained important "potentialities" capable of determining the early stages of development and that these potentialities likely came previously from the nucleus (Harvey, 1936). Although she could not have phrased it this way at the time (the role of RNA in protein synthesis would not be proposed for another 3 years; Caspersson and Schultz, 1939), Harvey provided the world its first clues that gene expression could be controlled downstream of transcription, a phenomenon now termed post-transcriptional regulation.

It would be nearly 30 years until Gross and Cousineau would show that transcriptional inhibition did not stymie new protein production and Spirin and Nemer

would find that polyribosomes, containing stored maternal mRNA accounted for the bulk of protein synthesis in early sea urchin (*Arbacia punctulata*) embryos (Gross and Cousineau, 1963; Gross and Cousineau, 1964; Spirin and Nemer, 1965). However, despite a slow beginning, there is now substantial evidence that in plants, fungi, and metazoa some species of RNA are stored and utilized long after their transcription (Dure and Waters, 1965; Grosfeld and Littauer, 1975; Spiegel and Marcus, 1975; Amaldi *et al*., 1977; Averbeck *et al*., 2005; Nakabayashi *et al*., 2005; Denis *et al.,* 2005; Malatesta *et al*., 2009; Flemr *et al*., 2010).

With regards to the storage and utilization of RNA in plants, seeds have undoubtedly received the most attention. The first demonstration of stored (then termed "long-lived") RNA in seeds, and one of the first demonstrations in any organism, came from studies carried out by Dure and Waters (1965) on cotton. Dure and Waters germinated seeds on control or experimental filters soaked with the transcriptional inhibitor actinomycin D. After different time intervals, polyribosomes were extracted, and they assayed *de novo* RNA synthesis, and protein production. They found that through 16 hours of germination, the inhibition of transcription had negligible effects on the levels of RNA, polyribosomes, and protein synthesis, and furthermore, that embryos treated with the transcriptional inhibitor could germinate at the same rate as untreated seeds for up to 36 hours (Dure and Waters, 1965). The most likely explanation for these observations was that preformed RNAs were stored within dormant seeds and were sufficient to serve as templates for any translation required during the early stages of seed germination. Later evidence demonstrated that while these RNAs were transcribed prior

to desiccation, they might not be fully mature since a sizable fraction of stored RNAs require polyadenylation during the first day of germination (Harris and Dure, 1978).

Since these initial investigations in cotton, the study of stored RNA in other seeds has expanded to several plant models including *Arabidopsis thaliana*. In 2004, Rajjou and coworkers demonstrated that treatment of *Arabidopsis* seeds with the RNA polymerase II inhibitor α-amanitin did not restrict germination (radicle protrusion), whereas the translational inhibitor cycloheximide inhibited growth. In 2005 researchers in Japan examined the transcriptome of dry *Arabidopsis* seeds. The dormant embryo was found to contain upwards of 12,000 different mRNA species (Nakabayashi *et al*., 2005). While these mRNAs represented many ontological categories, there was a statistical overabundance of abscisic acid-responsive elements in the promoters of highly abundant stored seed RNAs (Nakabayashi *et al*., 2005). This overrepresentation is unsurprising since the plant hormone abscisic acid is known to play a role in both dormancy (review: Koornneef *et al*., 2002) and in priming the seed for post-germination growth, which is likely mediated in part by inducing the transcription of RNAs destined for storage.

While seeds constitute a part of the spermatophyte life cycle that is desiccation tolerant, there are several anhydrobiotic plants (often called resurrection plants) whose sporophytic tissues are also desiccation tolerant. It is interesting, although perhaps not unexpected, that the long-term storage of RNA has also been identified in these plants. The bryophyte, *Tortula ruralis*, loses its translational ability during desiccation, an event that is concurrent with a loss of the majority of its polyribosomes (Dhindsa and Bewley, 1977). At the same time, RNA encoding "recovery" proteins are increasingly transcribed and accumulate in mRNP particles corresponding to the small ribosomal subunit region

in sucrose density gradients (Scott and Oliver, 1994; Oliver and Bewley, 1997). In addition, the recovery of leaves of the resurrection plant, *Xerophyta humilis,* has been studied (Dace *et al*., 1998). Dormant, desiccated vegetative tissue of this plant was incubated in water, water with the transcriptional inhibitor α-amanitin, or water with the translational inhibitor cyclohexamide (Dace *et al*., 1998). Recovery was assayed by measuring the quantum efficiency of photosystem II via chlorophyll fluorescence. In these experiments, transcriptional inhibition did not perturb recovery whereas translational inhibition did (Dace *et al.,* 1998). These findings from a broad range of different plant life and developmental phases suggest that the storage of RNA could be a widespread phenomenon for rapid recovery and development after quiescence.

Plants are not the only group of organisms that mediate rapid development through the use of stored RNAs. For example, in hibernating dormice, different pre-mRNAs are stored within nuclei at different stages of maturation in a tissue specific manner. In the liver of hibernating dormice, pre-mRNA is accumulated predominantly at the splicing stage, whereas in brown adipose tissue, pre-mRNA is found to be mostly stored in its cleavage stage of maturation. Interestingly, dramatic redistributions of pre-mRNA maturation machinery are seen in these tissues during hibernation, and the accumulation of this machinery mirrors the pre-mRNA storage state. This system of regulation assures that upon arousal, dormouse tissues high in pre-mRNA at the cleavage state also have abundant stores of cleavage and polyadenylation machinery within their subnuclear processing bodies. Likewise, tissues with high levels of pre-mRNA stored at the splicing stage of maturation have a large distribution of splicing machinery localized to the subnuclear processing bodies. In this way, stored pre-mRNAs are regulated such

that upon arousal, certain tissues are metabolically activated in a prioritized fashion, with brown adipose tissue (cleavage pre-mRNA) activating first, followed by the liver (splicing pre-mRNA), and then finally *de novo* transcription in all tissues (Malatesta *et al*., 2009).

Brine shrimp larvae are known to undergo desiccation and mediate their recovery through the use of stored transcripts. Upon hydration, these organisms utilize and translate stored mRNA (Muthukrishnan *et al.,* 1975). Analysis of transcripts isolated from imbibed larvae indicate that very few (if any) messages are associated with polysomes at the resumption of development, but by 22 hours, a large portion of transcripts become associated with the translational complexes (Amaldi *et al.,* 1977; Grosfeld and Littauer, 1975). A subset of messages has been identified as being pre-loaded with the 40S subunit, a condition that appears to destine transcripts for early translation.

Similar to sea urchin embryos mentioned above, maternal RNA mediates the early development of many organisms (*e.g., Drosophila* and *Xenopus* early embryogenesis). With many examples in the literature of stored RNA being essential for mediating a rapid burst of development following a period of quiescence, the question has shifted from whether stored RNA plays a role in regulating development, to what regulates the utilization of stored RNA essential for rapid development?

The role of RNA modification in post-transcriptional regulation of translation activation

The role of RNA modification is known in specific cases to be a major contributor to the regulation of translational activation of stored RNA. Both splicing and cytoplasmic polyadenylation have been shown in multiple systems to affect the translational activity of RNA. Often this control is seen in rapidly developing systems, as in the male gametophyte of *M. vestita*.

Translation can be activated or repressed for a specific transcript by the lengthening or shortening of its poly(A)+ tail (Gorgoni and Gray, 2004). This event is mediated by specific sequences within the 3' untranslated region of mRNA, which are recognized by cytoplasmic polyadenylation factors (Fox *et al.,* 1989). Polyadenylation of transcripts begins in the nucleus, where adenines are added to the 3' end of pre-mRNA before export to the cytoplasm. Once exported from the nucleus, transcripts with the correct cytoplasmic polyadenylation sequences may be further modified. Lengthening or stabilizing a long poly(A)+ tail, ~80-250 bases, increases transcript stability and translational activity, whereas shortening the tail to ~20-40 bases, promotes the storage and diminishes translational rates of mRNA (Paris *et al.,* 1988; Rosenthal *et al.,* 1983; Fox *et al.,* 1989; McGrew *et al.,* 1989). Thus, the degree of polyadenylation has a direct effect on the ability of a transcript to form into complexes with the translational machinery in a competitive manner (Proweller and Butler, 1994). In the microspore of *Marsilea,* the distribution of cytoplasmic poly(A) polymerase is restricted almost exclusively to the cytoplasm of spermatogenous cells, suggesting that cytoplasmic polyadenylation may be an important regulating factor for the translation of mRNA during spermatogenesis and cell fate determination in the rapidly developing male gametophyte (Tsai *et al*., 2004).

Canonical (U2) splicing of RNA requires the stepwise assembly and maturation of a multi-subunit ribonucleic protein complex known as the spliceosome. There are 5 small nuclear ribonucleic protein (snRNP) subunits that make up the spliceosome (review: Rino and Carmo-Fonseca, 2009). These snRNPs are named for the small nuclear RNA that is present in each: U1, U2, U4, U5, and U6. Introns present in pre-mRNA contain sequence elements that interact with various components of snRNPs and mediate spliceosome assembly. These sequences include, the 5' and 3' splice site (donor and acceptor sequences), the polypyrimidine tract, and the branch point.

The assembly of a functional spliceosome requires the progressive maturation of different spliceosomal complexes (for review: Wahl *et al*., 2009). The first step of spliceosome assembly is the recognition and binding of the 5' splice site by the U1 snRNP (and other non-snRNP factors), forming the so-called 'E complex.' One of the non-snRNP factors involved in E complex formation is U2AF. U2AF interacts with and recruits the U2 snRNP to the branch point, where the U2 snRNP binds forming the 'A complex' spliceosome (sometimes terms the commitment complex). After A complex formation the remaining snRNPs (U4, U5, and U6) are recruited to the spliceosome as a preexisting complex termed the tri-snRNP. Association of the tri-snRNP with the A complex promotes the maturation of the spliceosome to the 'B complex.' Next, a non-snRNP complex termed the 'nineteen complex' or NTC (the NTC is so named because one of its components is splicing factor Prp19) mediates the rearrangement of snRNPs within the B complex resulting in the dissociation of snRNPs U1 and U4, thus forming the 'activated B complex' (often represented as B* complex). Further rearrangements of the remaining snRNPs (U2, U5, and U6) result in so-called 'C complex' formation. The

C complex is the final stage of catalytic spliceosome maturation. Following C complex formation, two transesterification reactions occur, which result in the removal of the intron and joining of the 3' end of the upstream exon with the 5' end of the downstream exon.

The first splicing transesterification reaction occurs between the 5' splice site and the branch point. The 2'OH of the branch point adenine performs a nucleophilic attack on the first base of the intron (typically a guanine) resulting in the formation of a lariat. The second transesterfication reaction occurs when the 3'OH of the upstream exon performs a nucleophilic attack on the first base of downstream exon. The second transesterification results in the joining of the two exons and the release of the intron-lariat with the spliceosome C complex. Through some further rearrangements, the spliceosome dissociates from the intron and its subunits are recycled.

As the two exons are joined during the second transesterification reaction, the exon junction complex (EJC) can be found deposited ~22 bases upstream of the splice junction. The EJC is made up of several core proteins (such as Mago nashi, eIF4AIII, y14, and Aly) as well as more transiently associating factors. Since canonical U2 splicing is restricted to the nucleus, the EJC associates with mRNA prior to its export to the cytoplasm. EJCs remain associated with mRNA until the pioneering round of translation occurs, during which EJCs are removed (for review: Tange *et al*., 2004).

Splicing of transcripts increases translation significantly, a phenomenon that has been attributed to the deposition of Exon Junction Complex (EJC) proteins on the spliced mRNA, which are believed to mediate polysome association (Nott *et al.,* 2003). Splicing

has also been shown to have an effect on the rate or time at which a particular mRNA is exported from the nucleus to the cytoplasm, or in a few cases, to alter the subcellular distribution of specific transcripts (Luo and Reed, 1999; Zhou *et al.,* 2000; Le Hir *et al.,* 2001; Ryu and Mertz, 1989; Rafiq *et al.,* 1997). Unspliced transcripts with premature stop codons are targets of nonsense-mediated decay (NMD), and therefore undergo accelerated degradation resulting in negligible amounts translation products. Splicing of premature stop codons allows for passage of a transcript through the NMD pathway allowing for downstream translation of the message beyond the prioneering round (Maquat and Carmichael, 2001; Wilusz *et al.,* 2001; Wilkinson and Shyu, 2002).

So far, I have described where (in the nucleus) and how (through the stepwise assembly of a catalytic spliceosome) splicing occurs. An additional consideration is when splicing occurs. The textbook answer is that splicing occurs co-transcriptionally, while the downstream (3') end of the pre-mRNA is still being transcribed upstream (5') introns are being removed. Microscopic observation of individual nascent RNAs formed the bases for suggesting that splicing can occur co-transcriptionally (Beyer *et al*., 1981; Beyer and Osheim, 1988; Bauren and Wieslander, 1994). Numerous studies at the individual gene or global level have suggested that co-transcriptional splicing occurs the majority of the time (Neugenbauer and Roth, 1997; Pandya-Jones and Black, 2009; Khodor *et al*., 2011). Experimental evidence suggests that ~80% of splicing events in mammalian cell culture are co-transcriptional, with the remainder occurring post-transcriptionally (Girard *et al*., 2012). While co-transcriptional splicing is the predominant form of splicing, post-transcriptional splicing (splicing of a fully transcribed and 3' cleaved pre-mRNA) is known to exist (Wetterberg *et al*., 1996; Vargas *et al*.,

2011). While some post-transcriptional splicing could occur near the site of transcription (*e.g.,* just after the pre-mRNA has been 3' cleaved and dissociated from their DNA templates), evidence suggests that a large proportion occurs within subnuclear domains known as nuclear speckles (Girard *et al*., 2012). Few studies have been specifically conducted to characterize post-transcriptional splicing fully and the exact substrate(s) of post-transcriptional splicing have yet to be fully investigated. It is unknown whether targets of post-transcriptional splicing are transcripts that have yet to undergo any splicing (*i.e.,* they contain all their introns) or if post-transcriptional splicing occurs on pre-mRNAs harboring only a few retained introns.

Intron retention and post-transcriptional splicing

Alternative splicing (AS) is common in many eukaryotes (Campbell *et al*., 2006; Kim and Gladyshev, 2006; Ner-Gaon *et al*., 2007; Syed *et al.,* 2012). Intron retention (IR) is an AS event where the splicing of an intron is skipped, resulting in an otherwise mature transcript containing an unprocessed sequence. While IR events have been shown to be more common in plants than in animals (45.1% and ~30-47.9% of all AS products in rice and *Arabidopsis* versus 2-14.8% in humans), questions about the widespread function of IR in plants persist (Reich *et al*., 1992; Campbell *et al*., 2006; Ner-Gaon *et al*., 2004; Clark and Thanaraj, 2002; Kan and Gish, 2002; Iida *et al*., 2004; Wang and Brendel, 2006; Galante *et al*., 2004).

Intron retention has a mechanistic role in metazoans (Galante *et al*., 2004). The examination of intron retention in 21,106 human genes demonstrated that there is a large overrepresentation of retained introns in the UTRs of transcripts (47% observed *versus*

9% expected; Galante *et al*., 2004). The sequences of those introns retained within the coding region of human transcripts suggest that they have been selected against for protein coding potential, with approximately 1/4th of these encoding a conserved domain (Galante *et al*., 2004). Not surprisingly, up to 22% of intron retention events in human transcripts are estimated to be conserved in sequences obtained from mice (Galante *et al*., 2004). It is important to note that the conservation of retained introns is antithetical to the observation that intron loss has been a major phenomenon in eukaryotic gene evolution (Roy and Gilbert, 2004; Roy 2006), and further underscores the likelihood that these events play a mechanistic role(s) in development and cellular physiology.

A major function attributed to intron retention in animals is regulated unproductive splicing and translation (RUST), in which regulated retention of an intron negates a pre-mRNA's protein coding ability. Proinsulin pre-mRNA has been shown to be subjected to RUST as a means of regulating tissue development (Mansilla *et al*., 2005). During chick embryonic development, increasing levels of proinsulin transcript retaining the 1st intron (within the 5' UTR) coincide with decreasing levels of proinsulin protein (Mansilla *et al*., 2005). While retention of this intron was experimentally shown to inhibit nearly all translation of the transcript, it did not appear to affect longevity or nuclear export. As early embryogenesis transitions into organogenesis, the proportion of intron retaining proinsulin transcripts increase specifically within the heart tube and are absent in other developing regions, such as the pancreas (Mansilla *et al*., 2005). Tellingly, ectopic implantation of this isoform into a pre-organogenesis embryo resulted in the premature expression of cardiac marker genes, suggesting that intron retaining transcripts have the capacity to influence major developmental events (Mansilla *et al*., 2005).

Intron retention modulates development not only in animals but also in fungi. A study examining transcripts induced upon entry into meiosis in *Schizosaccharomyces pombe* identified 12 of 92 transcripts with retain introns during differentiation, and that these introns are removed at specific times during meiosis (Averbeck *et al*., 2005). The removal of these retained introns is specifically inhibited until the developmental time of their splicing, with the presence of retained introns serving as a block to translation (Averbeck *et al*., 2005). The studies by Averbeck and Mansilla show independently that intron retention is a mechanism for blocking translation, though Averbeck extends this further by implicating intron retention in forestalling translation, and thus allowing cells to accumulate preformed, but translationally-incompetent intron retaining transcripts, which can be utilized at a later time. The yeast cells can mediate the timing of release from translational inhibition through post-transcriptional splicing. Shortly after this phenomenon was described in yeast, it was found to occur in mammalian cell culture, albeit under specialized circumstances (Denis *et al*., 2005).

Since 2005, additional evidence for the post-transcriptional splicing of retained introns as a mechanism for forestalling translation has been found. The expression of genes encoding essential presynaptic components was shown to be regulated through incomplete 3' UTR splicing (Yap *et al*., 2012). Polypyrimidine tract binding protein 1 (Ptbp1) appears to interfere with splicing of 3' introns. In non-neuronal cells, this block leads to the initiation of nonsense-mediated decay (NMD), resulting in the degradation of these transcripts. However, in neuronal cells, nuclear surveillance machinery allows for nuclear retention and stabilization of these transcripts. As neurons continue to differentiate, Ptbp1 levels decrease, releasing transcripts with 3' retained introns from

splicing inhibition. These introns are spliced out and the resulting mRNAs relocalize to the cytosol for translation (Yap *et al*., 2012). This study along with the others mentioned above, highlight the potentially enormous role intron retention plays in priming cells for development while at simultaneously preventing precocious and ectopic translation of RNA.

Despite accumulating evidence of roles for intron retention in both animals and fungi, a function for intron retention has not been demonstrated in either of the other two eukaryotic kingdoms (*Plantae* and *Protista*) even though intron retention is the dominant form of AS in both (Campbell *et al*., 2006; McGuire *et al*., 2008).

While a mechanistic function for intron retention in plants has not been elucidated, studies of specific transcripts show regulation of intron retention that imply functionality. A line of investigation has identified an intron retention event arising from the cadmium-induced use of an upstream transcription initiation site in the maize Bz2 gene (Marrs and Walbot, 1997). In the presence of cadmium, Bz2 transcripts specifically (the effect does not extend to other pre-mRNAs: *e.g.,* Actin) retain an intron. The increase in intron retention is concomitant with an increase in the accumulation of Bz2 transcript. The increase in intron retention and transcript abundance does not affect levels of Bz2 activity (Marrs and Walbot, 1997). These results are specific to cadmium and do not extend of other stresses (Marrs and Walbot, 1997). Cadmium also induces the retention of an intron in several heat-shock protein (hsp) transcripts in various plants (Czarnecka *et al*., 1984; Rochester *et al*., 1986; Winter *et al*., 1988). In these cases, heat stress increases the abundance of fully spliced hsp transcripts, whereas cadmium stress not only increased the abundance of fully spliced hsp transcripts but also resulted in the

appearance of an intron retaining form of the transcript. It is interesting to note that stress-related inhibition of splicing could be rescued in *Drosophila* by the induction of hsps *via* brief pre-exposure to stress (Yost and Lindquist, 1986). This phenomenon was not observed in plants, suggesting that while hsps in animals may play a role in the stabilization of splicing during stress, in plants intron retention during stress is not influenced by hsp protection of splicing factors (Winter *et al*., 1988).

Other environmental stresses to plants have been shown to induce intron retention. Cold stress is known to induce the incomplete splicing of transcripts encoding a ribokinase as well as a C3H2C3 RING-finger protein (which was also induced to retain introns after dehydration stress) in wheat (Mastrangelo *et al*., 2005).

Mechanisms of intron retention

Despite not fully understanding the function of all intron retention events, we understand something about how they arise. Typical U2 introns have conserved sequences at both their 5' and 3' ends terms the donor and acceptor sequences respectively (Mount, 1982). Most IR events are thought to involve poorly defined splice signals that contribute to suboptimal splicing efficiency (Hampson and Rottman, 1987; Dirksen *et al*., 1995; Romano *et al*., 2001; Sterner and Berget, 1993; Talerico and Berget, 1994; McCullough and Berget, 1997; Romfo *et al*., 2000; Sakabe and de Souza, 2007). In addition, cis acting elements such as Intronic Splice Silencers (ISS) and Exonic Splice Silencers (ESS) modulate spliceosome activity and induce intron retention (for review: Wang and Burge, 2008). Short intron length has also been implicated in IR (Galante *et al*., 2004; Stamm *et al*., 2000; Sugnet *et al*., 2004; Zheng *et al*., 2005; Ohler *et al*., 2005;

Sakabe and de Souza, 2007). IR can be triggered by external stimuli, can be specific to developmental phases and tissue types, and can display sexual dimorphism (Marrs and Walbot 1997; Winter *et al*., 1988; Averbeck *et al.,* 2005; Mansilla *et al*., 2005; Gebauer *et al*., 1998). Retained introns affect the stability, function, localization, and translatability of the transcripts containing them (Altieri, 1994; Ebihara *et al*., 1996; Bor *et al*., 2006; Jaillon *et al*., 2008; Buckley, 2011).

As mentioned above, introns not spliced co-transcriptionally are capable of undergoing post-transcriptional maturation (Wetterberg *et al*., 1996; Averbeck *et al*., 2005; Denis *et al*., 2005; Vargas *et al*., 2011), the majority of which is thought to occur within nuclear speckles (Girard *et al*., 2012).

Nuclear speckles

Nuclear speckles, also referred to as SC35 speckles, interchromatin granules/particles, or less commonly, splicing speckles, or just speckles, are interchromatin domains highly enriched in pre-mRNA splicing factors (review: Lamond and Spector, 2003; Spector and Lamond, 2011). The first clues that nuclear speckles might be involved in RNA splicing came in 1979 when it was found that antibodies detected splicing ribonucleoproteins within intranuclear granules (Perraud *et al,* 1979). These structures had been observed long before their link to splicing was made. The first description of the structures we now refer to as nuclear speckles came in 1910 when the neuroscientist Ramón Y Cajal described "translucent clumps," that he observed within the nuclei of neurons (Cajal, 1910). The nature of these clumps remained elusive for the

next 40 years until these structures were rediscovered using electron and fluorescence microscopy, at which time the terms “interchromatin particles” and “speckles” came into use (Swift, 1959; Beck, 1961).

As mentioned above, nuclear speckles occur within the interchomatin space of the nucleus. They are typically clustered into 20-25nm regions (granules) and these regions number between 20-50 in a typical mammalian cell (Thiry, 1995). Though speckles usually cluster in regions devoid of DNA, they are often found adjacent to sites of high transcriptional activity (Thiry, 1995; Huang and Spector, 1991, Brown *et al*., 2008).

Nuclear speckles are highly enriched in snRNP and SR proteins (Fu, 1995). While localization to nuclear speckles has been used as a diagnostic marker of proteins involved in splicing, several other types of proteins without known roles in splicing are associated within these structures (Krause *et al*, 1994; Larsson *et al*., 1995; Dostie *et al*., 2000; Li *et al*., 1999; Nakayasu *et al*., 1984). In addition to protein constituents, nuclear speckles have also been found to contain a subset of Poly(A+) RNA (Carter *et al*., 1991).

While nuclear speckles by definition are thought of as intranuclear domains, many speckle constituents actually cycle between the nucleus and cytoplasm where they also form defined aggregates (Verheijen *et al*., 1986; Ferreira *et al*., 1994). This shuttling occurs in an orderly and predictable cycle, and is known as the speckle cycle. The cycle starts with the breakdown of the nuclear envelope, at which time many proteins formerly associated with nuclear speckles become homogenously distributed within the cytosol (Reuter *et al.,* 1985; Spector and Smith, 1986). While a portion of these proteins remain diffuse within the cytosol, a subset aggregates during metaphase into mitotic

interchromatin granules (MIGs), which by all analysis to date, appear to be analogous to interchromatin granules (ICGs), differing only in their cellular localization nuclear (ICGs) versus cytoplasmic (MIGs) (Ferreira *et al*, 1994; Leser *et al.,* 1989; Prasanth *et al*., 2003; Thiry 1995). As the mitosis progresses through anaphase and into telophase, the size and number of MIGs continually increases. MIG associated splicing factors are recycled back into the nucleus, but not until the nuclear envelope is reestablished (Prasanth *et al*, 2003). Initially, it was thought that intact MIGs might be transported into the nucleus (Leser *et al*., 1989; Thiry, 1995). However, more detailed investigations have uncovered that intact MIGs are not transported from the cytosol into the nucleus, but rather, there are at least two waves of MIG recycling (Ferreira *et al*., 1994; Prasanth *et al*., 2003). Initial clues came from Ferreira, who detected MIGs with antibodies against various splicing factors after some snRNPs had been imported into the nucleus (Ferreira *et al*., 1994). Later, it was found that the first wave of MIG import comes in late telophase when SF2/ASF, SC35, U2B", and other fully functional splicing factors are displaced from MIGs and enter the nucleus, forming nuclear speckles. The second wave occurs in G1 when residual SC35 and RNAPII enter the nucleus (Prasanth *et al*., 2003). While MIGs appear to be the cytoplasmic equivalent of nuclear speckles, a role for MIGs has not been established.

Because of the enrichment of splicing factors and their proximity to actively transcribing genes, nuclear speckles were once proposed to be the site of pre-mRNA splicing. However, it has been found that the majority of pre-mRNA species localize to perichromatin fibrils, which, due to their small size (~3nm) and close proximity to nuclear speckle granules, makes immunofluorescence differentiation difficult (Fakan and

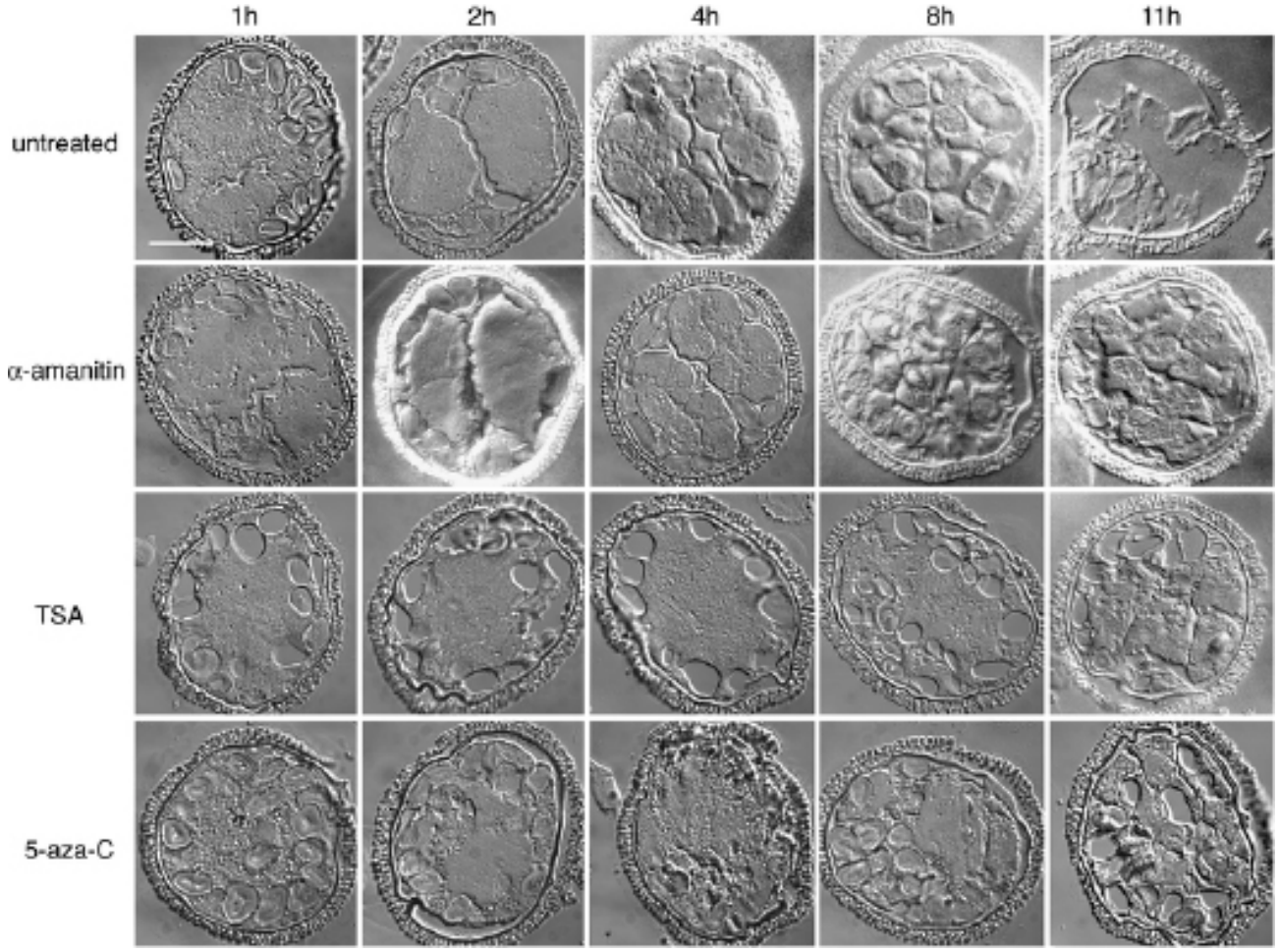
Bernhard, 1971; Cmarko *et al.,* 1999). Therefore, the current prevailing theory is that nuclear speckles mainly function as sites of storage, accumulation, and maturation of splice factors, and serve to bring high concentrations of these components close to their eventual site of activity (review: Lamond and Spector, 2003). Nevertheless, a small contingent of researchers contend that speckles play a direct role in splicing (for review: Hall *et al*., 2006). Recently, evidence has been published that could unify these two seemingly disparate theories. Girard and coworkers (2012) produced evidence that ~80% of splicing is co-transcriptional, but that ~20% is post-transcriptional. They identified active spliceosomes within nuclear speckles, suggesting that post-transcriptional splicing occurs in these structures (Girard *et al*., 2012).

The developing microspore is a transcriptionally-quiescent system

In *eukaryotes* different RNA polymerases (RNApol) are utilized to transcribe different classes of RNA (review: Kornberg, 1999). RNApol-II is responsible for the transcription of pre-mRNA, many pre-snRNAs, and pre-miRNA (precursors of microRNA, Sims *et al*., 2004). RNApol-I is responsible for transcription of 45S pre-rRNA, while RNApol-III transcribes 5S rRNA and tRNAs (Russell and Zomerdijk, 2006; Dieci *et al*., 2007). RNApol-IV and V both participate in making siRNA (Herr *et al*., 2005; Wierzbicki *et al*., 2009). Many rapidly developing systems depend on little or no new transcription for the proper completion of specific phases of development (Raff *et al.,* 1972; Capco and Jeffery, 1979; Davidson, 1986; Rosenthal, *et al.,* 1993; Curtis et al., 1995). Inhibitory drugs have historically been used to assay for the requirement of *de*

novo transcription during development. In addition, the induction of transcription and resulting effects has also been used to assess the role of *de novo* RNA production in development (review: Bensaude, 2011). In cases where blocking transcription does not inhibit development while inducing transcription arrests development progress, it is often speculated that endogenous transcription is not occurring. The effects of inhibiting transcription have been assessed in the microspore of *M. vestita* (Hart and Wolniak, 1998; Hart and Wolniak, 1999). α-amanitin is an inhibitor of RNApol-II, but will also inhibit other RNA polymerases at higher concentrations (Kedinger *et al*., 1970; Lindell *et al*., 1970; Weinman and Roeder, 1974). Microspore were incubated with α-amanitin at 1 mM and 10 mM (within the range for specific RNApol-II inhibition) followed by analysis of their development and levels of protein production (Hart and Wolniak, 1998; Hart and Wolniak, 1999). In the absence of RNApol-II mediated transcription, cell divisions and differentiation of spermatids was seen to progress normally and both general and specific protein blots were identical to non-treated controls (Figure I-3). In addition to α-amanitin, the transcriptional inhibitor actinomycin D has also been used to assess the requirement for transcription during spermatogenesis in *Marsilea* (unpublished results). Actinomycin D is a nonspecific inhibitor of DNA dependent RNA polymerases (Perry and Kelley, 1970). Like α-amanitin, microspores treated with with actinomycin D at the time of spore hydration divided and differentiated identically to control microspores (unpublished results). Reciprocal experiments have also been conducted; where drug induced transcription and its effects on spermatogenesis in *Marsilea* have been conducted (Wolniak *et al*., 2011). When transcription was artificially stimulated using TSA or 5-aza-C microspores did not complete their division cycles (Figure I-3).

Since spermatogenesis can progress in the absence of transcription, but not in the presence of induced *de novo* RNA production, the microspore of *M. vestita* has been



considered a transcriptionally silent system (review: Wolniak *et al*., 2011).

Figure I-3. *The developing male gametophyte of M. vestita is transcriptionally quiescent.* Gametophytes were grown for 1, 2, 4, 8, and 11 h at 20°C under various experimental conditions as indicated in the figure. Untreated gametophytes develop fully, and by 11 h many have released their mature spermatozoids, so the spore appears to be empty. α-Amanitin treatments develop normally and almost completely. At 8 hours of development in the presence of α-amanitin, the ciliary apparatus is being assembled, but at 11 h the spermatozoids have not been released from the antheridia. Gametophytes grown in the transcriptional activators, trichostatin A (TSA), and 5-aza-cytidine (5-aza-c), fail to develop for the first 8 h, but by 11 h some abnormal divisions are apparent. Bar = 25 μm. Wolniak *et al*., 2011.

Spermatogenesis in M. vestita is regulated post-transcriptionally

Instead of *de novo* transcription, developing spermatids rely on the translation of stored transcripts. Hart and Wolniak demonstrated that the dry microspores of *M. vestita* contain large quantities of stored proteins and mRNAs. Furthermore, translation of stored mRNA is essential for the development of the *Marsilea* male gametophyte (Figure I-4) Hart and Wolniak, 1998; Hart and Wolniak 1999).

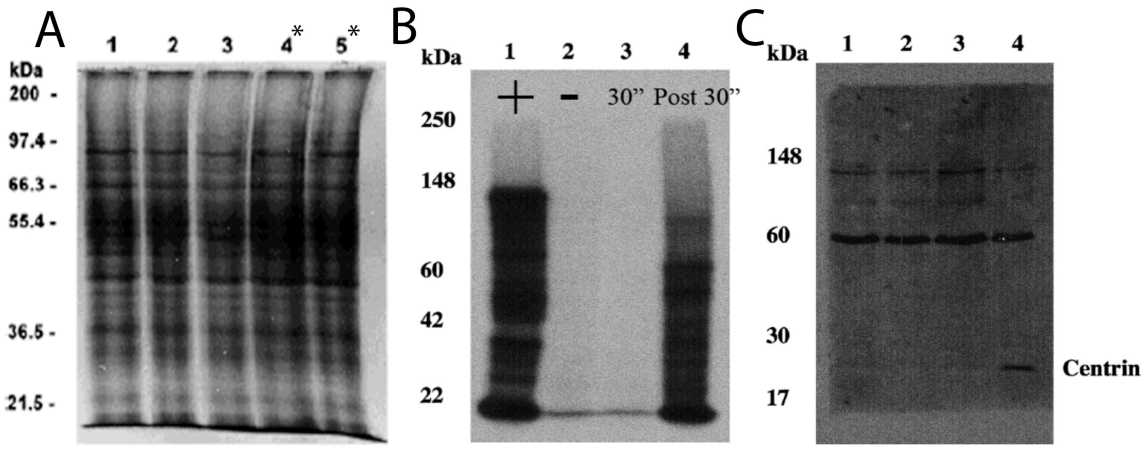


Figure I-4. The translation of a subset of stored transcripts is essential for spermatogenesis in *M. vestita*. (**A**), SDS-PAGE blots of total protein isolates made from: **1**) control, **2**) 1mM α-amanitin treated, **3**) 10mM α-amanitin treated, **4**) 1mM cycloheximide treated, and **5**) 10mM cycloheximide treated microspores. * denotes inhibited development. (**B**), SDS-PAGE blots of total protein resulting from *in vitro* translation of **1**) positive control (wheat germ RNA), **2**) negative control (no RNA), **3**) RNA isolated from *M. vestita* microspores prior to 30 minutes of development, and **4**) RNA isolated from *M. vestita* microspores after 30 minutes of development samples. (**C**), SDS-PAGE blot for Mv-Cen1 protein from *in vitro* translation of **1**) positive control (wheat germ RNA), **2**) negative control (no RNA), **3**) RNA isolated from *M. vestita* microspores prior to 30 minutes of development, and **4**) RNA isolated from *M. vestita* microspores after 30 minutes of development samples. Hart and Wolniak, 1998 & 1999.

Newly translated proteins were detectable as early as 2 hours after hydration. Specific proteins show different patterns of expression throughout spermatogenesis, with some being ubiquitously present during development, while others are present but with fluctuating abundances, and some not present in the dry spore but becoming abundant later (Figure I-5; Klink and Wolniak, 2001; 2003; Hart and Wolniak, 1998; Wolniak *et al.,* 2000).

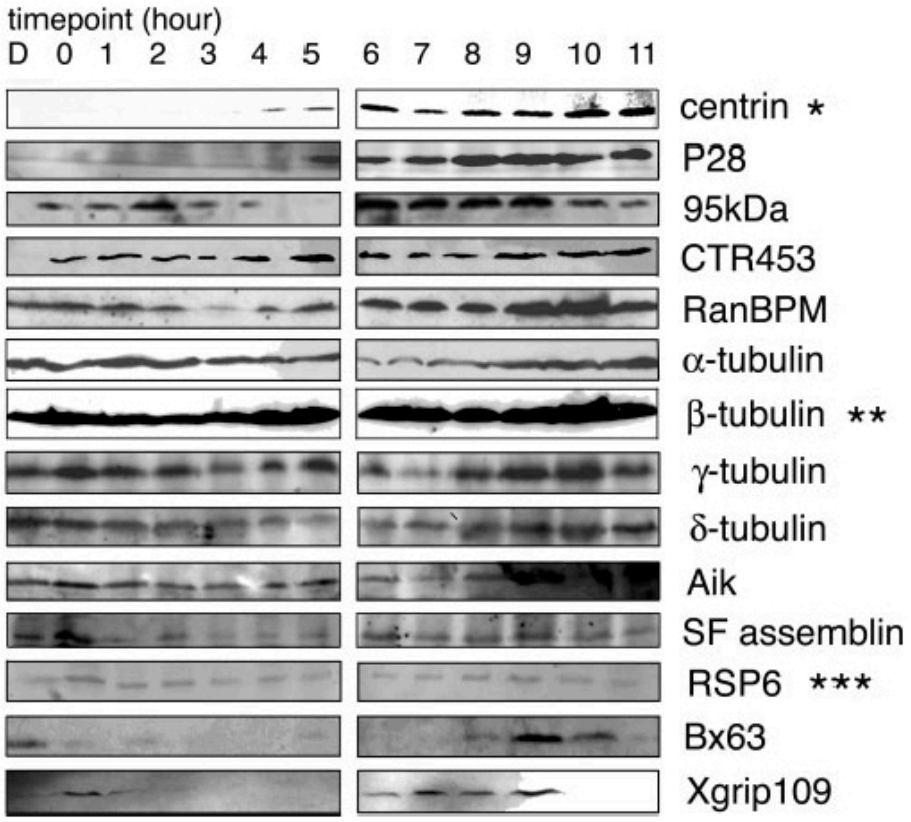
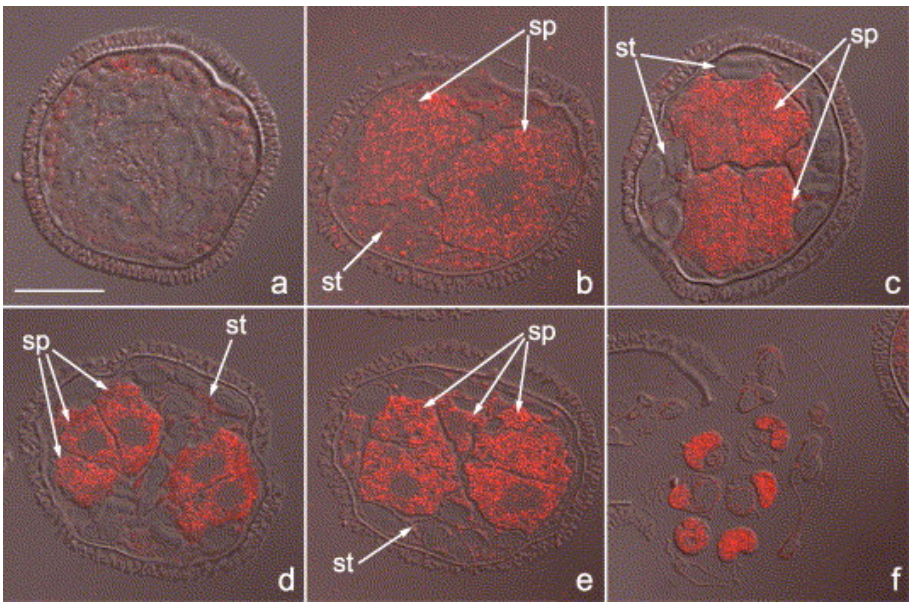


Figure I-5. *Antibodies directed against various centrosomal, cytoskeletal and axonemal proteins bind on immunoblots to similar polypeptides isolated from M. vestita.* M. vestita microspores were grown at 4°C for 30 min and transferred to 20°C and grown for various times up to 11 h. Soluble protein isolates were obtained and denatured by boiling in SDS. Thirty micrograms of protein from each isolate were separated electrophoretically on 10% SDS-polyacrylamide gels and blotted onto PVDF membranes. These immunoblots were probed with the antibodies listed in Table I. Here, immunoblots are presented for anti-centrin (Wolfrum and Salisbury, 1998); anti P28 (LeDizet and Piperno, 1995); anti-p95kDa (Geimer et al., 1998); anti-CTR453 (AKAP450) (Witczak et al., 1998); anti-RanBPM (Nakamura et al., 1998); anti-alpha-tubulin anti-beta-tubulin (Amersham); anti-gamma-tubulin, (Oakley et al., 1990); anti-delta-tubulin, (Dutcher et al., 1998); anti-Aik, (Kimura et al., 1997); anti-SF assemblin (Weber et al., 1993); anti-RSP6, (Curry et al., 1992); anti-Bx63, (Whitfield et al., 1988); anti-Xgrip109, (Martin et al., 1998). *: Centrin as the representative Group A antigen; **: beta-tubulin as the representative Group B antigen; ***: RSP6 as the representative Group C antigen. These proteins were detected either by ECL on Kodak film or by ECF with a Storm 840 Phoshorimager as described in the text. Klink and Wolniak, 2003.

Intriguingly, *in vitro* translation assays demonstrate that the mRNA isolated during the first 30 minutes of development could not undergo translation. However, transcripts isolated after 30 minutes of development were translated efficiently (Figure I-4; Hart and Wolniak, 1998).

Analysis of specific mRNAs revealed that transcripts display differential patterns of localization. Many transcripts are seen to be diffuse throughout the entire microspore, present in both sterile and spermatogenous cells, while others are confined solely to spermatogenous or jacket cells. Additionally, the localization of certain transcripts changes throughout development (Tsai *et al*., 2004; Deeb, 2009). Visualization of poly(A) RNA by *in situ* hybridization assays shows that these mature transcripts are sublocalized almost exclusively in spermatogenous cells after the jacket cell divisions have occurred. Proteins essential for RNA maturation and translation also display similar patterns of localization during development: for example, cytoplasmic poly-A-polymerase protein is detectable almost exclusively in spermatogenous cells (Tsai *et al*., 2004).

Preliminary evidence suggests that post-transcriptional modification of RNA may play a role in spermatid differentiation. Both protein and RNA encoding factors involved in polyadenylation (Figure I-6) and splicing (Figure I-7) of transcripts are predominantly localized in spermatogenous cells (Tsai *et al*., 2003).



The

Figure I-6. Immunolocalizations of gametophytes using anticytoplasmic poly-A RNA polymeras antibody (anti-PAP) during spermiogenesis. (**a**) Just after spore hydration, there is a weak antibod label scattered throughout (**b**) At 2 h of development, the anti PAP antibody label is mor cells (sp) of the gametophyte, and les abundant in the sterile cel n area in the spermatogenous cell o the right is occupied by a the spermatogenous cells (sp) exhibi high levels of anti-PAP a hacket cells (st) exhibit almost n detectable labeling with th pment, the spermatogenous cells (sp are heavily labeled (the nu development, the PRP-19 mRNA appears to be mor exhibit almost no staining concentrated in the central portion of the gametophyte intense in the cytoplasmic the antibody) and the sterile cells (st PAP antibody labeling is particularly (**C**, **D**) At 6 h of development, the PRP-19 mRNA almost entirely localized in the spermatogenous cell (sp) of the gametophyte, though a small amount o labeling is sometimes detectable in the sterile jacke cells (st). (**E**) RNAi treatment, employing dsRNA mad from MvU89 (PRP19) (**E**), reveal anomalies in the siz of some of the spermatids, but the numbers of cel division cycles in the gametophyte appear to be normal Scale bar = 25 µm. Tsai *et al*., 2004.

polyamine spermidine plays a

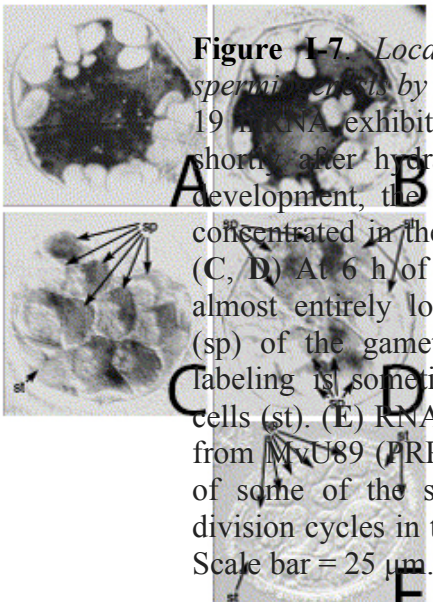


Figure I-7. *Localizations of PRP-19 mRNA during spermiogenesis by in situ hybridization assays.* (**A**) PRP 19 mRNA exhibits a uniform cytoplasmic distributio shortly after hydration of the spores. (**B**) At 1 h o development, the PRP-19 mRNA appears to be mor concentrated in the central portion of the gametophyte (**C**, **D**) At 6 h of development, the PRP-19 mRNA almost entirely localized in the spermatogenous cell (sp) of the gametophyte, though a small amount o labeling is sometimes detectable in the sterile jacke cells (st). (**E**) RNAi treatment, employing dsRNA mad from MvU89 (PRP19) (**E**), reveal anomalies in the siz of some of the spermatids, but the numbers of cel division cycles in the gametophyte appear to be normal Scale bar = 25 µm. Tsai *et al*., 2004.

role in the endogenous unmasking of stored transcripts

Spermidine is a polycation found in both *pro* and *eukaryotes* (Cohen, 1998; Tabor and Tabor, 1984). Spermidine, along with other polyamines are known to influence rapid cell growth, differentiation, transcription, protein synthesis, cell divisions, development of reproductive organs, and embryogenesis (Yatin, 2002; Kakkar abd Sawhney, 2002). In *Marsilea,* spermidine is detectable mainly in jacket cells up to 2 hours post-imbibition. At 6 hours of development, levels of spermidine within spermatogenous cells has increased, and by 8 hours spermidine is abundant throughout all regions of the microspore (Figure I-8 and I-9; Deeb *et al*., 2011). Spermidine synthase transcripts display a pattern of localization that presages the distribution of the polyamine. Spermidine synthase transcripts become localized within jacket cells through the first 4 hours of development. By 6 hours, spermidine synthase mRNAs appear in *in situ* hybridization assays as dense punctate localizations of particles in the cytoplasm of spermatogenous cells, and by 8 hours, spermatogenous cells exhibit intense cytoplasmic labeling of the *in situ* probe, indicating the presence of large amounts of this transcript in the spermatogenous cells (Figure I-8 and I-9; Deeb, 2009). Since this increase in labeling occurs in the presence of α-amanitin, it is likely that the transcripts were not newly transcribed in these cells, but rather, were present as stored mRNAs (and inaccessible to the *in situ* probe because of transcript masking).

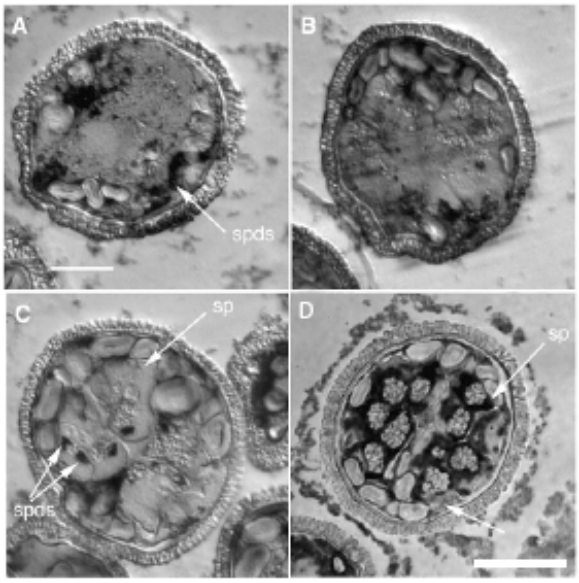


Figure I-8. *Temporal distribution of SPDS transcripts during spermatogenesis in M. vestita.* Gametophytes were allowed to develop normally for variable time increments, fixed, embedded in methacrylate and sectioned. *In situ* hybridization assays for spermidine synthase were performed on thin sections (1-2 µm). Images were acquired with a bright field microscope. (A) 2 h sample. (B) 4 h sample. (A, B) Spermidine synthase tra n scripts were detected in the jacket cells only up until 4 h of development. (C) 6 h sample. Spermidine synthase transcripts became detected in distinctly localized dots in the spermatogenous cells (yellow arrowhead). (D) 8 h sample. Transcripts were detected throughout the microspore. **SP:** Spermatogenous cells. **JC:** Jacket cells. Bar = 20 µm. Adapted from Deeb *et al.*, 2010.

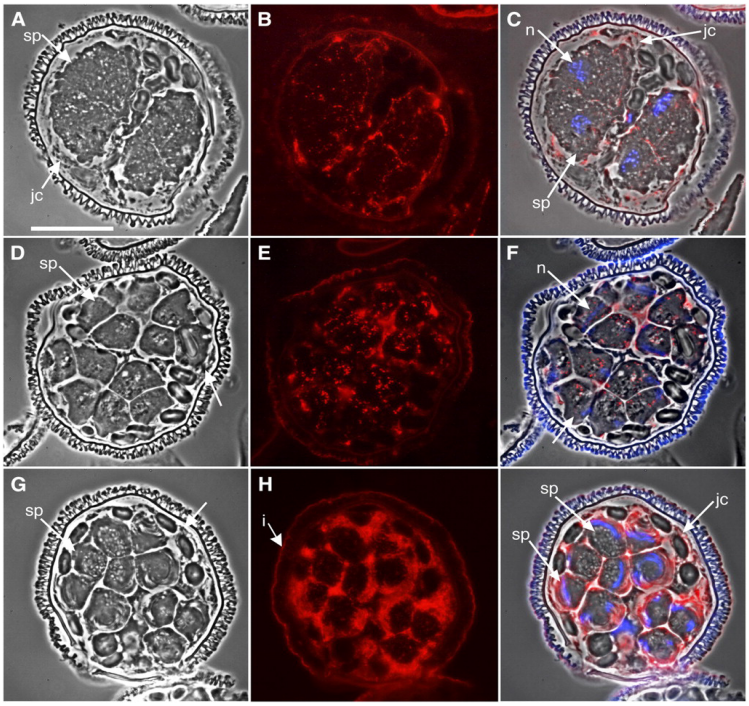


Figure I-9. *Immunolocalizations in Normal Gametophytes Show That Spermidine Levels Increase Dramatically in the Spermatids after They Are Formed*. Gametophytes were allowed to develop normally for various time intervals (**A**-**C**, 2hr; **D** to **F**, 6hr; **G** to **I**, 8hr) and then fixed, embedded in methacrylate, and sectioned. sp, spermatogenous cells; jc, jacket cells; n, nucleus. Bar = 25 µm. (**A**), (**D**), and (**G**) Phase contrast images showing the morphology of the gametophytes. (**B**), (**E**), and (**H**) Immunolabeling with antispermidine primary antibody and Alexa Fluor 594–conjugated secondary antibodies (red). (**C**), (**F**), and (**I**) Phase contrast images (gray) were overlaid with antibody (red) and DAPI nuclear staining (blue) images to establish the relative localizations of spermidine in the gametophyte. The spore wall autofluoresces brightly (i, intine). Deeb *et al.*, 2010.

Exogenous additions of spermidine cause a halt in development. Concurrent with perturbed gametophyte development is the apparent and precocious unmasking of transcripts within the nuclei of undeveloped spores (Figure I-10). Spermidine and other polyamine additions were found to unmask several transcripts including; spermidine synthase, PRP-19, gamma tubulin, centrin and MvU-620 (Deeb, 2009). Taken together, these data indicate that the post-transcriptional regulation of stored transcripts may play an important role in spermatogenesis in *M. vestita*.

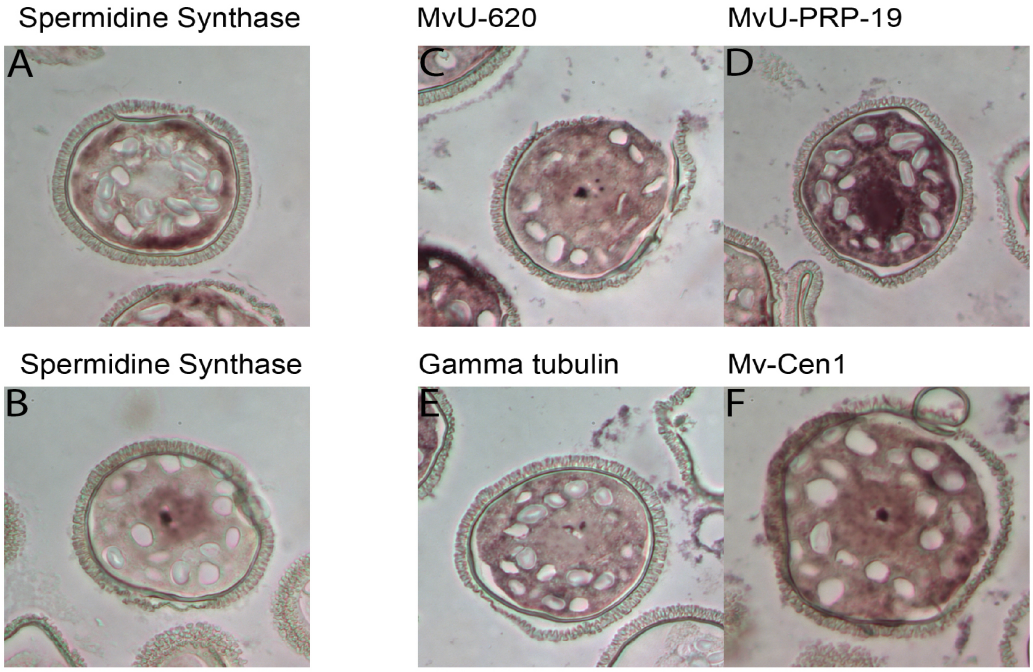


Figure I-10. *Exogenous additions of SPD result in the unmasking of subnuclear RNA*. (**A**) Representative control microspore labeled with SPDS *in situ* probe. (**B-F**), microspores treated with 10 mM SPD, fixed, sectioned, and probed with *in situ* probes directed against the indicated targets. Adapted from Deeb, 2009.

Identification of post-transcriptional regulatory events mediating spermatogenesis in M. vestita

The observations that male gametophyte and gamete development does not rely on transcription but does require translation, involves the regulated distribution of splicing and differential polyadenylation factors, masking of stored RNA, and temporal control of translation, all point to the fact the post-transcriptional regulation of RNA could be a major mechanism mediating spermatogenesis. To address how the development male gametophyte regulates the use of stored RNA, I first wanted to identify as many species of RNA present during spermatogenesis in *M. vestita* as possible and to have some idea of their temporal distribution. To achieve this I performed deep sequencing on cDNA fragments derived from poly(A)+ RNA isolated from distinct time ranges during development. I assembled a reference transcriptome for the microspore of *M. vestita* as well as 3 time range specific transcriptomes. Analysis of these transcriptomes revealed major categories of enriched transcripts. I found that transcripts encoding splicing and nuclear speckle factors are highly enriched. Another lab member (Richard S. Zipper) and I identified a large subset of transcripts for which intron retaining and fully spliced isoforms exist in our reference transcriptome. Using molecular and cell biological approaches I demonstrate that fully spliced isoforms arise from IRTs in a spliceosome-dependent fashion. The retention of an intron in a transcript blocks translation, while the splicing of this intron promotes translation. I show that a subset of RNA stored in the microspore is associated with nuclear speckles. Nuclear speckles and associated RNA exit the nucleus and are only observed in spermatogenous cells under normal conditions. Intron retaining transcripts appear to be contained within this subset

of speckle associated RNA. Inhibition of splicing results in the nuclear retention of poly(A)+, total, and stabilized IRTs. Together, these findings indicate that intron retention and post-transcriptional splicing in nuclear speckles are mechanisms regulating the spatial and temporal utilization of stored transcripts in the developing male gametophyte of *M. vestita*.

Chapter II – Transcriptome sequencing, assembly, and analysis

Partially adapted from Boothby et al., Developmental Cell, 2013

Background

The complete compliment of transcribed products (RNA) within a cell or organism is known as a transcriptome. This includes mRNA and non-coding RNA as well as different isoforms that might be present as well as correlated information about the abundance of these transcripts. Unlike a genome, transcriptomes are not static and may change at different developmental stages or under different physiological conditions. Knowledge of an organism's transcriptome promises to improve our understanding of how different biological processes are mediated (Morozova *et al*., 2009).

To gain insight into the transcriptomes of whole organisms and cells, researchers have used various methods. Early Sanger-sequencing-based approaches have several limitations for modern large-scale transcriptome projects. These include their expense, time and effort required, and general lack of quantifiable data produced (Boguski *et al*., 1994; Gerhard *et al.,* 2004). To address quantification issues, serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE), and massively parallel signature sequencing (MPSS) approaches were developed (Velculescu *et al*., 1995; Kodzius *et al*., 2006; Brenner *et al*., 2000). However, these methods come with their own limitations, such as analyzing only a small portion of a particular transcript and loss of isoform specificity.

To gain relatively inexpensive, quantifiable data on transcripts, researchers have historically used various microarray approaches. High throughput hybridization of

labeled cDNA to oligo microarrays has proven useful in measuring relative abundance and isoform structures of transcripts. However, the downsides to this approach are poor signal to noise ratios because of the nature of nucleic acid hybridization, difficulty in comparing data from different experiments, and a reliance of existing sequence knowledge, which is not always available (Royce *et al*., 2007; Okoniewski and Miller, 2006).

RNAseq is a so-called 'next generation' sequencing approach that addresses many of the shortcomings of previous transcriptome production methods (Morozova *et al*., 2009). There are several RNAseq technologies currently in use, which in general, can be conducted without prior genomic sequencing, and are capable of generating precise isoform information. These methods produce negligible background, they provide quantifiable results, and being massively parallel, they can be conducted at relatively inexpensive costs.

For the past decade, the Wolniak lab has been involved in the construction of a cDNA library, which to date contains ~2800 sequenced cDNAs constituting ~1300 contigs of legitimate gene products. While this cDNA library has proven to be essential to many worthwhile studies, its coverage of transcriptomic diversity and transcript abundance information is lacking. The advent of RNAseq technology provided us with the opportunity to gain vastly greater insight into the diversity of molecular mechanisms necessary for spermatogenesis in the microspore of *M. vestita*.

The developing male gametophyte of *M. vestita* is particularly well-suited for such RNAseq experiments. When rehydrated at 20 ^{0}C, the male gametophytes contained

within the microspores of this fern develop synchronously, allowing for RNA isolates to be made for precise times or developmental periods. The microspore is transcriptionally quiescent, so that temporal fluctuations in RNA abundance can be assumed to be due to post-transcriptional regulation (*e.g.,* unmasking, differential polyadenylation, degradation, or sequestration). Additionally, since the *M. vestita* genome has yet to be sequenced, next generation RNAseq provides great enough coverage to construct *de novo* transcripts.

Introduction

To our knowledge, there is only one published RNAseq-based transcriptome study that has been conducted in ferns (Der *et al*., 2011). This study was done using RNA extracted from a mix of *Pteridium aquilinum* gametophytes (including vegetative and sexually mature male, female, and hermaphroditic gametes of various ages). In contrast to this relatively non-specific approach, we wanted to construct transcriptomes from precise non-overlapping time ranges of male gametophyte development in *M. vestita.* Not only would time range-specific transcriptomes greatly increase the number of known *Marsilea* transcript sequences, it would provide insights on temporal specificity of available mRNA abundance that we suspected would mirror the developmental use of these transcripts. More specific to my dissertation project, I hoped that transcriptome sequencing would help in identifying potential post-transcriptional regulatory mechanisms mediating the use of stored RNA during spermatogenesis. To construct these transcriptomes, we isolated and sequenced poly(A)+ RNA from 3 distinct time ranges covering both the division and differentiation phases of spermatogenesis. RNAseq reads were assembled using the *de novo* transcriptome assembly program Trinity (Grabherr *et*

al., 2011), allowing us to make both a combined reference transcriptome, and time range-specific transcriptomes. These transcriptomes were annotated with putative identities and estimated abundance data. This annotation was followed by Gene Ontology enrichment analysis to identify potential developmentally important categories of transcripts. As expected, we saw that the enrichment of transcripts involved in or associated with specific cellular processes, functions, or components mirrored their presumptive use during development. Furthermore, during the first half of spermatogenesis, Gene Ontology terms related to RNA splicing are highly enriched and revealed that microspores contain a subset of transcripts with retained introns, suggesting that post-transcriptional splicing may be a mechanism mediating the use of these stored transcripts.

Results

RNAseq

Poly(A)+ RNA was isolated from microspores grown for specific time ranges (1-2 h, 3-5 h, 6-8 h) in triplicate and prepared for Illumina RNAseq (Figure II-1). Multiplexed 100bp paired end RNAseq was carried out using the Illumina HiSeq 1000 platform at the IBBR DNA Sequencing Facility, housed in the Plant Sciences Building at the University of Maryland campus (http://www.ibbr.umd.edu/facilities/sequencing).

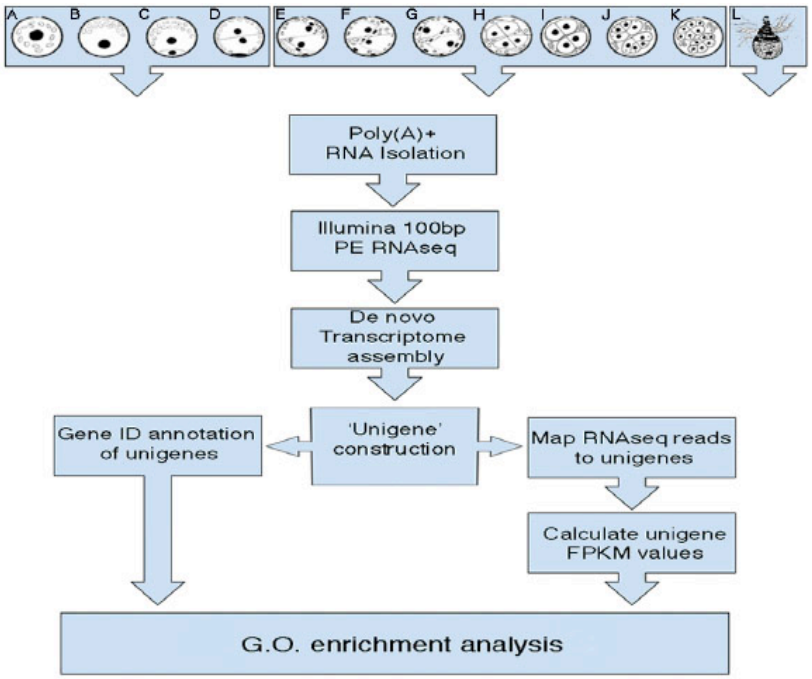


Figure II-1. *Pipeline for RNAseq, transcriptome assembly, and analysis.* Microspores were rehydrated and allowed to develop for 1-2 h (A-D), 3-5 h (E-K), and 6-8 h (L). Poly(A)+ RNA isolates were obtained from microspores within these time ranges. RNA was prepared for Illumina 100bp paired end sequencing and sequenced. Sequenced reads were assembled *de novo* using the program Trinity. For our reference transcriptome (1-8 h) unigenes were constructed first by selecting the longest isoform from each Trinity compilation group and then by assembling contigs from these isoforms using the program CAP3. Refseq_protein databases were used to annotation unigenes with putative identities. RNAseq read mapping was used to assign FPKM values to each unigene. Finally, annotations were used in conjunction with the program GOrilla to conduct gene ontology enrichment analysis. Putative identity and FPKM annotations were assigned for time specific transcriptomes, for which gene ontology enrichment analysis was also carried out. T.C.C. Boothby, unpublished.

The average number of 100bp reads generated from each sample was 65,6444,004. For *de novo* transcriptome assembly, all RNAseq sample reads were concatenated and quality-filtered (Figure II-2). These reads were used as input for the *de novo* transcriptome assembly program Trinity (Figure II-3; Grabherr *et al*., 2011). Time range specific transcriptomes were assembled using RNAseq reads from 1-2, 3-5, or 6-8 h. For construction of our reference transcriptome (spanning 1-8 h) 590,796,040 paired end reads were assembled into 234,427 transcripts with an average length of 1,162 nucleotides resulting in an estimated coverage of 728X.

```
#!/bin/bash

##Concatenate RNAseq reads
cat 1-2hr_2_R1* > 1-2hr_2_R1_combined.fastq.gz.Z
cat 1-2hr_2_R2* > 1-2hr_2_R2_combined.fastq.gz.Z

##Filter concatenated files
zcat 1-2hr_2_R1_combined.fastq.gz| grep -A 3 '^@.* [^:]*:N:[^:]*:' > 1-
2hr_2_R1_combined.filtered.fastq
zcat 1-2hr_2_R2_combined.fastq.gz| grep -A 3 '^@.* [^:]*:N:[^:]*:' > 1-
2hr_2_R2_combined.filtered.fastq
```

Figure II-2. *Representative script for concatenation and quality filtering of Illumina RNAseq reads.* Boothby *et al*., 2013.

```
#!/bin/bash

##Run trinity
ulimit -s unlimited

../Trinity.pl --seqType fq --left 1-2hr_2_R1_combined_filtered.fastq --right 1-
2hr_2_R2_combined_filtered.fastq --SS_lib_type RF --paired_fragment_length 280  --
min_contig_length 305 --run_butterfly --output 1-2hr_2_output.dir --CPU 3 --
bflyHeapSpace 95360M --bfly_opts "--compatible_path_extension --stderr "
```

Figure II-3. *Representative script for running Trinity*. *Boothby* et al*., 2013.*

In order to reduce potential artifacts and construct reference transcripts ('unigenes'), that would aid in downstream analysis and cross-referencing, we used a multistep contig assembly approach. The Trinity *de novo* assembly program generates and groups isoforms (Grabherr *et al*., 2011). We selected the longest transcript from each isoform group, which reduced our transcriptome to 127,744 transcripts with an average length of 803 bases. These transcripts were further condensed using the contig construction program CAP3 (Huang and Madan, 1999). CAP3 contig assembly resulted in 34,704 contigs with an average length of 1210 bases and 48,502 singlets with an average length of 544 bases for a total of 83,206 unigenes with an average length of 822 bases and estimated coverage of 864X.

Assigning Putative Identities

To assign putative identities to sequences from our reference transcriptome, NCBI refseq_protein databases for *Arabidopsis thaliana* and *Homo sapiens* were downloaded and used for BLASTx analysis. We opted to use these two databases since downstream analysis would require Gene Ontology (G.O.) annotation. *A. thaliana* is the plant model with the best available G.O. annotation, but since this organism does not produce ciliated male gametes, we reasoned it was necessary to include in our database comparisons, annotated organisms that have the capacity to make ciliated cells. Obviously, *M. vestita* transcripts will be more distantly related to *H. sapiens*, but because of the level of G.O. annotation associated with *H. sapiens* reference sequences, we opted to use this database as well. The *A. thaliana* database contained 34,875 reference sequences and BLASTx

analysis resulted in 23,919 best hits with E-value less than 1E-10. The *H. sapiens* database contained 35,668 reference sequences and BLASTx analysis resulted in 11,600 best hits with E-values less than 1E-10. It was reassuring to see that in assigning putative identities, our transcriptome shared more components with another plant than with a metazoan.

Gene Ontology Enrichment Analysis

G.O. enrichment analysis was performed for best hits with E-values below 1E-10. Background adjusted G.O. enrichment was performed using the GOrilla G.O. enrichment analysis program (Eden *et al*., 2009). Enrichment analysis of *M. vestita* transcripts with best hits to *A. thaliana* sequences associated with G.O. terms resulted in 1,199 overrepresented and 12 underrepresented G.O. terms (Appendices i-v). Enrichment analysis of *M. vestita* transcripts with best hits to *H. sapiens* sequences associated with G.O. terms resulted in 876 overrepresented and 499 underrepresented G.O. terms (Appendix vi-xi). Enriched G.O. categories of interest (Figure II-4) included categories associated with post-transcription regulation of RNA (*e.g.,* RNA processing P-value 5.08E-81, mRNA catabolic processes P-value 3.13E-50, ncRNA processing P-value 1.37E-36, RNA splicing P-value 8.88E-36, spliceosomal complex P-value 3.34E-30, nuclear speckle P-value 2.96E-07), Cell cycle regulation (*e.g.,* mitotic cell cycle P-value 3.93E-26, M/G1 transition of mitotic cell cycle P-value 6.11E-17), cytoskeleton and cilia (*e.g.,* microtubule organizing center P-value 3.35E-07, centrosome P-value 3.17E-05, cilium P-value 2.01E-04) as well as negative regulation of transcription (*e.g.,* negative

Overrepresented			
	G.O. term	Description	P-value
PTR of RNA	GO:0006396	RNA processing	5.08E-81
	GO:0006402	mRNA catabolic process	3.13E-50
	GO:0008380	RNA splicing	8.88E-36
	GO:0043631	RNA polyadenylation	1.42E-04
	GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	1.26E-06
PTR of proteins	GO:0043687	post-translational protein modification	6.13E-10
	GO:0000151	ubiquitin ligase complex	6.48E-15
	GO:0051084	'de novo' posttranslational protein folding	3.74E-09
	GO:0006625	protein targeting to peroxisome	8.47E-05
	GO:0006446	regulation of translational initiation	1.31E-04
Cell cycle	GO:0000278	mitotic cell cycle	3.93E-26
	GO:0000084	S phase of mitotic cell cycle	4.52E-26
	GO:0000216	M/G1 transition of mitotic cell cycle	9.81E-18
	GO:0000082	G1/S transition of mitotic cell cycle	6.11E-17
Cytoskeleton/Ciliary	GO:0007017	microtubule-based process	1.64E-19
	GO:0005815	microtubule organizing center	3.35E-07
	GO:0005813	centrosome	3.17E-05
	GO:0005929	cilium	2.01E-04
Negative regulation of transcription	GO:0031048	chromatin silencing by small RNA	2.70E-12
	GO:0045814	negative regulation of gene expression, epigenetic	2.96E-18
Underexpression			
	G.O. term	Description	P-value
Positive regulation of transcription	GO:0045944	positive regulation of transcription from RNA polymerase II promoter	9.73E-09
	GO:2001141	regulation of RNA biosynthetic process	2.06E-10
	GO:0045893	positive regulation of transcription, DNA-dependent	1.02E-06
Transcription factors	GO:0000976	transcription regulatory region sequence-specific DNA binding	3.99E-08
	GO:0000975	regulatory region DNA binding	2.73E-06
	GO:0001067	regulatory region nucleic acid binding	2.73E-06

Figure II-4. Representative o*ver and underrepresented G.O. terms of interest.* Presented here are examples of representative G.O. terms that were either over or underrepresented in our reference transcriptome. For full lists of G.O. terms see Appendices i-xi. T.C.C. Boothby, unpublished.

regulation of gene expression, epigenetic P-value 2.96E-18). Underrepresented G.O. terms comprised many categories that one would expect to not be associated with plant gametophyte development (*e.g.,* skeletal system morphogenesis P-value 4.16E-06) but also many categories involved in positive transcriptional regulation (*e.g.,* DNA-dependent positive regulation of transcription P-value 1.02E-06, transcription factor regulatory region DNA binding P-value 2.73E-06). The overall theme that emerges from this analysis is that the microspore stores RNA involved in major cellular and morphological changes that occur during spermatogenesis such as cell cycle progression,

cytoskeletal formation, and ciliogenesis. In addition, transcripts encoding proteins involved in RNA processing are present. This analysis reconfirms that little, if any, transcription takes place during development. Underrepresentation of positive regulators of transcription and enrichment of negative regulators of transcription underscore the fact that the microspore is a determinate system whose development is mediated post-transcriptionally.

Time range-specific transcriptomes

In contrast to our reference transcriptome, which provides a broad overview of the transcripts present during spermatogenesis, time range-specific RNAseq allowed us to construct and analyze temporally restricted transcriptomes. We reasoned that these transcriptomes should consist of only transcripts that have been unmasked, polyadenylated and not degraded and therefore these transcripts likely mediate progression through or processes occurring during the time range(s) in which they are found.

Assembly of these transcriptomes was conducted using the same quality control filtering and assembly parameters that were used for construction of our reference transcriptome (Figure II-2 and II-3). However, unlike with our reference transcriptome, time range-specific transcriptomes were not subjected to additional contig construction steps, since we wanted to preserve as much temporal information about isoforms as possible. Our time range transcriptomes consisted of 56,744, 77,882, and 85,133 transcripts (1-2 h, 3-5 h, 6-8 h, respectively).

Time range-specific analysis – Quantification, identity assignment and G.O. enrichment analysis

To aid in consistency and cross-referencing of quantification between time ranges, we estimated time range FPKM values by using our reference transcriptome as a mapping scaffold for abundance comparisons. Time range specific RNAseq reads were mapped against our reference using Bowtie (Langmead *et al*., 2009). Bowtie read mapping was used in conjunction with Cufflinks to estimate transcript abundances (Trapnell *et al*., 2010). As with our reference transcriptome, time range-specific transcripts were assigned putative identities generated using BLASTx searches against *A. thaliana* and *H. sapiens* refseq_protein databases. Transcript identities were cross-referenced with Cufflinks abundance estimations in order to create ranked lists for G.O. enrichment analysis. Both *A. thaliana* and *H. sapiens* best hits with E-values below 1E-10 were used to carry out ranked G.O. enrichment analysis using GOrilla (Eden *et al*., 2009).

Ranked G.O. analysis of *H. sapiens* gene identities (7,507 with G.O. terms) for our 1-2 h time range showed that 267 process, 35 functional, and 36 cellular component gene ontology terms are enriched (Appendix xii-xiv). The same analysis of our 3-5 h transcriptome (8,414 identities with G.O. terms) showed enrichment in 300 processes, 33 functional, and 39 cellular component Gene Ontology terms (Appendix xv-xvii). Finally, our late time range (6-8 h) transcriptome showed enrichment for 118 process, 41 functional, and 20 cellular component terms from 9,579 identities with G.O. term annotations (Appendix xviii-xx).

Ranked G.O. analysis of *A. thaliana* gene identities (5,870 with G.O. terms) for our 1-2 h time range showed that 93 process, 8 functional, and 8 cellular component gene ontology terms are enriched (Appendix xxi-xxiii). Analysis of our 3-5 h transcriptome (6,688 identities with G.O. terms) showed enrichment in 54 processes, 8 functional, and 13 cellular component Gene Ontology terms (Appendix xxiv-xxvi). Finally, our late time range (6-8 h) transcriptome showed enrichment for 16 process, 5 functional, and 11 cellular component Gene Ontology terms from 7,090 identities with G.O. terms (Appendix xxvii-xxix). For all temporal transcriptome ranked G.O. enrichment analysis, a P-value threshold of 10E-4 was used as a cutoff.

The results from this study are clearly extensive and should provide future researchers with ample avenues of interest to pursue. Extensive molecular analysis and reverse genetic dissection of spermatogenesis in *M. vestita* will be necessary to confirm these findings and such analysis is beyond the scope of this dissertation. Complete lists of time range enriched terms can be found in Appendices xxii-xxix. The following paragraphs provide a cursory qualitative analysis and interpretation of the data generated from our temporal G.O. enrichment study, while the subsequent chapters serve as experimental validation and expansions of several RNAseq derived findings.

By cross-referencing enriched G.O. terms across time ranges, it is possible to obtain a snapshot of what processes, functions, and cellular components are potentially essential for a given developmental time range. In many cases G.O. terms associated with a particular developmental period were not surprising, but seemed to mirror cellular events occurring at that time (Figure II-5). For example, G.O. terms associated with mitotic cell divisions (*e.g.,* MCM, origin of replication, and spindle) were exclusively

enriched in the 1-2 h time range while terms associated with ciliogenesis (*e.g.,* cilia and axoneme) were exclusively enriched in the 6-8 h time range.

Despite these seemingly obvious results, we were surprised by several initial findings from our G.O. enrichment analysis. The first surprise was a general lack of G.O. component enrichments exclusive to the 3-5 h time range. This is the time range when the transition from the division to differentiation phase occurs and when the blepharoplast arises and forms basal bodies. Our initial expectation was that we would see G.O. terms involved in these processes exclusively enriched during this time period. However, our data indicate that this middle time range is a rich mix of 1-2 h and 6-8 h transcripts. G.O. terms that are enriched exclusively within the 3-5 h time range are associated with turnover of proteins (*e.g.,* proteasome and ubiquitin ligase, Figure II-5). We believe this indicates that one of the biological functions unique to the middle time range is the destruction of biological material required for the division phase of spermatogenesis but not requisite for later differentiation of the spermatids. Turnover might be essential for this transition either by halting ongoing processes and/or by recycling biological macromolecules required for later processes.

Another surprise was that G.O. terms enriched during the first half of spermatogenesis (1-5 h) included those involved in splicing of pre-mRNA (*e.g.,* catalytic spliceosome and nuclear speckles, Figure II-5). In retrospect, this might have been predicted, as previous studies had identified a few factors related to splicing in the microspore of *M. vestita* (*e.g.*, PRP19 and Mago nashi, Tsai *et al*., 2003; van der Weele *et al*., 2007). However, if enrichment within the transcriptome is any indication of developmental importance, then these results reaffirm the idea that spermatogenesis relies

extensively on the maturation of stored RNA. Not only are splicing G.O. terms enriched, they are statistically the most enriched groups during this time interval. Terms associated with splicing of RNA are the first 6 most confidently scored process enrichments at 1-2 h (P-values 3.97E-27 to 5.34E-22). Likewise, splicing associated G.O. terms are the second and third most enriched cellular components during the 1-2 h time range (spliceosomal complex and catalytic step 2 spliceosome, P-values of 3.61E-21 and 5.29E-16, respectively).

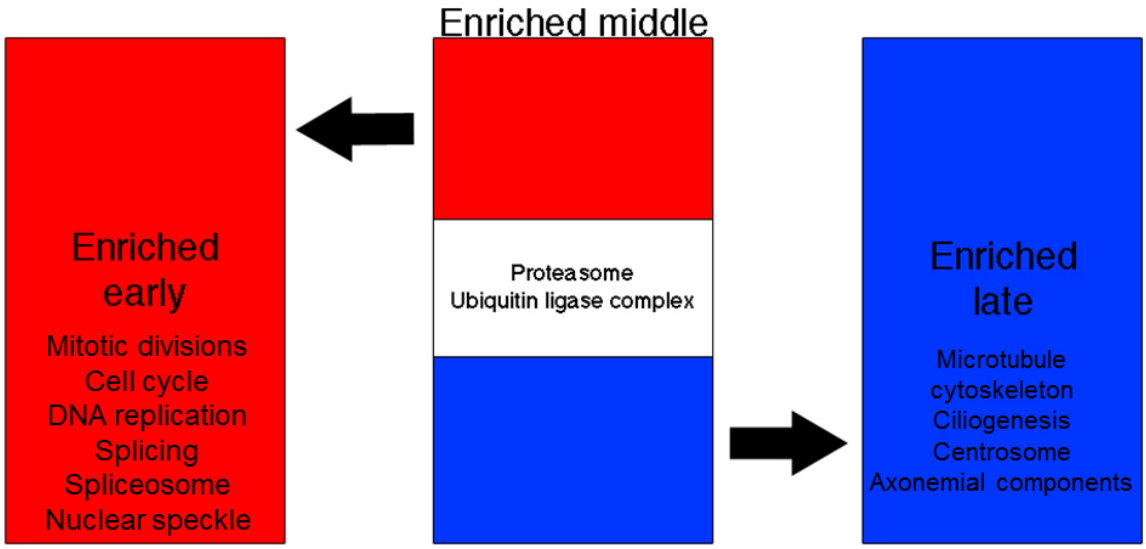


Figure II-5. *G.O. enrichments correspond to specific time ranges and correlate with development processes occurring at those times.* Enriched G.O. terms unique to different time ranges correspond to developmental processes occurring during those time ranges. Enrichment of splicing terms occurs mainly between 1-5 h. T.C.C. Boothby, unpublished.

Identification of Intron Retaining Transcripts (adapted from Boothby et al., 2013).

The statistical overrepresentation of factors involved in splicing indicate that maturation of pre-mRNA is essential for spermatogenesis in *M. vestita*, and furthermore

that we would expect to find pre-mRNA present in our transcriptomes. To determine if *M. vestita* male gametophyte transcriptome contains pre-mRNA, we performed *in silico* analysis to identify pairs of intron retaining transcripts (IRTs) and their fully spliced isoforms. Since the *M. vestita* genome has not been sequenced or annotated, we were unable to utilize a mapping based approach for identifying IRTs (Galante *et al*., 2004). Instead, we opted to use and parse results from the Basic Local Alignment Search Tool, megaBLAST (similar to Wang *et al*., 2003; Figure II-6).

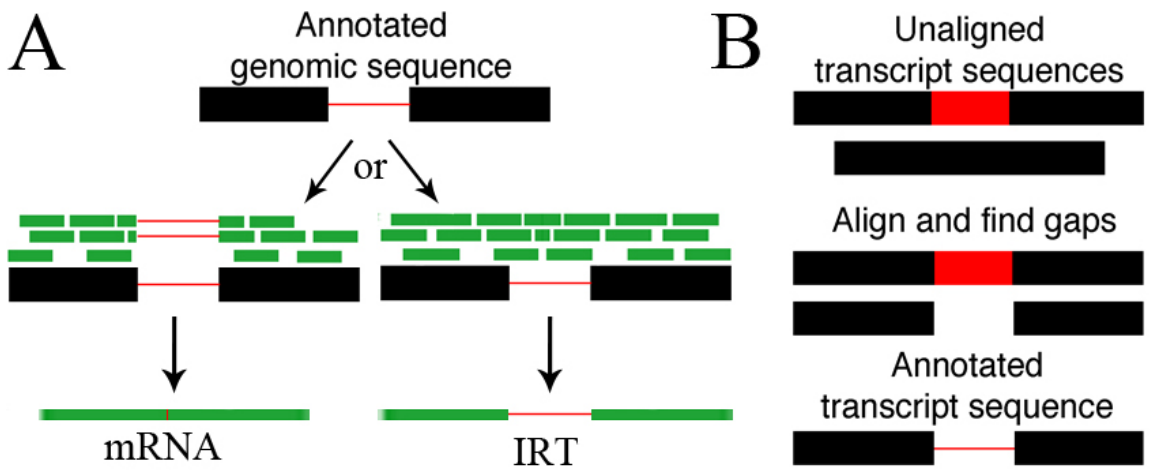


Figure II-6. *IRT identification schematic.* (**A**), schematic of an annotated genomic reference mapping based approach. Reads are mapped against a genomic reference with known exon:intron junctions. Mapped reads are examined for intron coverage as a means of identifying retained introns. (**B**), schematic of a BLAST and gapped alignment parsing based approach to identify retained introns. Trinity output transcriptomes are compaired with themselves. MegaBLAST alignments with gaps in them are selected as putative introns for downstream validation and analysis. T.C.C. Boothby unpublished.

We identified 12,589 retained introns and 10,115 IRT pairs. Retained introns had an average length of 179 bases with the shortest being 38 bases and the longest being 4,014 bases (Figure II-7).

To get a broader view of what functions IRTs might play in development, we performed gene ontology (G.O.) analyses of IRTs using the program Blast2GO (Conesa *et al*., 2005). We annotated our IRT database with G.O. terms for cellular components, cellular processes, and cellular functions. This analysis showed that IRTs are not unique for any particular process, function, or part of the cell. However, our analysis revealed that cellular components and processes associated with later stages of development had IRT representation, including IRTs present or associated with centrosomes, flagella, cilia, or axonemes, cell motility and locomotion, spermatozoid and male gamete differentiation, and cell death (Figure II-8).

Examining the number of introns in each IRT shows the majority (8,217/10,115) of IRTs appear to contain only 1 retained intron (Figure II-7). While most IRTs contain only 1 intron, we suspected that the genes encoding these transcripts have more then 1 intron. This could mean that newly transcribed pre-mRNAs undergo partial maturation during spore desiccation and that there is developmental control over the retention of specific introns and the removal of others. Genomic DNA was isolated and sequencing was conducted for four genes encoding IRT transcripts. We compared introns contained within the genomic sequences to those retained in IRTs (Figure II-7). In all cases examined (4/4), genes exhibited at least 1 intron that was not present in IRTs, demonstrating that the desiccating microspore is capable of pre-mRNA splicing. Additionally, when RNAseq reads were mapped to a subset of IRTs the average maximal coverage of mapped reads within retained introns was found to be 124X, suggesting that we are not missing other retained introns because of low sequencing coverage, or if we are, their retention rate is substantially lower than the retained introns identified by our

method. Taken together, these data indicate that the microspore of *M. vestita* retains specific introns within a subset of IRTs, which are stored during spore quiescence.

Typical U2 introns contain consensus sequences at their 5' and 3' ends, known as the donor and acceptor sequence, respectively (Mount, 1982; Mount *et al*., 1992; for review: Padgett *et al*., 1986). The donor sequence is characterized by an almost invariant GU dinucleotide at the start of the intron, while the acceptor sequence is characterized by an AG dinucleotide at the last two bases of the intron. For our presumptive introns, we looked for the presence and position of known U2 splice signals within 10+/- bases of both the 5' and 3' exon:intron junctions. We found a high degree of correlation between these conserved sequences and exon:intron junctions relative to a set of control (randomly generated) sequences (Figure II-7).

Additionally, we selected 17 representative IRTs for downstream analysis and constructed 5' and 3' *M. vestita* IRT consensus sequences (Figure II-7) that correspond well with conserved U2 dinucleotide signals, but may contain unique conservations at other bases. In particular the conservation of C at the +8 position, T at the +6 position, and W at the -6 position is interesting and if real might represent a potential regulatory signal.

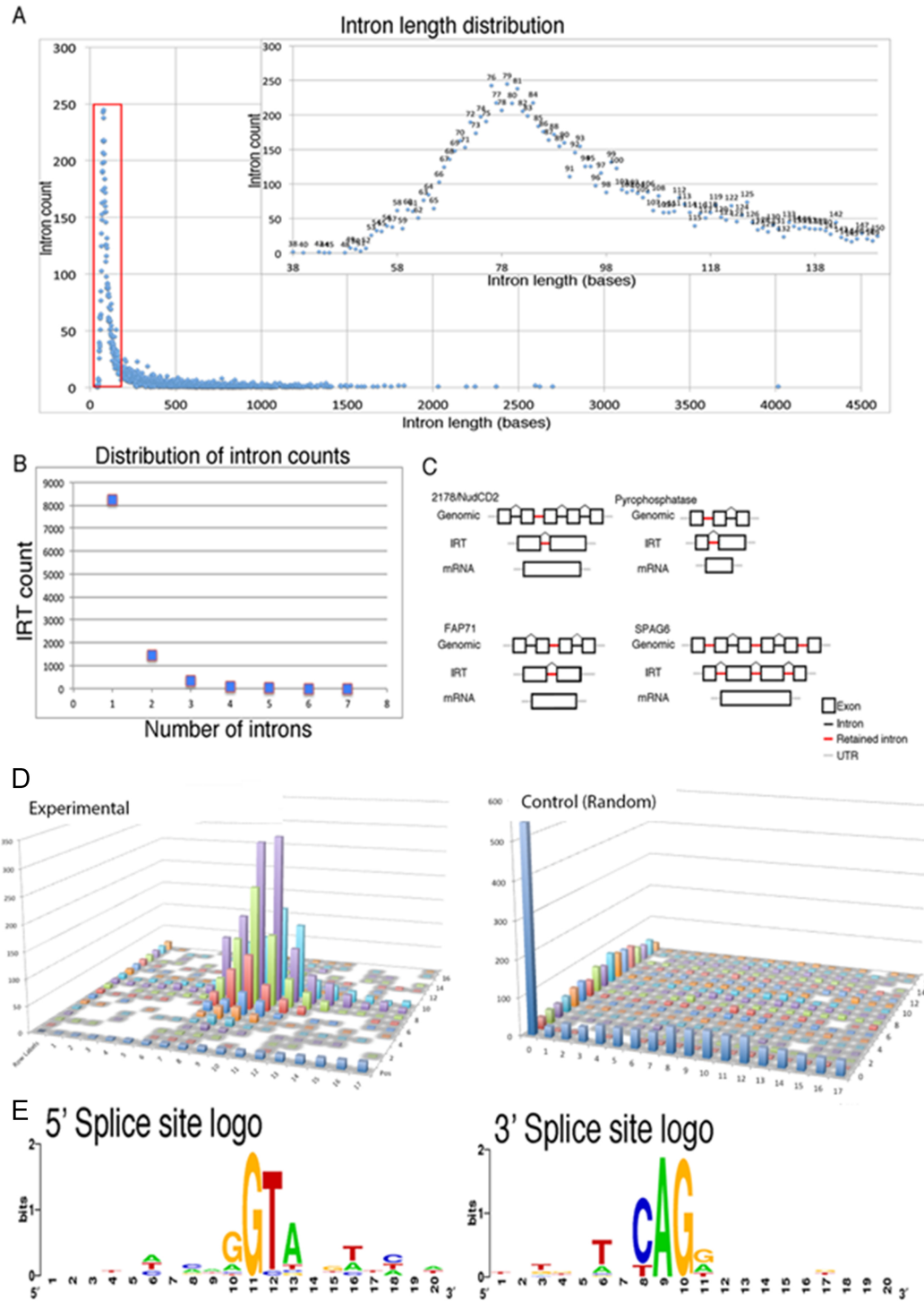


Figure II-7. *Identification of intron retaining transcripts (IRTs).* (**A**), Distribution of intron sizes. (**B**), Distribution of the number of retained introns per IRT. (**C**), Schematic representations of genomic, IRT, and fully spliced isoform architectures. (**D**), Correlation of conserved U2 splice signals with the ends of presumptive introns. R.S. Zipper, unpublished. (**E**), Consensus sequence of 5' and 3' splice signals constructed from representative IRTs. Boothby *et al.*, 2013.

Centrosomes
GO:0005813
GO:0031616
Flagella, cilia, axonemes
GO:0019861
GO:0030990
GO:0030992
GO:0009434
GO:0044442
GO:0020017
GO:0044460
GO:0005929
GO:0072372
GO:0031513
GO:0044447
GO:0005879
GO:0005858
GO:0042995
GO:0042384
GO:0001539
GO:0060271
Cell motility and locomotion
GO:0040011
GO:0048870
GO:0051234
Sperm and male gamete differentiation
GO:0048235
GO:0010480
GO:0045595
GO:0007276
GO:0048232
Cell death
GO:0010941

Figure II-8. *Representative IRT G.O. terms.* Boothby *et al*., 2013.

Conclusions/Discussion

Quality of our transcriptome

Our sequencing of RNA derived libraries from 3 time ranges during spermatogenesis in triplicate generated a reference transcriptome comprising 34,704 contigs and 48,502 singletons with an average transcript length of 822 bases and estimated coverage of 864X. By comparison, the *P. aquilinum* gametophyte transcriptome has an average estimated coverage of 7X (Der *et al*., 2011). We believe such deep coverage indicates that we have captured a sizeable portion if not all the transcripts present in the microspore of *M. vestita*. Future genomic sequencing will aid in this assessment.

To reduce errors in assembly resulting from poor quality reads, we conducted filtering on our concatenated RNAseq reads to remove those results that did not pass Illumina's automated quality control. We opted to generate 100bp paired end reads and used the paired end read setting for Trinity assembly. Paired end reads are only assembled into transcripts in cases where both mates fulfill incorporation requirements, adding an additional level of quality control. These approaches should greatly reduce incorrect assembly of RNAseq reads into *de novo* generated transcripts.

Identity assignment and G.O. enrichment analysis

The goal of this transcriptome assembly and analysis was threefold. Firstly, we sought to generate a more extensive transcriptome for the developing male gametophyte of *M. vestita*. Secondly, we aimed to identify potentially key mediators of spermatogenesis encoded by stored RNA in the microspore of *M. vestita* with temporal

specificity. Thirdly, and most importantly for the current study, we aimed to identify potential post-transcriptional regulatory mechanisms mediating spermatogenesis in *M. vestita*.

On all counts, our initial attempts have been successful. Our reference transcriptome consists of 83,206 transcript sequences (34,704 contigs and 48,502 singletons) which is a ~64 fold increase compared to our previous cDNA library. The assignment of identities through comparison of our sequences to curated reference protein databases allows us to predict putative functions for a large portion of transcripts that make up our reference transcriptome. G.O. enrichment analysis allows us to predict processes, functions, and cellular components that are essential for spermatogenesis in *M. vestita* with temporal specificity. Identity assignment and G.O. analysis indicates that splicing of pre-mRNA is a major post-transcriptional mechanism mediating development of male gametes in *M. vestita* and parsing of megaBLAST results identified a large subset of IRTs as well as their fully spliced isoforms.

There is clearly more that can be done to expand our knowledge of the transcriptome of male gametophyte of *M. vestita*. Our transcriptome was derived from poly(A)+ RNA isolated in 3 distinct time ranges and prepared in a non-directional fashion. As such, our current transcriptome provides no appreciable information about noncoding-RNA, micro-RNA, or antisense-RNA. Noncoding-RNA could be sequenced (with and/or without depletion of rRNA) using total RNA isolation methods, and specific protocols (with kits) exist for the isolation as well as sequencing of micro-RNA. Antisense-RNA could be identified through directional sequencing, which provides unambiguous information as to the 5’ and 3’ orientation of transcripts. In addition, more

temporally precise RNA isolation and sequencing could be performed, although the extent to which this would improve our understanding of male gametophyte development in *M. vestita* is probably less than the other options mentioned above. In addition to further RNAseq experiments, genomic sequencing of *M. vestita* would provide a better framework from which to investigate extents of alternative splicing and RNA modification within our existing transcriptome.

Additional analysis of our existing transcriptome could also offer additional biological insights on mechanisms regulating cytomorphogenesis. Since the male gametophyte of *M. vestita* is a determinate system that carries out a handful of specialized processes (*e.g.*, *de novo* basal body formation in a blepharoplast, and ciliogenesis) comparative transcriptomics with similarly specialized systems could help in identifying essential or new genes involved in these activities.

Gene ontology enrichments correspond to processes occurring during the developmental period in which they are found

In initiating our transcriptome assembly and analysis project, we had reasoned that different sets of stored transcripts would be unmasked during discrete periods of development. This appears to be true, at least for the majority of transcripts. During the division phase of the gametophyte, transcripts encoding products involved in various aspects of mitosis and the cell cycle are enriched. Similarly, during spermatid differentiation, transcripts encoding products involved in cilogenesis and cytoskeletal organization are enriched. During the transition between the division and differentiation phase, transcripts encoding products responsible for protein turnover were enriched and

may serve to 'wipe the slate clean' in order for the developing spermatids to begin differentiating. Our analysis reveals that post-transcriptional splicing may be a prominent mechanism controlling rates of development, as transcripts encoding products involved in splicing at highly overrepresented from 1-5 h (Figure II-9).

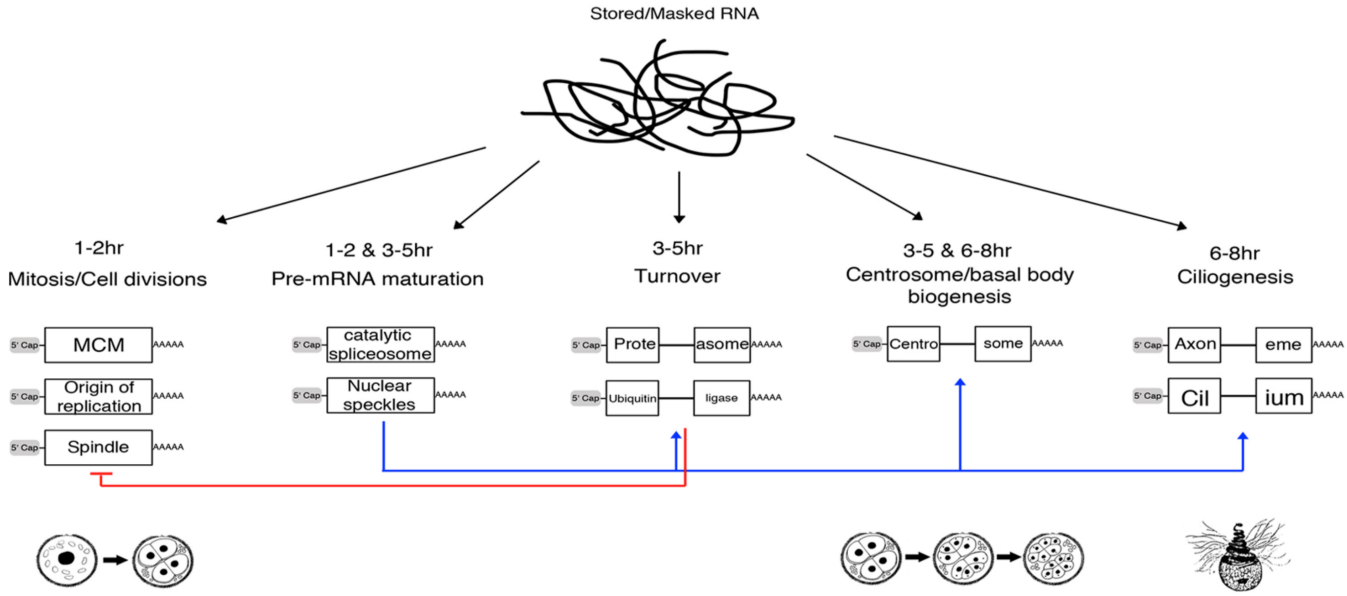


Figure II-9. *Working model*. The developing male gametophyte of *M. vestita* stores translationally-masked transcripts that are unmasked and utilized at precise times during spermatogenesis. Transcripts are unmasked in a temporally specific fashion, such that they are detectable during their period(s) of function. During divisions, transcripts encoding products involved in mitosis are abundant. Later as spermatogenous cells complete divisions and begin their differentiation phase, transcripts encoding products involved in protein turnover are unmasked. During the differentiation phase, transcripts encoding products involved in cytoskeletal organization and ciliogenesis are unmasked and mediate the differentiation of spermatozoids. Early on, there is an enrichment of transcripts encoding splicing factors. This coupled with the identification of a large subset of intron retaining transcripts, indicates that transcripts utilized later in development may require additional maturation. T.C.C. Boothby, unpublished. Lower images of microspores and spermatozoid adapted from drawings by S.M. Wolniak; Sharp 1914.

Splicing as a post-transcription mechanism regulating spermatogenesis

The enrichment of RNA encoding splicing factors and nuclear speckle components as well as the identification of a large subset of IRTs suggests that post-

transcription splicing plays a major role in development of spermatozoids in *M. vestita*. Nuclear speckles are the known site of post-transcription splicing of retained introns and will be discussed in detail in Chapter IV. Intron retention is a widespread form of alternative splicing (AS) in all eukaryotic kingdoms, and in plants, AS is the predominant form (Campbell *et al*., 2006, McGuire *et al*., 2008). Post-transcriptional splicing of retained introns has been shown in several cases to play a role in development of metazoans and fungi. However, despite being the predominant form of AS, information about the functional role of intron retention is lacking in plants (Reich *et al*., 1992; Campbell *et al*., 2006; Ner-Gaon *et al*., 2004; Clark and Thanaraj, 2002; Kan *et al*., 2002; Iida *et al*., 2004; Wang and Brendel, 2006; Marquez *et al.,* 2012). In the next chapter, validations of results from our transcriptome analysis are presented. IRTs and fully spliced isoforms are confirmed by RT-PCR using isoform specific primers. IRT maturation and spermatogenesis are shown to be dependent on the maturation of splicesomes. Furthermore, IRT maturation is shown to precede translation and developmental function, suggesting that intron retention acts to block the utilization of a transcript and the splicing of introns removes this block. Intron retention and post-transcriptional splicing are mechanisms regulating the use of stored RNA in the developing male gametophyte of *M. vestita*.

Chapter III - Intron retention and post-transcriptional splicing as a temporal mediator of translation

Note: the following has been adapted from Boothby et al., Developmental Cell, 2013

Background

The preceding chapter provides details on RNA sequencing, and the assembly and analysis of transcriptomes from the male gametophyte of M. vestita. RNAseq was used to obtain transcript information for the entire span of spermatogenesis as well as for specific portions of the developmental process. Several findings from this analysis suggest that splicing plays a role in regulating the use of stored transcripts essential for male gamete formation in *M. vestita*. Firstly, transcripts encoding proteins involved in splicing are enriched, especially during the early and middle portions of spermatogenesis. Secondly, a large subset of transcripts present in our transcriptome consists of intron retaining transcripts (IRTs), which encode proteins essential to many later developmental processes. In this chapter, details of experiments addressing the requirement and function of intron retention and post-transcriptional splicing of retained introns during spermatogenesis are presented.

Introduction

Post-transcriptional regulation and intron retention

Alternative splicing (AS) is common in many eukaryotes (Campbell *et al*., 2006; Kim *et al*., 2007; Ner-Gaon *et al*., 2007; Syed *et al.,* 2012). Intron retention (IR) is an AS event where the splicing of an intron is skipped, resulting in an otherwise mature

transcript harboring an unprocessed sequence. While IR events have been shown to be more common in plants than in animals (45.1% and ~30-47.9% of all AS products in rice and *Arabidopsis* versus 2-10% in humans), questions about the widespread function of IR in plants persist (Reich *et al*., 1992; Campbell *et al*., 2006; Ner-Gaon *et al*., 2004; Clark and Thanaraj, 2002; Kan *et al*., 2002; Iida *et al*., 2004; Wang and Brendel, 2006; Marquez *et al.,* 2012). Most IR events are thought to involve poorly defined splice signals that contribute to suboptimal splicing efficiency (Hampson and Rottman, 1987; Dirksen *et al*., 1995; Romano *et al*., 2001; Sterner and Berget, 1993; Talerico and Berget, 1994; McCullough and Berget, 1997; Romfo *et al*., 2000; Sakabe and de Souza, 2007) and cis acting elements (for review: Wang and Burge, 2008). Short intron length has also been implicated in IR (Galante *et al*., 2004; Stamm *et al*., 2000; Sugnet *et al*., 2004; Zheng *et al*., 2005; Ohler *et al*., 2005; Sakabe and de Souza, 2007). IR can be triggered by external stimuli, it can be specific to developmental phases and tissue types, and it can display sexual dimorphism (Marrs and Walbot 1997; Winter *et al*., 1988; Gebauer *et al*., 1998; Averbeck *et al.,* 2005; Mansilla *et al*., 2005; Filichkin *et al*., 2010). Retained introns have been shown to affect the stability, function, localization, and translatability of the transcripts containing them (Altieri, 1994; Ebihara *et al*., 1996; Bor *et al*., 2006; Jaillon *et al*., 2008; Buckley *et al.,* 2011).

The splicing of introns from pre-mRNA is a likely mechanism that may regulate the timing for translation of stored pre-mRNA in the microspore. The mRNA encoding the *M. vestita* homolog of splicing factor PRP-19 is specifically localized to the cytoplasm of spermatogenous cells during later stages of divisions (Tsai *et al*., 2004), a localization that can be disrupted through RNAi mediated depletion of Mv-Mago (van

der Weele *et al*., 2007). Furthermore, RNAi depletion of PRP-19 results in the failure of spermatid differentiation (Tsai *et al*., 2004). Cell-specific localization of splicing factors and cell type specific anomalies resulting from their depletion hint that the splicing of stored transcripts could serve as a post-transcriptional mechanism controlling transcript utilization during differentiation of the microspore of *M. vestita*.

Here, we have tested whether the splicing of stored transcripts plays a regulatory role in controlling the timing of translation for specific mRNAs during rapid development of this male gametophyte. In the preceding chapter I have discussed the construction of *de novo* transcriptomes using RNAseq data generated with Poly(A)+ RNA isolated from gametophytes during specific time intervals in development. Additionally, I have detailed the *in silico* identification and analysis of intron retaining transcripts (IRTs) and their fully spliced isoforms.

In this chapter, data are presented confirming the presence and splicing of these IRTs during development. RT-PCR is used to confirm that IRTs temporally precede their fully spliced isoforms during development. We used the spliceosome inhibitor Spliceostatin A (Kaida *et al*., 2007), and the transcriptional inhibitor α-amanitin (Hart and Wolniak, 1999; Klink and Wolniak, 2001) to confirm that the maturation of IRTs to mRNA is spliceosome-dependent but independent of transcription. RT-PCR was used to assess the timing of splicing for a subset of IRTs; the analysis reveals that introns are removed from these RNAs at distinct times during development. RNAi was used to deplete cells of specific IRTs and fluorescence microscopy used to assess the timing of resulting developmental perturbations, which in all cases examined (17/17), occurred after the predicted time that splicing would have occurred. We cross referenced our RT-

PCR and RNAi time course data with previously obtained temporal protein abundance data for the gametophyte (Klink, 2001; Klink and Wolniak, 2003; Deeb *et al*., 2010) and we find in all (4/4) cases that spliced isoforms precede or are contemporary with translation of a specific transcript, whereas translation precedes or is contemporary with RNAi-induced perturbations. In all (13/13) instances examined where protein data were not available, we found that the experimentally determined time of splicing preceded or was contemporary with RNAi-induced perturbations. Based on our data, we hypothesize that a subset of transcripts stored in the spore contains at least 1 unspliced intron, which is removed later in development. Retention of an intron results in, or contributes to, translational inhibition of the IRT and removal of the intron releases this block and thereby allows translation to proceed.

Results

Intron Retaining Transcripts are Stored and Later Spliced in the microspore of M. vestita

To confirm the existence of IRTs as well as fully spliced isoforms, isoform specific primers were constructed for RT-PCR. Isoform specific primers offer several advantages over primers designed to amplify multiple splice variants. Firstly, they allow for non-competitive detection of rare isoforms whose presence might be obscured by more abundant isoforms because of the exponential nature of PCR. Secondly, by not amplifying multiple splice isoforms in the same reaction, isoform specific primers allow for PCR products free from stem-loop artifacts that can arise from the annealing of

intron-containing and complimentary intron-free strands. For all such primer pairs, IRT specific primers were designed to hybridize within the intron of an IRT, while fully spliced isoforms had 1st strand primers designed to hybridize to the exon-exon junction formed by the removal of the retained intron. In concert with strict PCR parameters, this strategy allowed us to amplify specific isoforms of IRT pairs. RNA was isolated from microspores after different intervals of development, and RT-PCR was then conducted (Figure III-1). Isoform specific primers made to recognize IRT (2178A) and fully spliced (2178B) variants of 2178 (NUDCD2) show a temporal shift in abundance, where 2178A was present from the onset of development, and 2178B became detectable after 4 hours of development (Figure III-1). Standardized gel intensities were graphed for both the IRT and spliced isoform of MvU2178 as well as non-IRT Mv-Cen1 (loading control), whose transcript has previously been shown to be present from the onset of development (Hart and Wolniak, 1999) (Figure III-1). Quantification demonstrates that IRT 2178A RNA is abundant early and declines later in development. The inverse appears to be true for fully spliced 2178B RNA, which in contrast to 2178A, increases over development (Figure III-1). This trend can be visualized by tracking the ratio of 2178B:2178A RT-PCR gel intensities, which show a rapid increase at 4 h and nearly double between 4 and 8 h (Figure III-1).

Since the IRT isoform 2178A was still detectable after the appearance of the fully spliced isoform, we wanted to confirm that the appearance of 2178B was dependent on the splicing of 2178A by treating microspores with the splicing inhibitor Spliceostatin A (SSA) (Kaida *et al*., 2007). RNA extracted from SSA treated microspores was used for RT-PCR using isoform specific primers for 2178A and B (Figure III-1). SSA-mediated

inhibition of splicing resulted in the persistence of the 2178A isoform and there was no increase in the abundance of 2178B over 8 hours.

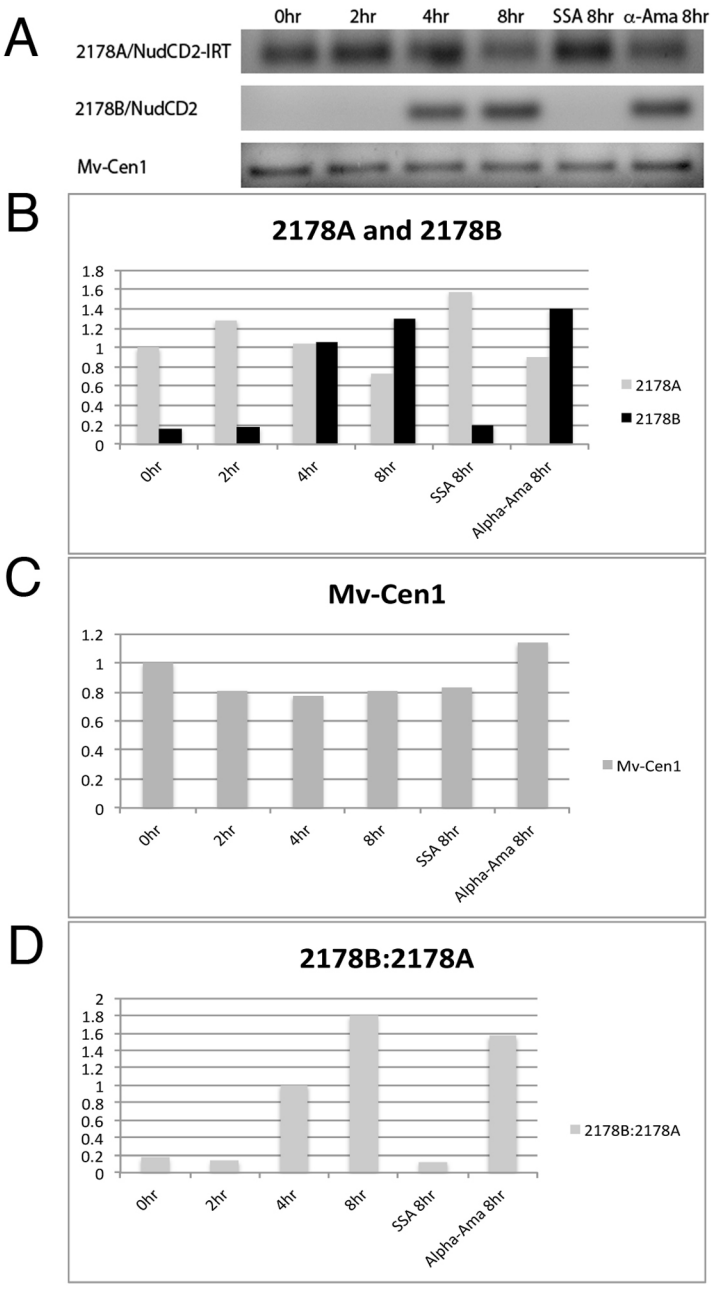


Figure III-1. *Splicing of IRTs is spliceosome dependent and transcription independent event.* (**A**), RNA isolated from microspores grown for the indicated times was isolated, normalized, and used with isoform specific primers to amplify IRT (2178A) and fully spliced (2178B) species. RNA from microspores treated with the splicing inhibitor SSA or transcriptional inhibitor α-amanitin was isolated after 8 hours of development and used with isoform specific primers to amplify IRT (2178A) and fully spliced (2178B) species. Mv-Cen1 was used as a control for normalization. (**B**), Standardized gel band intensities were obtained for both 2178A and B isoforms and relative intensities graphed for comparison (2178A 0 h was used as the standard). (**C**), Standardized gel bands intensities for Mv-Cen1 (Mv-Cen1 0 h was used as a standard). (**D**), Ratio of 2178B and 2178A standardized gel intensities. Boothby *et al*., 2013.

Microspores treated with SSA and fixed after 8 hours of development were viewed to determine the timing and extent of their developmental arrest (Figure III-2). SSA treatment appears to arrest cells during or just before the final cell division, corresponding to approximately 4.5-5 hours of gametophyte development as evidenced by the number and size of cells, lack of cellular elongation, and lack of nuclear elongation and coiling evident in treated cells as compared to control cells fixed after 8 hours (Figure III-2).

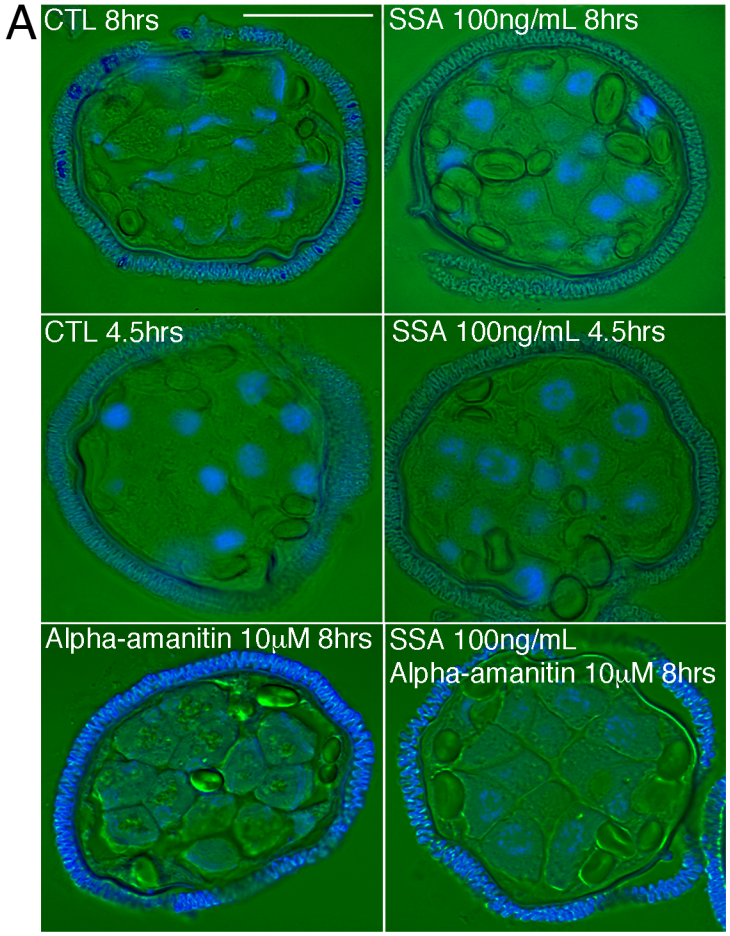


Figure III-2. *Spliceostatin A perturbs development of the microspore of* <u>*M. vestita*</u>*.* (**A**), control microspores were hydrated and allowed to develop for 8 or 4.5 hours. Microspores were treated with SSA, α-amanitin, or both at the time of their hydration and grown for 8 or 4.5 hours as indicated. Chromatin was stained with DAPI (blue) to aid in assessment of nuclear morphology and developmental progression (Bar = 25μm). Boothby *et al.*, 2013.

Additionally, we examined the effect of inhibiting splicing on several developmental markers whose patterns of distribution have been well characterized (Figure III-3). The addition of SSA inhibited the dissociation of the blepharoplast and the alignment of basal bodies along the nucleus (Figure III-3), a reduction and mislocalization of α-tubulin (Figure III-3), as well as a reduction and mislocalization of spermidine (SPD), which is normally abundant in the spermatids at 8 hours (Figure III-3).

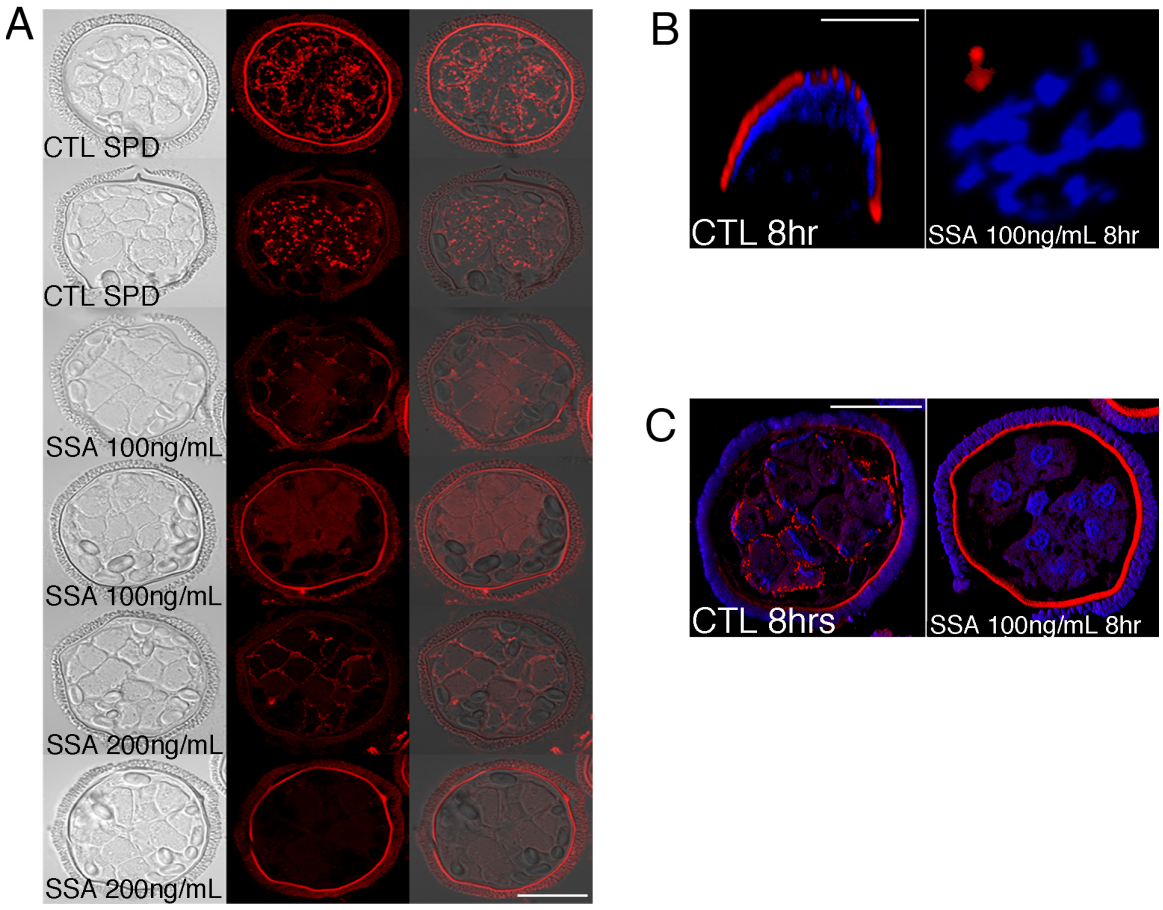


Figure III-3. *Inhibition of splicing disrupts the abundance and distribution of several developmental markers.* (**A**), Confocal imaging of subcellular distribution of the polyamine spermidine (SPD, red) in 8 h control and disruption in SSA treated microspores (Bar = 25μm). (**B**), Confocal projections showing nuclear morphology (DAPI, blue) and Centrin protein (red) distribution in 8 h control and disruption in SSA treated microspores (Bar = 2.5μm). (**C**), Confocal projections showing subcellular distribution of α-tubulin protein (red) in 8 h control and disruption in SSA treated microspores. Nuclei were stained with DAPI (blue) (Bar = 25μm). Boothby *et al.*, 2013.

We were concerned that SSA treatment might simply slow rates of development, so that SSA treated cells grown for 8 hours would appear as if they had developed for only ~4.5 h. To assess this, microspores were treated with SSA at the time of hydration and grown for 4.5 h, along with control microspores (Figure III-2). Both 4.5 h SSA treated and control spores developed at the same rate for the first 4.5 h suggesting that SSA treatment does not cause development to proceed at a slower rate, but rather arrests it at a specific stage of development.

While the microspore of *M. vestita* has previously been reported to be a transcriptionally quiescent system (review: Wolniak *et al.,* 2011) with many RNA species being 'masked' at the onset of development (Deeb *et al*., 2010), we wanted to confirm that spliced isoforms arise from IRTs stored and present in the desiccated microspore. We treated microspores from the time of hydration with the transcriptional inhibitor α-amanitin. Consistent with previous results (Klink and Wolniak, 2001), the development of microspores treated with α-amanitin proceeded normally (Figure III-2). RNA was isolated from α-amanitin treated microspores and 2178A- and 2178B-specific primers were used to perform RT-PCR (Figure III-1). RT-PCR results were similar to control reactions, showing that neither of these RNAs is produced by new transcription in the gametophyte. RNAseq reads from our 6-8 h time range as well as RNAseq reads from 8 h microspores treated with transcriptional inhibitors α-amanitin or actinomycin D were mapped to a subset of IRT sequences (Figure III-4). While the number of mapped reads fluctuates between all three samples to some degree, there is not an overall negative trend with regard to IRT abundance after treatment with transcriptional inhibitors. Instead,

there is a slightly positive trend with mapped reading being most abundant in one of the inhibitor samples for 10/17 IRTs (Figure III-4).

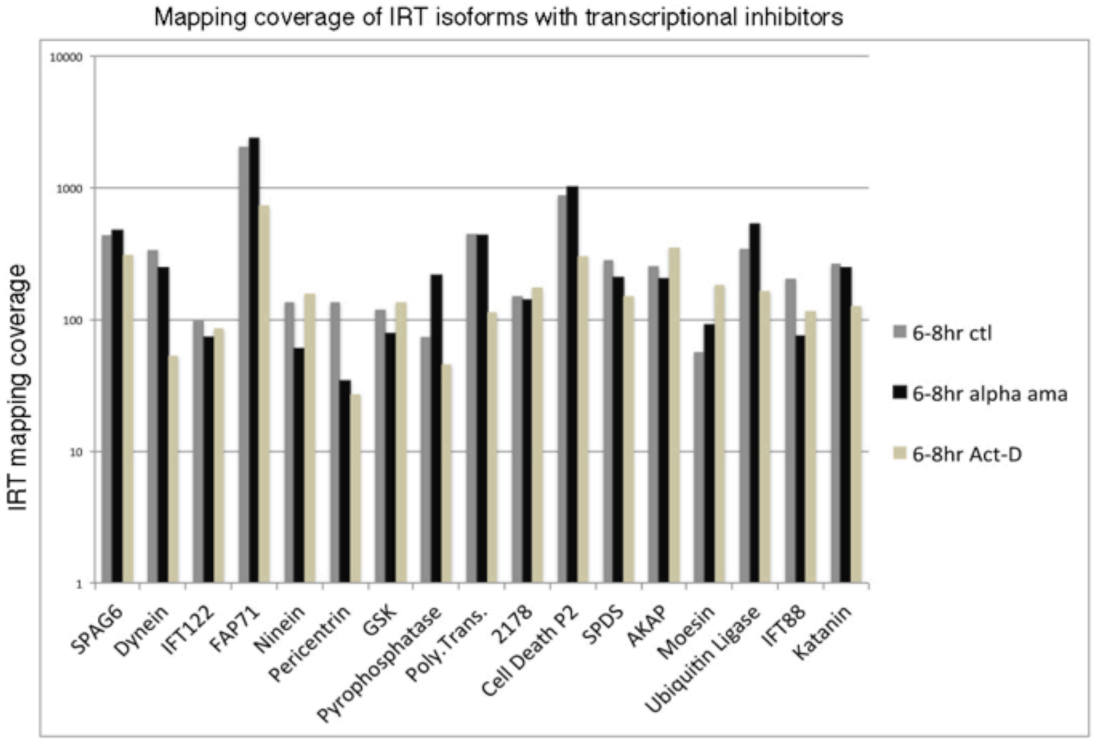


Figure III-4. *Mapping coverage of IRT isoforms with and without transcriptional inhibitors*. Microspores were treated with the indicated inhibitors and developed for 8 h. RNA from these microspores was isolated and used for RNAseq. RNAseq reads from these as well as control (6-8 h) samples were used to estimate FPKM values for representative IRT isoforms. Boothby *et al.*, 2013.

Together, these results indicate that IRTs are stored in the desiccated microspore and present from the onset of gametophyte development. These IRTs are precursors that are spliced to generate fully mature transcripts during gametophyte development. Both IRT maturation as well as a transition from the division to differentiation phase during spermatogenesis appear to depend on spliceosomal maturation, but do not require *de novo* transcription.

Temporal Variance in the Detection of Fully Spliced Transcripts

As a means to understand how IR and splicing could regulate the temporal utilization of transcripts during spermiogenesis, we selected a subset of IRTs and examined the timing of their splicing. As noted above, primers were designed to amplify specific spliced isoforms using RNA extracted from different times during development of the gametophyte. RNA concentrations were quantified spectrophotometrically and normalized to a standard concentration prior to use in RT-PCR.

Our pre-mRNA subset consisted of 17 different IRTs and 1 fully spliced control (Mv-Cen1) transcript. For our IRT subset, we selected several transcripts known or predicted to be involved in centrosome formation or ciliogenesis (SPAG6, Ninein, IFT88, Dynein Heavy Chain 1b, IFT122, Katanin, FAP71, and Pericentrin) in addition to transcripts we believed might also function at later stages of development.

Mv-Cen1 served as a baseline control, since it is fully spliced and its mRNA has previously been shown to be present at consistent levels throughout development of the microspore (Hart and Wolniak, 1999, Figure III-1 & III-5). In contrast, RT-PCR results for our IRT subset revealed variance in the timing of the appearance of corresponding spliced isoforms (Figure III-5). Interestingly, 6 transcripts (SPDS, E3-Ligase, Katanin, AKAP2, GSK3, and Moesin) exhibited spliced isoforms at 1 hour of development. While the spermidine synthase (SPDS) spliced isoform was detected throughout development, its abundance increased dramatically after 5 hours, a time that coincides with the appearance of this transcript in spermatogenous cells (Deeb *et al*., 2010). For fully

spliced transcripts present at 1 hour of development, isoform specific RT-PCR was performed on 0-hour RNA to see if these isoforms exist in the desiccated spore or whether they are spliced from IRT precursors (Figure III-5).

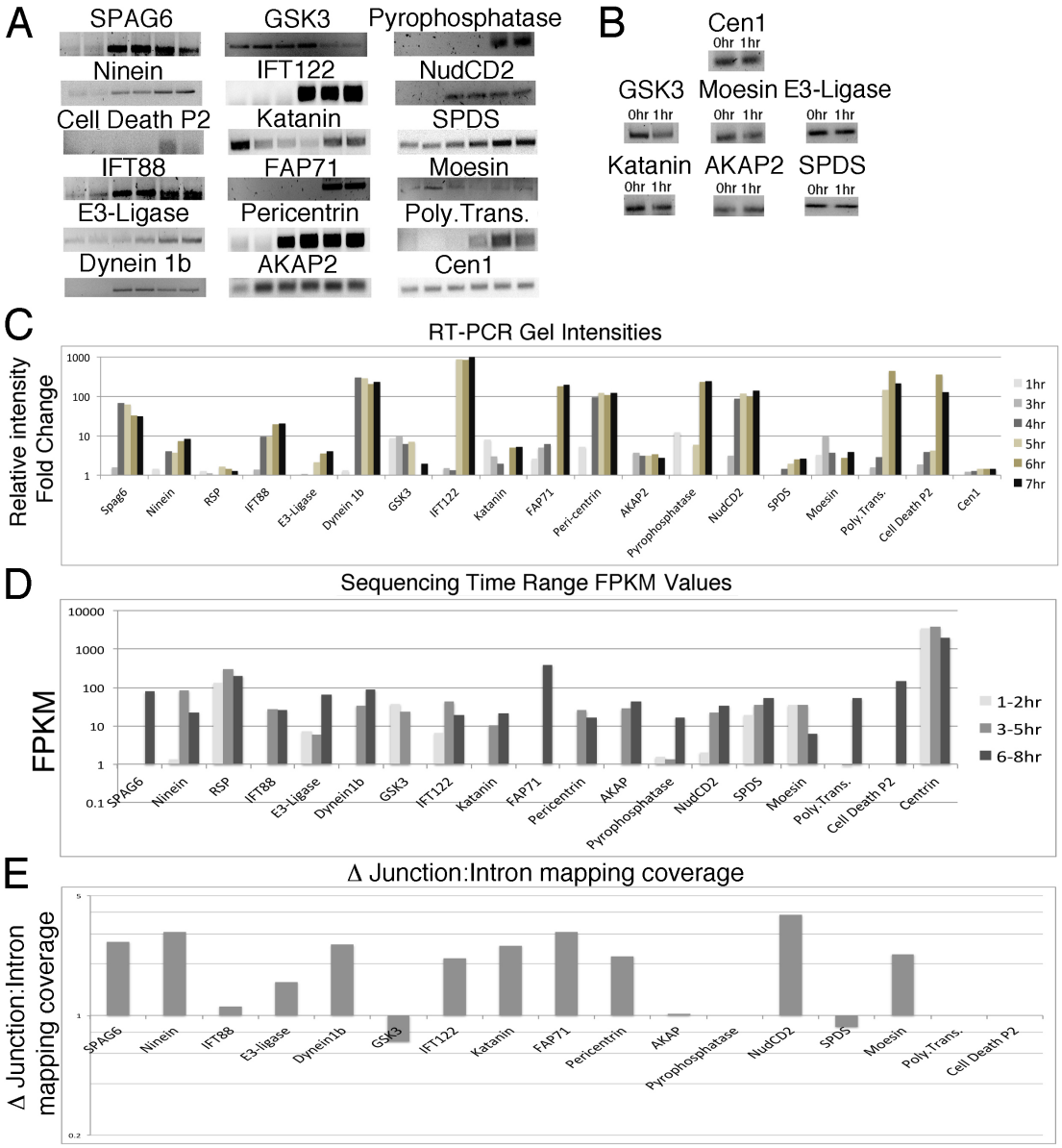


Figure III-5. *Temporal Variance in Intron Removal*. (**A**), RNA isolated from microspores grown for 1, 3, 4, 5, 6, and 7 hours was normalized and used with isoform specific primers for fully spliced transcripts for RT-PCR. (**B**), For IRTs with high abundances at 1 h, the presence of these isoforms was assayed for in the dry spore (0 h) using isoform specific primers. Cen1 serves as a none-IRT control. (**C**), Temporal RT-product relative fold change of gel intensities for IRT subset. (**D**), Fragments per kilobase per million fragments mapped (FPKM) values from IRT isoforms for each time range sequenced. (**E**), Change in the ratio of fully spliced to IRT isoforms. Boothby *et al*., 2013.

Fragments mapped per kilobases of exon per million reads (FPKM) values were graphed for each fully spliced isoform (Figure III-5). While our FPKM values time intervals (1-2 h, 3-5 h, and 6-8 h) do not have the same temporal resolution as our RT-PCR data (0 h, 1 h, 3 h, 4 h, 5 h, 6 h, 7 h), both data sets appear to conform well, adding further evidence that the appearance of fully spliced isoforms of different transcripts are temporally distinct events. If IRTs are serving as precursors to fully spliced isoforms, we expect to see the ratio of fully spliced to IRT isoforms increase during development. To see if this is true, we measured the number of RNAseq reads corresponding to splice junctions and introns and calculated the change in this ratio over time (Figure III-5). In the majority of cases (12/17) the ratio of fully spliced isoform to IRT increased over development (Figure III-5). For one transcript (Polyamine Transporter: Poly.Trans.) there was no change in the ratio of fully spliced to IRT transcript during development. In two cases (Pyrophosphatase and Programed Cell Death P2), RNAseq reads corresponding to introns and splice junctions were unmasked only in the final time interval, making it impossible to evaluate changes in the ratio of fully-spliced to IRT transcripts with the time-range resolutions currently available to us. In two cases (SPDS and GSK3), the ratio of fully spliced to IRT transcript decreased. Interestingly, our RT-PCR results for these two transcripts indicate that their fully spliced isoforms are present from the onset of development (Figure III-5). These observations lead us to question if these fully mature isoforms of these transcripts are spliced from IRTs after rehydration. To test this, microspores were treated with SSA and allowed to develop for 1 or 4 hours before RNA was isolated from each sample (Figure III-6). In both cases, RNA extracted at 1 hour of development did not indicate any effect on the presence of spliced isoform because of

splicing inhibition; however, at 4 hours, fully spliced isoforms were not detectable (Figure III-6). This indicates that some IRTs may be spliced in waves or that IRTs might be spliced to replenish supplies of fully spliced isoforms as they are utilized during development. Another alternative is that there is cell type specific splicing, which occurs after spermatogenous cells are formed. Collectively, these RT-PCR and *in silco* results indicate that different fully spliced isoforms arise at different times during development.

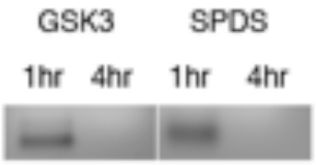


Figure III-6. *Inhibition of splicing does not affect the abundance of early fully spliced isoforms but causes their disappearance later*. Microspores were treated with 100ng/mL SSA at the time of hydration and allowed to develop for 1 or 4 hours. After the appropriate time, RNA was isolated from microspores and concentrations normalized. Isoform specific primers for GSK3 and SPDS were used to amplify fully spliced isoforms of these transcripts. Boothby *et al*., 2013.

RNAi Depletion of Intron Retaining Transcripts does not Perturb Development Until After the Time of Splicing

To determine whether IRTs play a functional role in development, or merely exist as storage precursors of their fully spliced isoforms, we performed dsRNA-mediated RNAi on our subset of IRTs. We previously found that the delivery of dsRNA into the microspore of *M. vestita* for RNAi depletion of stored transcripts is an effective method for studying whether and when a transcript plays an essential role in development (Klink and Wolniak, 2001). For these experiments, we transcribed and constructed dsRNA molecules specific to the introns present in our subset of IRTs. These dsRNA constructs were introduced to microspores at the time of hydration. Microspores were grown for 8

hours, fixed, embedded in plastic, and sectioned. Sectioned material was stained with DAPI and examined using incident light fluorescence and phase contrast microscopy (Figure III-7).

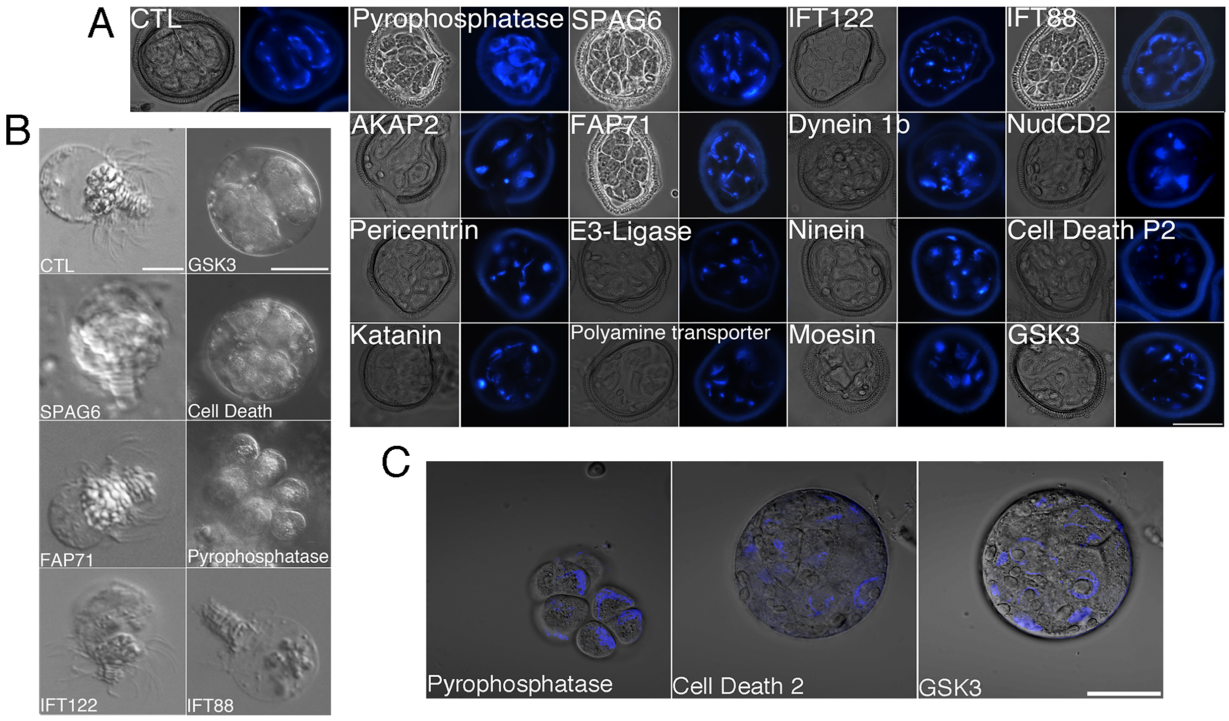


Figure III-7. *Perturbations by RNAi depletion of IRTs Manifest after the Predicted Time of Splicing.* (**A**), Microspores were treated with 200µg/mL of dsRNA targeting specific IRTs for RNAi. Treated and control microspores were grown for 8 hours, fixed, embedded, and sectioned. Sectioned material was stained with the nucleic acid stain DAPI (blue) and imaged using fluorescence and phase contrast microscopy. (**B**), For those IRTs whose depletion via RNAi was not seen to perturb maturation before 8 hours of development RNAi treatments were conducted and microspores grown until spermatozoid emergence (~10.5 hours). Imaging of spermatozoids or microspores was conducted to assess the effect of RNAi on mature spermatozoid (Right Bar (for microspores) = 25µm, Left bar (for spermatozoids) = 5µm). (**C**), Fluorescence confocal microscopy on microspores treated with dsRNA and grown for 16 hours, fixed, and stained with DAPI (DAPI) and overlaid on DIC images (Bar = 25µm). Boothby *et al*., 2013.

Several of our initial RNAi treatments failed to elicit developmental perturbations within the first 8 hours. For these transcripts, we extended RNAi treatments, allowing

microspores to progress through the time of normal spermatozoid release. In control samples, spermatozoids emerge after 10.5-11 h of development at 20^{o}C (Figure III-7). Several RNAi treatments allowed for the release of spermatozoids displaying various developmental abnormalities, while other RNAi treatments arrested development prior to spermatozoid release (Figure III-7). These latter RNAi treatments were allowed to develop further, up to 16 hours, to account for any potential delay in development caused by RNAi treatment. In all (3/3) cases, spermatozoids still failed to emerge (Figure III-7).

In this study, each dsRNA construct produced a different phenocopy, indicating a specific effect rather then general toxicity caused by adding excessive amounts of dsRNA to the spores. In each case examined (17/17), we found that the RNAi-induced perturbation of development was manifested after the proposed time of splicing for the transcript that had been silenced. This result, combined with the observation that fully spliced isoforms arise in a splicing dependent fashion (Figure III-1) leads us to believe that IRTs do not directly function in development, but rather, that they serve as precursors of fully spliced transcripts, whose maturation and subsequent translation is necessary for the proper completion of spermatogenesis.

Correlation between splicing, developmental arrest, and translation

The temporal changes in abundance of several proteins encoded by IRTs have previously been reported (Klink, 2001; Klink and Wolniak, 2003; Deeb *et al*., 2010). We cross-referenced our RT and RNAi time course data with previous protein abundance data (Klink, 2001; Klink and Wolniak, 2003; Deeb *et al*., 2010). The results are shown in

Figure III-8. In all cases (4/4) with available protein data, splicing of transcripts preceded or was contemporary with protein production. Furthermore, in all case examined (17/17), RNAi mediated arrest of development was preceded by or was concurrent with both splicing and protein production of IRT-containing pre-mRNAs (Figure III-8).

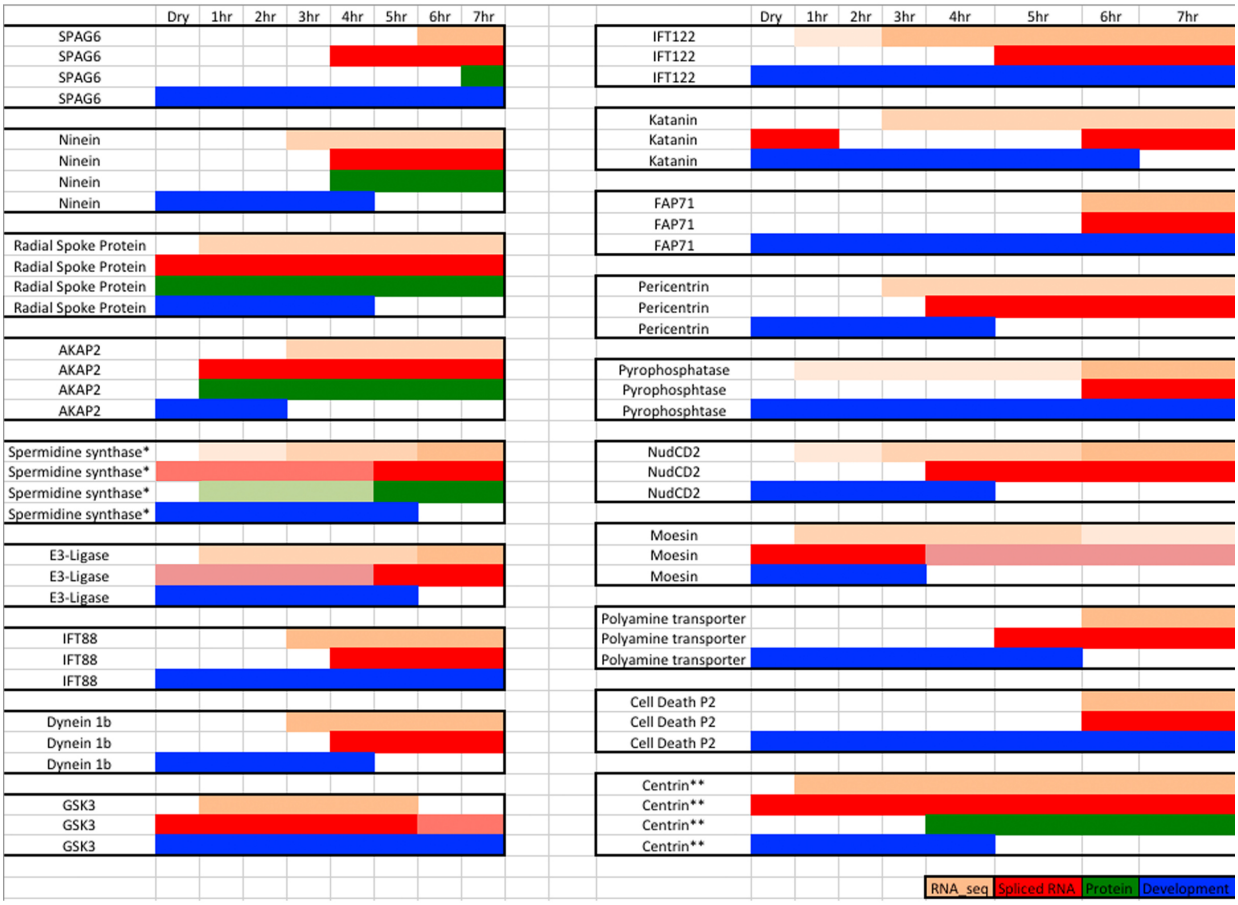


Figure III-8. *Splicing of IRTs precedes functionality and translation.* Temporal plot showing the times at which fully spliced isoforms (RNA_seq: tan, RT-PCR: red) and proteins (green) are present. Also included is the duration of developmental viability after RNAi depletion of IRTs (blue). Shading for RNAseq (tan) and RT-PCR (red) indicates relative abundance. * Temporal protein data are for SPD, the product of SPDS; RNAi development arrest time point previously reported (Deeb *et al*., 2010). ** Non-IRT control. Boothby *et al*., 2013.

Cross-referencing of our RT-PCR, *in silico* analysis, protein abundance, and RNAi data, leads us to suspect that retained introns present in IRTs prevent their precocious translation, and that the gametophyte utilizes a mechanism to control the timing of splicing of these introns. Once the retained introns are removed, the fully spliced transcript isoforms are competent for translation. In this way, the rapidly developing gametophyte is able to regulate the translation of a subset of stored transcripts through the removal of retained introns.

Conclusions/Discussion

The use of stored RNA is important in many rapidly developing systems. Here, we investigated the mechanistic role of intron retention in regulating the translation of stored RNA in the microspore of *M. vestita*. These data indicate that the microspore of *M. vestita* regulates the retention of specific introns during its desiccation, a time when the spores undergo expansive transcription of stored pre-mRNAs (Boothby and Wolniak, 2011). After spore rehydration, the stored IRTs are translationally inhibited, and the gametophyte regulates the removal of retained introns at specific times during development as a means of promoting translation.

While many RNAs in the microspore of *M. vestita* are fully mature, a subset of transcripts retains an intron. Examples of inefficient splicing are known and often attributable to imperfect splice signals (Hampson and Rottman, 1987; Dirksen *et al*., 1995; Romano *et al*., 2001; Sterner and Berget, 1993; Talerico and Berget, 1994; McCullough and Berget, 1997; Romfo *et al*., 2000; Sakabe and de Souza, 2007). The fact that many of the retained introns present in the microspore of *M. vestita* contain 5' and 3'

U2 splice signals (Figure II-7) that conform well to accepted consensus sequences from other plants and animals leads us to suspect that these retention events do not arise because of inefficient splicing. Instead, the retention of the intron appears to be precise and regulated. Lending further support to this idea is the fact that analysis of genomic sequences revealed that in all (4/4) cases examined, splicing of other non-retained introns had previously occurred in IRTs. A variety of other organisms and cell types display intron retention events, and though the role of retained introns in the regulation of developmental progression has not been widely reported, a precedent exists for such observations (Averbeck *et al.*, 2005). In instances where this process has been described, regulated retention of introns appears to be specific to a developmental stage (Averbeck *et al.,* 2005) and to correlate with a decrease in protein product while not compromising the stability of the IRT (Mansilla *et al*., 2005). Post-transcriptional splicing of retained introns has also been observed (Denis *et al*., 2005), suggesting that stable IRTs could undergo splicing long after their transcription. In the case of the dry microspore of *Marsilea*, this interval could extend for many years.

Our data indicate that IRTs are spliced at different times. We believe that this reveals an additional level of regulation, one that specifies the removal of specific retained introns at specific times. Two scenarios are plausible to explain the control over timing of intron removal. In the first, IRTs contain sequences or sequence structures that interact with some temporal regulator of splicing (review: Wang and Burge, 2008). That temporal regulator is translated, activated, or uninhibited at a specific time and is then able to interact with and promote the splicing of IRTs containing the correct sequence/structure. By having multiple sequences and multiple temporal regulators, the

cell could effectively splice different subsets of IRTs at different times. In the second scenario, the cell is fully capable of splicing all IRTs; however, the splicing substrates (the IRTs) are masked (Deeb *et al*., 2010; Boothby and Wolniak, 2011) or otherwise blocked from being spliced. The unmasking of a subset of IRTs would make them available to the splicing machinery of the cell. Temporally regulating this unmasking for subsets of IRTs would lead to the splicing of different subsets of IRTs at different times of development.

It appears that for several IRT pairs in our subset, a fully spliced isoform is present in the dry spore. In these cases, IRTs may serve as backup pools for these transcripts, or as is likely the case for SPDS, spliced isoforms may arise at different times in different cell types at successive stages of gametophyte development (Deeb *et al*., 2010). This theory is supported by the observation that spliceosomal inhibition does not influence the abundance of these products at 1 hour of development but leads to their disappearance at 4 hours of development (Figure III-6).

Some IRTs also have fully spliced isoforms present early in development, and the depletion of these IRTs via RNAi results in arrested development prior to the completion of the division phase (*e.g.,* Moesin, which arrests development around 3 hours), while drug induced inhibition of splicing via SSA results in arrested development at the end of the division cycles, approximately 4.5-5 hours after development is initiated. SSA affects the maturation of the spliceosome by interaction with the spliceosomal component SF3b and by blocking the complex A to complex B transition (Kaida *et al*., 2007). A possible explanation of these seemingly disparate observations is that IRTs whose splicing is essential for early development may already have mature complex B spliceosomes

associated with them, so that the addition of SSA does not affect their splicing. Another possibility is that spliceosomes associated with these transcripts undergo the A to B complex transition very rapidly after spore hydration, and as a result, SSA introduced at the time of rehydration lacks sufficient time to block the maturation of these spliceosomes. Irrespective of the mechanism of regulation of intron removal, it is apparent that there are levels of control that dictate the timing of splicing of several subsets of IRTs.

RNAi knockdown of IRTs does not affect development until after the time point at which the transcripts are normally spliced and previous western blot, immunofluorescence analysis and RT-PCR time course data reveal that intron removal precedes translation in the microspores. The persistence of introns in a transcript has been seen to affect rates of translation negatively in a variety of other organisms (Braddock *et al*., 1994; Gebauer *et al*., 1998) and the post-transcriptional splicing of retained introns has been demonstrated to make these transcripts translationally viable (Denis *et al*., 2005). Collectively, these results indicate that the retention of introns arrests translational events necessary for spermatogenesis in *M. vestita* to proceed normally.

While it appears likely that IR blocks translation of IRTs in the gametophyte, the mechanism by which this occurs remains elusive. Furthermore, intron retention does not appear to trigger nonsense-mediated decay (NMD) or otherwise negatively affect the stability of these transcripts. It has been shown that in plants the majority of alternatively spliced isoforms are not subjected to NMD (Kalyna *et al*., 2012). Since targeting of transcripts for NMD occurs after the pioneering round of translation, it is likely that stable IRTs are not translated until after they are spliced (review: Chang *et al*., 2007). In

the microspore of *M. vestita* IRTs might not interact with translational machinery for a variety of reasons. Nuclear retention of IRTs could occur or alternatively sequestration in ribonucleoprotein particles could block IRT association with translational machinery.

With post-transcriptional splicing appearing to play a role in the temporal regulation of stored transcripts, the question remains as to how the use of stored transcripts is controlled spatially. As detailed above and in the preceding chapter, a large subset of stored transcripts in the microspore of *M. vestita* consists of IRTs that must undergo post-transcriptional splicing to become functional mRNAs. Logic dictates that these IRTs and spliceosomes must at some point be in the same place at the same time for splicing to occur, and this colocalization could control spatial utilization of post-transcriptionally spliced transcripts. In the next chapter, results from experiments tracking splicing factors during development are detailed. During desiccation, splicing factors are found to coalesce into a large subnuclear particle that contains markers unique to nuclear speckles (now the known site of post-transcriptional splicing, Boothby and Wolniak, 2011; Girard *et al.,* 2012). Aggregated speckles are shown to contain a subset of masked RNA, and intron specific probes show retained transcripts to have a similar pattern of distribution (Boothby and Wolniak, 2011). During development, speckles and their associated RNA transit from the nucleus to the cytoplasm of the antheridial mother cell and in the following cellular divisions are asymmetrically distributed to spermatogenous but not sterile cells. Inhibition of splicing results in the stabilization of IRTs within the nucleus of spermatogenous cells in a transcriptionally independent fashion. Thus, it appears that a likely mode of spatial regulation for stored IRTs is their association with

spliceosomes in nuclear speckles and the asymmetric distribution of these splicing factors during development.

Chapter IV - Spatial regulation of stored RNA through association of nuclear speckles

Note: the following has been adapted from Boothby and Wolniak, BMC Cell Biology, 2011 and Boothby et al., Developmental Cell, 2013

Background

The microspores of *M. vestita* are transcriptionally silent and rely on the use of stored transcripts to mediate spermatogenesis (for review: Wolniak *et al*., 2011). Not surprisingly, the translation of specific stored transcripts is under tight temporal and spatial control (Klink and Wolniak, 2001; Klink and Wolniak, 2003; Tsai and Wolniak, 2001; Tsai *et al*., 2004). One example of this spatial and temporal regulation of stored transcripts is centrin mRNA. Centrin is a calcium-binding phosphoprotein that has been shown to be essential in motile apparatus formation in the microspore of *M. vestita* (Klink and Wolniak, 2001). Centrin mRNA is uniformly distributed throughout the cytoplasm of the microspore from the onset of gametophyte development, but centrin protein levels are barely detectable during of the first four hours after the spores are hydrated. Beyond that time point, Centrin protein levels increase dramatically, but only in the spermatogenous cells, where they remain elevated through the completion of gamete formation (Hart and Wolniak, 1999; Klink and Wolniak, 2001; Klink and Wolniak 2003). Thus, the translational capacity for Centrin protein synthesis is asymmetric, because centrin mRNA is present in the cytoplasm of both sterile and spermatogenous cells in the gametophyte, but Centrin is translated only in spermatogenous cells (Tsai *et al*., 2004). Centrin RNA was examined in this study because of the extensive amount of preexisting

knowledge regarding its spatial and temporal dynamics during microspore development (review Wolniak *et al*., 2011). Similarly, temporal and spatial control over translation has been observed for a number of other transcripts (Tsai *et al*., 2004) and proteins (Klink and Wolniak; 2003) in these gametophytes.

An important mechanism regulating gametophyte development is the unmasking of stored transcripts for translation (review Sommerville and Ladomery, 1996). Within this context we define "masked RNA" as mRNA whose translational state is initially inhibited, but later is "unmasked" to become translationally competent. This pool of masked mRNA is stored in the nucleus of the desiccated spore (Deeb *et al*., 2010). We refer to mRNA that is uniformly distributed in the cytoplasm of all cell types in the gametophyte, but does not appear to be translated at any time during development as quiescent cytoplasmic mRNA (qc-mRNA).

Recently, we found that the polyamine, spermidine (SPD), acts as a temporal regulator for releasing the masked, stored transcripts in the gametophyte (Wolniak *et al*., 2011, Deeb *et al*., 2010). Exogenous additions of SPD and other polyamines at the time of spore hydration cause the precocious unmasking of spermidine synthase (SPDS) mRNA in addition to other masked transcripts including centrin, PRP-19, and gamma-tubulin (Deeb *et al*., 2010). High concentrations of SPD also arrest division cycles, presumably because of premature transcript unmasking. Precociously unmasked transcripts display an intriguing pattern of distribution, and appear as distinct particles in the nucleus (Deeb *et al.,* 2010). These findings led us to hypothesize that a subset of masked transcripts is stored within the nucleus of the microspore and that the temporal regulation of these transcripts is dependent on unmasking as a prerequisite for translation

essential to the proper completion of spermatogenesis. Since masked transcripts appear to be stored within the nucleoplasm of the microspore as it undergoes desiccation, we were also interested if these masked transcripts are associated with known nuclear bodies.

The preceding chapters (II and III) describe the identification of intron retaining transcripts (IRTs) and demonstrate that intron retention and post-transcriptional splicing of retained introns are mechanisms controlling the temporal utilization of these RNAs. Stored transcripts utilized in the developing male gametophyte of *Marsilea* are also subject to spatial regulation (Tsai *et al*., 2004). To understand the spatial regulation of stored transcripts and to identify the localization of IRTs that undergo post-transcriptional splicing, we examined nuclear speckle dynamics, as these structures are the known site of post-transcriptional splicing (Girard *et al.*, 2012).

Introduction

In most eukaryotic cells, the majority of introns are removed co-transcriptionally, while downstream (3') portions of the nascent transcript are still being transcribed (for review: Neugebauer, 2002). Like other subnuclear bodies that are observed to be enriched with splicing factors (*e.g.,* Cajal Bodies), the identification of spliceosomal components in nuclear speckles initially led to the hypothesis that they might be the site of splicing (review: Lamond and Spector, 2003). However, the finding that the majority of nascent transcripts localize to perichromatin fibrils (Monneron and Bernhard, 1969; Fakan and Bernhard, 1971; Fakan and Nobis, 1978; Cmarko *et al*., 1999) rather than nuclear speckles severally derailed research in this direction, despite logical arguments in favor of speckle mediated splicing (review: Hall *et al*., 2006). Recent investigations have

revealed that a subset of splicing takes place within nuclear speckles, but that this splicing is post-transcriptional, as opposed to the more common co-transcriptional form (Girard *et al*., 2012).

Nuclear speckles are small (~1 μm) aggregations of 20-25 nm granules that occupy the interchromatin space of many eukaryotic nuclei (Thiry, 1995). Several types of pre-mRNA processing proteins are constituents of nuclear speckles (Fu, 1995), and speckles also contain a subset of polyA)+ RNA (Huang *et al.,* 1994; Carter *et al*., 1991; Visa *et al*., 1993). Within the interchromatin space of the nucleus, speckles are often localized adjacent to genes with high transcriptional activity (Huang and Spector, 1991; Johnson *et al*., 2000; Moen *et al*., 2004; Smith *et al*., 1999, Xing and Lawrence, 1993; Xing *et al*, 1995).

While nuclear speckles are defined as intranuclear domains, many speckle constituents actually cycle between the nucleus and cytoplasm where they also form defined aggregates (Verheijen *et al*., 1986; Leser *et al*., 1989; Ferreira *et al*., 1994). This shuttling occurs in an orderly and predictable cycle, known as the speckle cycle. The cycle starts with a breakdown of the nuclear envelope as mitosis begins, at which time many proteins formerly associated with nuclear speckles become homogenously distributed within the cytosol (Reuter *et al.,* 1985; Spector and Smith, 1986). While a portion of these proteins remain diffuse within the cytosol, a subset aggregates during metaphase into mitotic interchromatin granules (MIGs), which by all analysis to date, appear to be analogous to interchromatin granules (ICGs), differing only in their cellular localization (cytoplasmic for MIGs versus nuclear for ICGs) (Ferreira *et al*, 1994; Leser *et al.,* 1989; Prasanth *et al*., 2003; Thiry, 1995). As the cell cycle progresses through

anaphase and into telophase, the size and number of MIGs continually increases. MIG associated splicing factors are recycled back into the nucleus but not until the nuclear envelope is reestablished (Prasanth *et al*, 2003). Initially, it was thought that intact MIGs might be transported into the nucleus (Leser *et al*., 1989; Thiry, 1995). However, more detailed investigations have uncovered that intact MIGs are not transported from the cytosol into the nucleus, but rather there are at least two waves of MIG component import (Ferreira *et al*., 1994; Prasanth *et al*., 2003). Initial clues came from MIGs labeled with antibodies against various splicing factors after some snRNPs had already been imported into the nucleus (Ferreira *et al*., 1994). Later, it was found that the first wave of MIG import comes in late telophase when SF2/ASF, SC35, U2B", and other fully functional splicing factors are displaced from MIGs and enter the nucleus forming nuclear speckles. The second wave occurs in G1 when residual SC35 and RNAPII enter the nucleus (Prasanth *et al*., 2003). Intriguingly, though protein constituents of nuclear speckles and MIGs appear to be the same, a cellular function for MIGs has not been established.

Another interesting aspect of speckle dynamics is that upon inhibition of transcription, speckles enlarge and assume a rounded morphology, which has been suggested to result from the storage of pre-mRNA splicing factors (Melcak *et al*., 2000; Spector *et al.,* 1983; Spector *et al*., 1991). In addition to pre-mRNA splicing factors, a subset of poly(A)+ RNA remains associated with enlarged nuclear speckles within the nucleus under conditions where transcription has been inhibited (Huang *et al*., 1994). The purpose of this nuclear retention of RNA is unknown, but it seems clear that while associated with these aggregated speckles, RNAs are not translated.

Stress conditions such as hypoxia, inhibition of respiration, transcription, phosphorylation and ethanol treatment have been shown to cause the sequestration of the Exon Junction Complex (EJC) core component eIF4A-III to several subnuclear structures including nuclear speckles (Koroleva *et al*., 2009a; Koroleva *et al*., 2009b). It is likely that mRNA associated with these sequestered components will be retained within the nucleus and not translated. The association of core components of the EJC with subnuclear RNA and/or nuclear speckles could play a vital role in regulating a subset of cellular processes. Previous experiments have demonstrated the ubiquitous expression of EJC core component Mago-nashi in other plant systems and that the loss of this expression has extensive effects on development (Park *et al*., 2009).

Since the microspore of *M. vestita* is a transcriptionally silent system that utilizes post-transcriptional splicing as a mechanism mediating temporal control over a subset of stored transcripts and since a subset of masked RNAs exist within the nucleus of the quiescent spore (Deeb *et al.,* 2010), we suspected that examining nuclear speckle dynamics might lead to insights into developmental control in the maturing gametophyte.

In this study, we have examined nuclear speckle dynamics during microspore entry into dormancy and transcriptional quiescence. We show that in addition to cytoplasmic stores, aggregated nuclear speckles serve as sites of poly(A)+ RNA storage. We developed a novel variation on fluorescence *in situ* hybridization (FISH), in an assay designed to distinguish between masked and unmasked (qc-mRNA) populations of the same transcripts in fixed cells. We demonstrate the utility of this assay by tracking the movements of specific transcripts initially stored in association with nuclear speckles into the cytoplasm of the antheridial mother cell. We show subsequent movements of masked

transcripts into spermatogenous cells, but not jacket cells of the developing gametophyte, and that this asymmetry may be regulated via the EJC component Mv-Mago. Furthermore, *in situ* labeling of IRTs (with probes targeting their introns) reveals that IRT dynamics are identical to those of speckle associated RNA. IRTs are seen to be localized within the nucleoplasm during prior to the first division during spermatogenesis. As nuclear speckles exit the nucleus, IRTs appear to be displaced as well and enter the cytoplasm. Asymmetric localization of speckles and IRTs occurs, resulting in spermatogenous, but not sterile cells inheriting pre-mRNA isoforms. After the time when specific IRT species are supposed to be spliced, we fail to detect their introns via *in situ* hybridization. The detection of IRTs can be stabilized by inhibiting splicing, which causes the sequestration of RNA species within the nuclei of spermatogenous cells. Thus, it appears that IRTs are associated with nuclear speckles and that this is their subcellular site of splicing. Asymmetric speckle distributions mediate the spatial utilization of associated IRTs and present a mechanism not only for the spatially precise translation of a subset of stored RNA but also for the regulation of cell fate determination.

Results

Nuclear speckles coalesce during desiccation and transcriptional silencing to form a single nuclear speckle aggregate

In other systems, when transcription is inhibited nuclear speckles are seen to form enlarged aggregates containing a subset of poly(A)+ RNA (review Lamond and Spector, 2003). We examined nuclear speckle dynamics during microspore desiccation, since this

is the developmental period when the gametophyte becomes transcriptionally silent. We harvested microspores from green, mostly-submerged sporophytes prior to the beginning of dehydration and subsequently every 2 weeks thereafter, as the ponds containing the sporophytes were allowed to dry out.

For all experiments using sectioned material presented here, thousands of microspores were grown, fixed, and embedded. Each experiment was conducted independently a minimum of 2 times. Between 30-40 sections were used per slide so that many hundreds of sectioned spores were observed for each experimental trial, of which a no less then 20 representative photographs were taken to ensure accurate analysis. The number of independent trails per experiment, and total number of photographs taken for each set of experiments are listed in Additional file 1, Table S1.

SC35 is a non-snRNP splicing factor that is often used as a specific marker of nuclear speckles (for review Lamond and Spector, 2003). We used the distribution of SC35 protein to assay for the presence of nuclear speckles and to assess putative speckle dynamics in drying spores (Figure IV-1). Because there is a high level of autofluorescence in drying and desiccated microspores, making immunofluorescence difficult, we instead used alkaline phosphatase conjugated secondary antibodies and the nitro-blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate detection method to label cells during this stage of development. Prior to desiccation, SC35 exhibited a typical speckled pattern throughout the nucleus (Figure IV-1). After 2 weeks of sporophyte drying, these speckles had increased in labeling intensity (Figure IV-1). By 4 weeks of drying, the speckled appearance of SC35 labeling appeared to have been replaced by larger aggregates of the protein within the nucleus (Figure IV-1). With total

desiccation of the spore, SC35 was detectable as a single, large subnuclear aggregation (Figure IV-1).

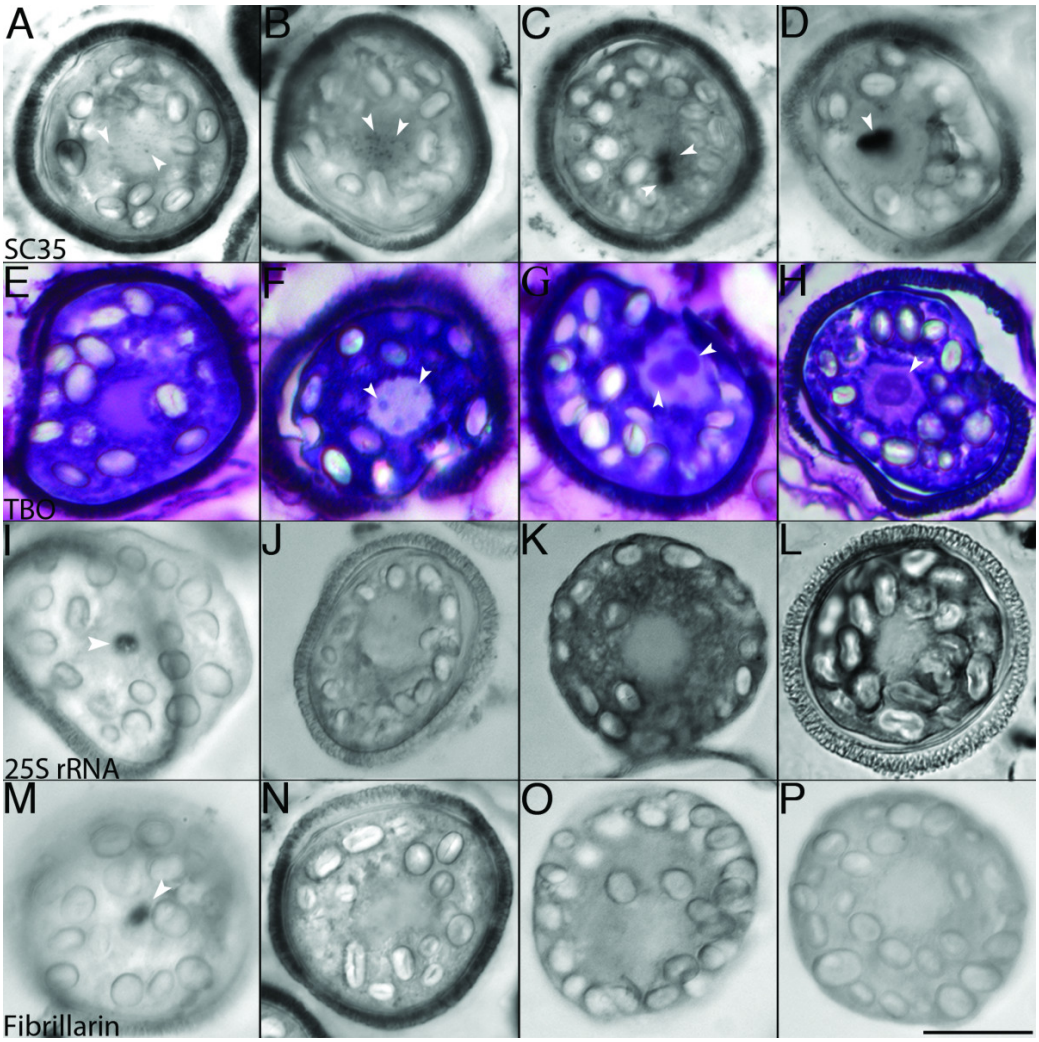


Figure IV-1. *Nuclear speckles aggregate during desiccation of the microspore of M. vestita.* Bright field microscopy: **A-D**, Labeling of microspores with SC-35 mAb against nuclear speckle marker SC35. **E-H**, TBO staining. **I-L**, ISH with probe for 25S rRNA. **M-P**, Labeling of microspores with mAb 72B9 against nucleolar marker Fibrillarin. Microspores were harvested from plants that were watered (**A**, **E**, **I**, & **M**), had not been watered for 2 weeks (**B**, **F**, **J**, & **N)**, had not been watered for 4 weeks (**C**, **G**, **K**, & **O)** and had not been watered for 6 weeks (**D**, **H**, **L**, & **P)**. Arrows denote subnuclear aggregations. Bar = 25 mm. Boothby and Wolniak, 2010.

We used the metachromatic dye Toluidine Blue O (TBO; Figure IV-1) to confirm our SC35 nuclear speckle labeling. In spores harvested prior to desiccation, TBO staining did not reveal the presence of any accumulated subnuclear material in the microspores Figure IV-1). In samples taken from ponds that had not received water for 2 weeks, TBO staining revealed small, multiple subnuclear aggregations in microspore nuclei (Figure IV-1). As the desiccation process continued, these small aggregations coalesced into larger particles (Figure IV-1), until a single large aggregation occupied most of the volume of the nucleoplasm.

Since the single aggregate present in dry spores superficially resembles a nucleolus, we used markers known to associate with nucleoli to confirm we were looking at aggregated speckles. To distinguish between speckles and nucleoli, we analyzed 25S rRNA distributions by *in situ* hybridization (ISH) in the desiccating microspore, which we expected to label large nucleoli. This rRNA was present as a conspicuous nuclear particle (Figure IV-1), and exhibited some cytoplasmic staining in the microspores prior to the onset of desiccation. Nuclear 25S rRNA persisted during the first 2 weeks of desiccation (Figure IV-1), but became undetectable thereafter and for the remainder of the dehydration process (Figure IV-1). Concurrent with the loss of nuclear 25S rRNA, levels of 25S rRNA increased in the cytoplasm as dehydration progressed (Figure IV-1), consistent with the export of newly-assembled ribosomes to the cytoplasm. Similarly, Fibrillarin, an rRNA processing factor and nucleolar marker, was apparent in a conspicuous nuclear inclusion (Figure IV-1) identical in morphology to 25S rRNA ISH patterns obtained at this time point compare with (Figure IV-1). The abundance of

Fibrillarin within desiccating microspores declined over time, and was weakly detectable up to but not after 2 weeks of drying (Figure IV-1). The disappearance of the nucleolar markers from the nuclei of desiccating microspores shows that ribosomal synthesis reaches completion before the spores are totally dry. As the nucleoli become less prominent, the nuclear speckles become more conspicuous, thereby revealing a developmental shift from rRNA synthesis to pre-mRNA synthesis. Moreover, continued desiccation and a transition to transcriptional quiescence triggers the aggregation of nuclear speckles in the microspore of *M. vestita*.

Nuclear speckles remain aggregated in the newly hydrated microspore

Rehydration of the microspore triggers the commencement of spermatogenesis, but does not disrupt transcriptional silencing. We were interested in seeing if speckles remained aggregated after microspore rehydration. We examined the distribution of nuclear speckle markers, SC35 and U2B", a snRNP splicing factor known to be present in nuclear speckles. Since a subset of PolyA)+ RNA is retained in speckles during transcriptional inhibition (Huang *et al*., 1994), we used a biotinylated poly(T) probe for Poly(A)+ RNA as an additional marker for aggregated speckles. SC35 protein and poly(A)+ RNA localized as an aggregate within the nucleus of newly hydrated microspores (Figure IV-2). U2B" protein also overlapped with polyA)+ RNA within the nucleus (Figure IV-2).

Similar to our observations of later stages of spore desiccation, nucleolar markers did not reveal the presence of a nucleolus in the rehydrated microspores. As expected, Fibrillarin was barely detectable within the nucleus of newly hydrated microspores and appeared diffuse throughout the nucleoplasm (Figure IV-2), whereas anti-Fibrillarin

antibody labeled the nucleoli of sporophytic cells (Figure IV-2) and microspores prior to desiccation (Figure IV-1) robustly label nucleoli, demonstrating the efficacy of this antibody in *M. vestita*. In ISH assays, 5S and 25S rRNA were not detected within the nuclei but were abundant throughout the cytoplasm of newly hydrated microspores (Figure IV-2). These data confirm that nuclear speckles remain aggregated within the nucleus of newly hydrated microspores of *M. vestita*, and, as seen in mammalian systems (Huang *et al*., 1994), these enlarged speckles contain a subset of poly(A)+ RNA.

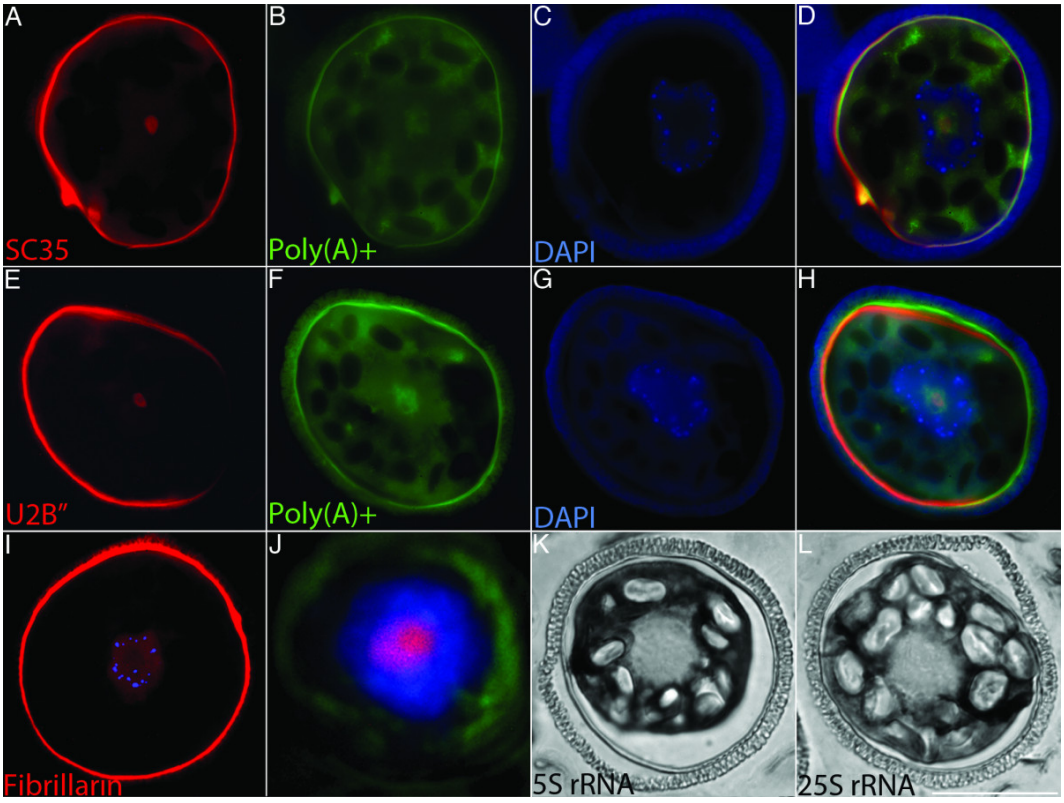


Figure IV-2. *Nuclear speckles remain aggregated within the nuclei of newly hydrated microspores*. **A-I**, **K**, and **L**, microspores were fixed and sectioned after 30 minutes of development. **A-D**, dual detection of (**A**) SC35 (red) and (**B**) Poly(A)+ RNA (green). **C**, DAPI (blue). **D**, Merge of **A-C**. **E-H**, Dual detection of (**E**) U2B" (red) and (**F**) Poly(A)+ RNA (green). **G**, DAPI (blue). **H**, Merge of **E-G**. **I**, Microspore labeled with monoclonal antibody 72B9 against Fibrillarin (red) and counter stained with DAPI (blue). **J**, Monoclonal antibody 72B9 labeling (red) of the nucleolus in a sporophytic cell, counter stained with DAPI (blue) and overlain on the phase contrast image (green) of the gametophyte. **K**, Representative 5S rRNA ISH labeling of the cytoplasm of a microspore (dark staining). **L**, Representative 25S rRNA ISH labeling of the cytoplasm of a microspore (dark staining). Bar = 25 µm. Boothby and Wolniak, 2010.

Nuclear poly(A)+ RNA leaves the nucleus during the first division

Earlier work showed that spermatogenesis in *M. vestita* relies on the translation of stored mRNA (Hart and Wolniak, 1998; Klink and Wolniak, 2001; Klink and Wolniak; 2003; Tsai *et al*., 2004; Hart and Wolniak, 1999), but this translation cannot take place within the first 30 minutes after spore hydration (Hart and Wolniak; 1998). We were interested in whether poly(A)+ RNA present in the nuclear speckle aggregate could be utilized by the microspore after this initial period of quiescence in early gametophyte development. A prerequisite for the translation of speckle associated RNA would be its movement into the cytosol. We studied the localization and storage of masked mRNA in microspores after developing a fluorescent DNA/RNA differential detection assay based on the methyl green/pyronin Y histochemical staining technique. Histochemical dyes such as Pyronin Y lack exclusive specificity for RNA, so off-target binding to other nucleic acids can cause erroneous results (Toba *et al*., 1995). One solution to this problem is to use a DNA-binding dye, such as methyl green, as a competitor dye in conjunction with an RNA-binding dye, such as Pyronin Y, with the aim of eliminating or decreasing off target binding (Kurnick, 1955). We found that the commonly used DNA preferential method employing methyl green/pyronin staining is ineffective in *M. vestita* because the chromatin becomes tightly condensed in the spermatids. We recently showed that DAPI labels chromatin in the developing spermatids of *M. vestita* (van der Weele *et al*., 2007), so we developed an assay similar to methyl green/pyronin Y staining by substituting DAPI for methyl green. This substitution greatly reduced the off-target labeling of both DNA and RNA (Figure IV-3) allowing us to examine RNA distributions

during microspore desiccation (Figure IV-4) and gametophyte development (Figure IV-5).

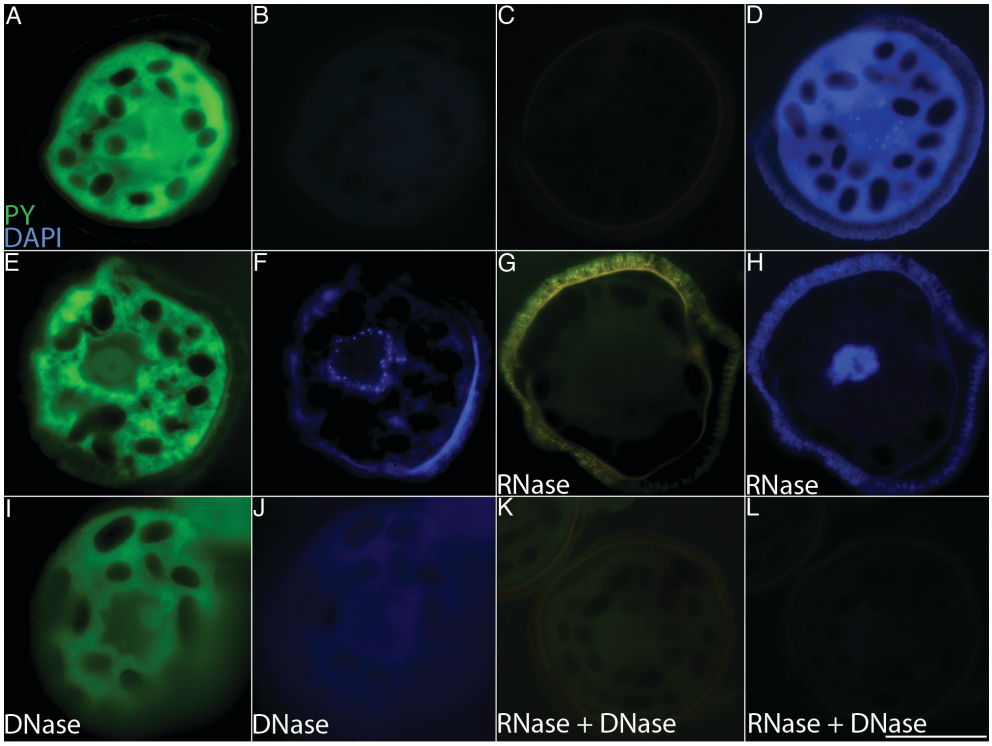


Figure IV-3. *Differential fluorescent labeling of DNA and RNA via dual DAPI/PY staining.* Microspores fixed at 30 minutes of development. PY (green) (**A**, **C**, **E**, **G**, **I**, **K**) and DAPI (blue) (**B**, **D**, **F**, **H**, **J**, **L**) detected via 488 nm and UV illumination respectively. **A-B**, Microspore stained with PY (green). **C-D**, Microspore stained with DAPI (blue). **E-F**, Microspore double-stained with DAPI (blue) and PY (green). **G-H**, Microspore pretreated with RNase and double-stained with both DAPI (blue) and PY (green). **i-j**, DNase pretreated samples were double-stained with DAPI (blue) and PY (green). **K-L**, DNase and RNase treated samples were double-stained with DAPI (blue) and PY (green). Bar = 25 μm. Boothby and Wolniak, 2010.

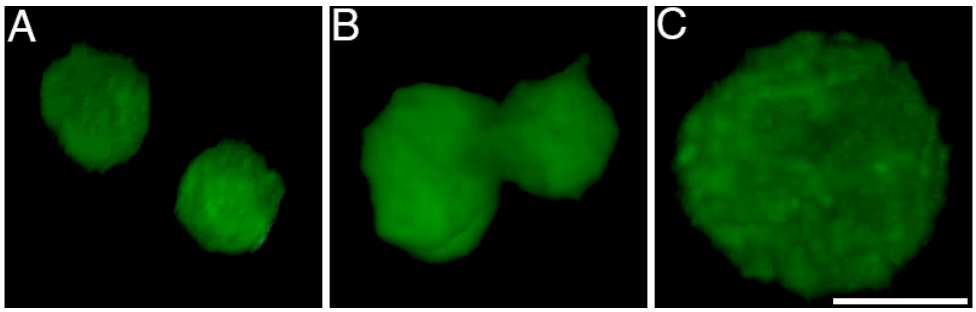


Figure IV-4. *PY labeling of subnuclear RNA in desiccating microspores.* (**A**), Microspores collected and fixed after 2 weeks without watering were sectioned and double stained with DAPI and PY. Subnuclear PY (green) signal was detected in successive confocal slices and these were rendered as a 3D model. (**B**), Microspores collected and fixed after 4 weeks without watering were sectioned and double stained with DAPI and PY. Subnuclear PY (green) signal was detected in successive confocal slices and these were rendered as a 3D model. (**C**), Microspores collected and fixed after 6 weeks without watering were sectioned and double stained with DAPI and PY. Subnuclear PY (green) signal was detected in successive confocal slices and these were rendered as a 3D model. Bar = 2.5 μm. Boothby and Wolniak, 2010.

We used this dual staining assay to assess the distribution of stored RNA during male gamete development (Figure IV-5). At all time points examined (30 minutes, 45 minutes, 1.5 hours, and 5 hours), high concentrations of RNA were seen within the cytoplasm of both spermatogenous and sterile cells. Prior to the first division, a large aggregation of RNA was observed within the nuclei of newly hydrated microspores (Figure IV-3, IV-4, IV-5). After the first division cycle had occurred (Figure IV-5), the large RNA aggregate was no longer detected within the nucleus, and PY staining was absent from the nucleoplasm (Figure IV-5, IV-6).

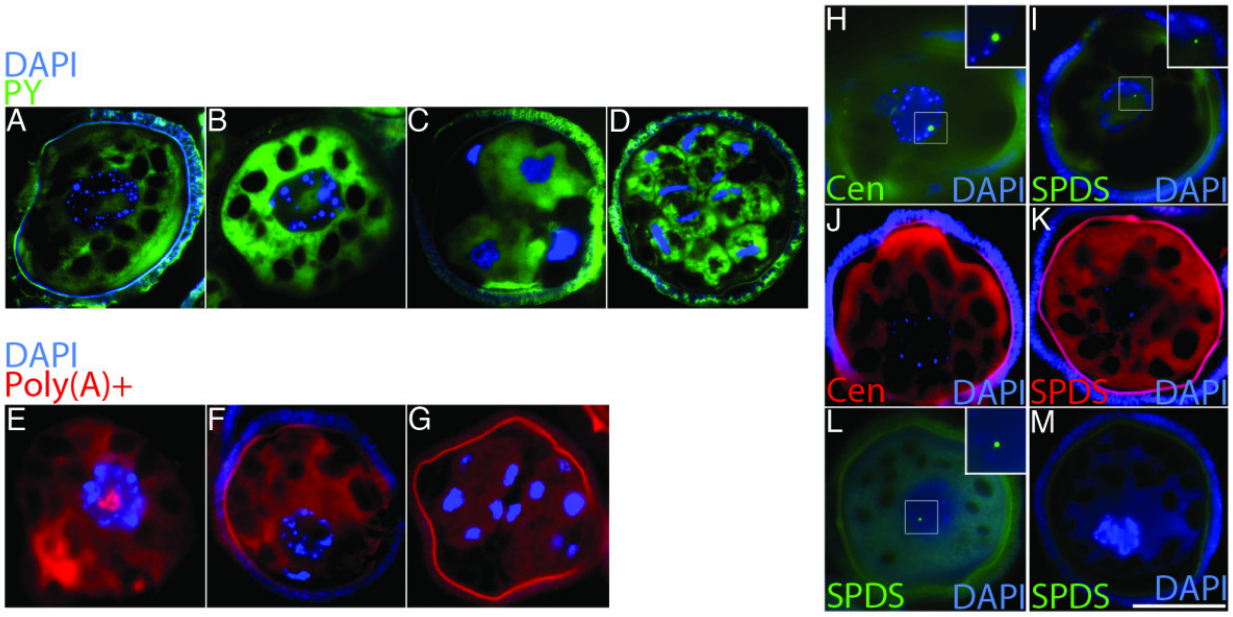


Figure IV-5. *Masked mRNA is stored within the microspore of M. vestita.* **A**, Microspore fixed after 30 minutes of development, double-stained with DAPI (blue) and PY (green). **B**, Microspore fixed after 45 minutes of development (just prior to the prothallial division) double-stained with DAPI (blue) PY (green). **C**, Microspore fixed after 1.5 hours of development double- stained with DAPI (blue) PY (green). **D**, Microspore fixed after 5 hours of development double-stained with DAPI (blue) PY (green). **E-G**, Distribution of Poly(A)+ RNA during spermatogenesis. A biotinylated poly(T) probe (red) was used for FISH on microspores fixed at (**E**) 30 minutes, (**F**) 1 hour, and (**G**) 4 hours of development. FISH probes were detected using avidin bound TexasRed. **H-M**, Microspores fixed after 30 minutes of development. **H**, FISH using 25mer biotinylated centrin specific probes, and detected using avidin bound Fluorescein (green). **I**, FISH using 25mer biotinylated SPDS specific probes, and detected using avidin bound Fluorescein (green). **J**, FISH using traditional biotinylated centrin specific probes, and detected using avidin bound TexasRed (red). **K**, FISH using traditional biotinylated SPDS specific probes, and detected using avidin bound TexasRed (red). **L-M**, Microspores pretreated with (**L**) DNase or (**M**) RNase and incubated with 25mer biotinylated SPDS probes, which were detected using avidin bound Fluorescein (green). Bar = 25 µm. Boothby and Wolniak, 2010.

To confirm these results, we used our biotinylated poly(T) probe to detect poly(A)+ RNA movements during development. At 30 minutes of development, a large aggregate of poly(A)+ RNA was seen within the nucleoplasm of undivided microspores (Figure IV-5). At the completion of the first (prothallial) division, poly(A)+ RNA was no longer detected within the nuclei of cells (Figure IV-5) and this lack of nucleoplasmic poly(A)+ RNA persisted through all the division cycles (Figure IV-5). Poly(A)+ RNA relocalized in an identical pattern as total RNA (Figure IV-5A-D compare with IV-5E-G); poly(A)+ RNA that had been stored in association with aggregated nuclear speckles exits the nucleus during the first division of the gametophyte . This is the earliest time point in development when proteins can be translated *in vitro* from gametophyte mRNA isolates (Hart and Wolniak; 1998).

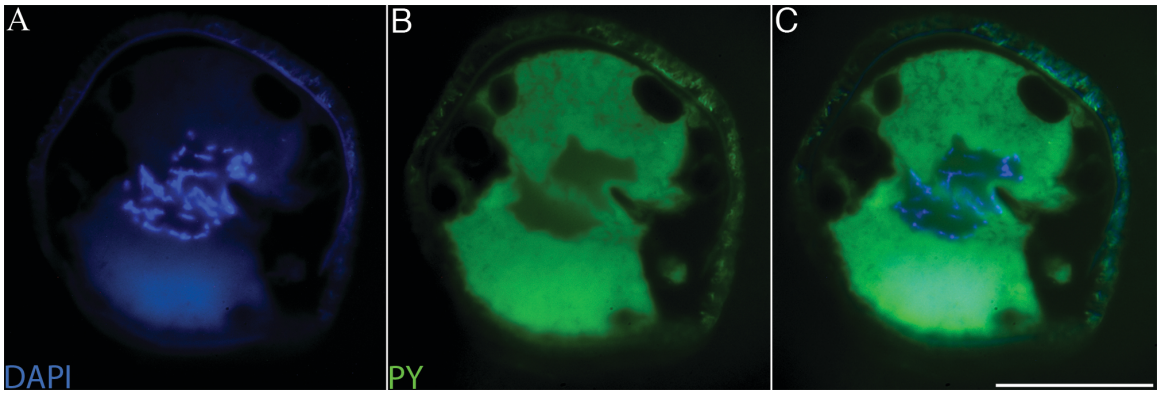


Figure IV-6. *RNA is not detectable within the nuclei of microspores after the first division.* **A-C**, Microspore fixed and sectioned after 1.5 hours of development. **A**, DAPI (blue). **B**, Pyronin Y (green). **C**, merge of **A** and **B**. Bar = 25 µm. Boothby and Wolniak, 2010.

Masked mRNA species localize to discrete foci within the nuclear speckle aggregate

Since precociously unmasked RNAs are detectable within the nucleus of the newly hydrated spore after 10 mM polyamine additions (Deeb *et al*., 2010), we reasoned that masked transcripts could constitute a subset of speckle associated poly(A)+ RNA. Since additions of SPD cause mitotic arrest and the precocious unmasking of transcripts (Deeb *et al*., 2010), it is reasonable to assume that if these transcripts are associated with nuclear speckles, then SPD should perturb speckle aggregation within the microspore nucleus. We tested this hypothesis by adding 10 mM SPD to microspores, which were then allowed to develop for 4 hours. The effect of SPD on speckle aggregation was examined during PY-DAPI labeling, and we found that SPD causes the partial or complete dissociation of speckles (Figure IV-7). In cases of partial dissociation (Figure IV-B) the subnuclear PY signal was observed as one or more amorphous masses within the nucleoplasm of the microspore. Total dissociation of the speckle aggregate was seen as the loss of subnuclear organization; the speckle aggregate no longer occupied a defined central portion of the nucleoplasm but rather, the PY signal was dispersed throughout the nucleus but apparently contained by the nuclear envelope (Figure IV-7A).

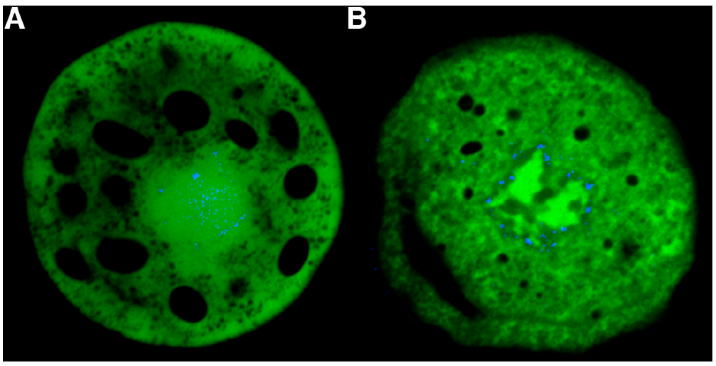


Figure IV-7. *Unmasking of stored subnuclear RNA causes the total or partial dissociation of nuclear speckles.* Microspores were incubated with 10 mM SPD for 4 hours, fixed, sectioned and double stained with DAPI (blue) and PY (green). Boothby and Wolniak, 2010.

To confirm that masked transcripts are associated with aggregated speckles, and then, to examine their role in spermatogenesis, it was essential to find a method for visualizing specific masked transcripts in unperturbed microspores. We found that traditional ISH and FISH methods work well for determining the localization of qc-mRNA transcripts, but these strategies fail to label masked transcripts (Tsai *et al*., 2004; Deeb *et al*., 2010) present in the speckles, presumably because masking agents obscure hybridization sites. We found that a 'short' 30 base poly(T) probe would label subnuclear RNA in the newly hydrated spores (Figure IV-2, IV-5), thereby suggesting that small FISH probes might reveal the localization patterns of both masked and qc-mRNA by hybridizing between masking agents.

We made short 25mer biotinylated DNA probes complementary to SPDS and centrin mRNAs, transcripts that become detectable by ISH within the nucleus after treatments of spores with 10 mM SPD (Deeb *et al*., 2010). The 25mer probes were added to sections of fixed gametophytes that had been developing for 30 minutes (Figure IV-5). These 25mer probes showed diffuse cytoplasmic labeling in addition to intense subnuclear labeling at conspicuous foci in newly hydrated microspores (Figure IV-5). Longer hybridization probes derived from SPDS and centrin were incubated with successive sections from the same specimen block, containing the same gametophytes (Figure IV-5). The longer probes hybridized with qc-mRNAs present throughout the cytoplasm, but these transcripts were undetectable in the nuclei of the cells. Samples treated with DNase and assayed with "short" probes retained their cytoplasmic and subnuclear hybridization patterns for the tested transcripts (Figure IV-5), while samples pretreated with RNase before short probe hybridization lacked any detectable RNA

labeling (Figure IV-5). The fluorescent foci detectable with short-probe hybridizations were usually centrally situated in the nucleus and did not overlap with any of the chromosomes (Figure IV-8). Thus, the short FISH probes are not hybridizing to chromosomes.

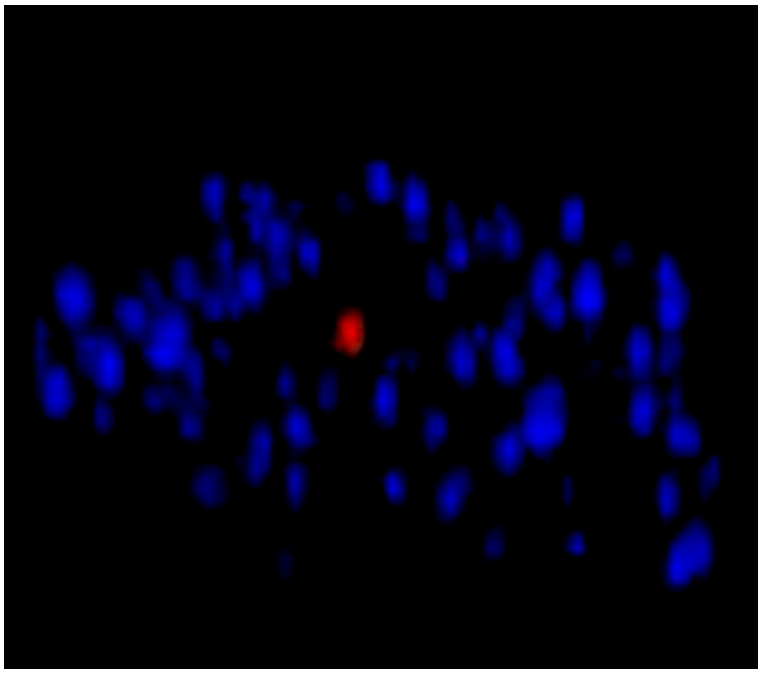


Figure IV-8. *'Short' FISH probes detect foci of subnuclear masked transcripts that are distinct from chromatin.* FISH against masked SPDS (red) was conducted on 20 μm sections taken from microspores fixed after 30 minutes after hydration. Probes were detected using avidin bound TexasRed. Sections were counterstained with DAPI (blue). Boothby and Wolniak, 2010.

Dual FISH labeling with a poly(T) probe and a 25mer SPDS probe showed that foci of specific masked transcripts localize with nuclear speckle aggregate-associated poly(A)+ RNA (Figure IV-9). Double-labeling of samples with short probes for SPDS and centrin showed distinct foci for each transcript within the nucleoplasm (Figure IV-9). These data taken with our previous finding that polyamine additions can cause the unmasking of subnuclear RNA (Deeb *et al*., 2010) demonstrate that masked mRNAs are

stored in discrete foci associated with aggregated nuclear speckles, and that the addition of SPD disrupts both the aggregation of nuclear speckles and the masking of these mRNAs.

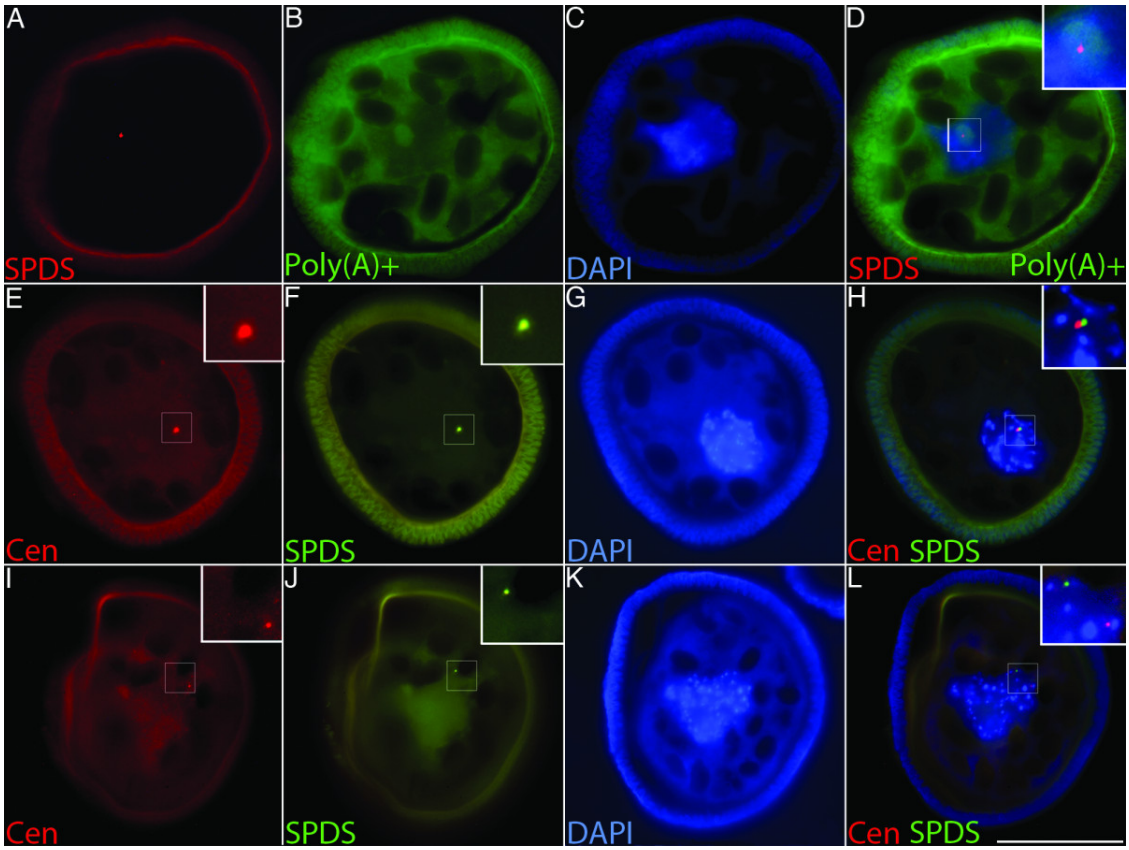


Figure IV-9. *Subnuclear poly(A)+ RNA localizes with discrete foci of masked mRNA in the microspore of M. vestita.* **A-L**, Developing microspores were fixed at 30 minutes of development. **A-D**, Dual FISH labeling of microspores with (**A**) 25mer biotinylated probe against SPDS (red) and (**B**) a poly(T) probe (green). SPDS probe was detected using avidin conjugated TexasRed (red) (**A**). An avidin/biotin blocking step was performed, followed by labeling and detection of the Poly(T) probe using an avidin conjugated Fluorescein (green) (**A**). **C**, DAPI (blue). **D**, Merge of **A-C**; inset shows enlarged view of subnuclear poly(A)+ RNA (green) and masked SPDS (red). 25mer biotinylated probes for centrin (red) (**E** & **I**) and SPDS (green) (**F** & **J**) were used to label sections sequentially. Centrin probe was detected using avidin conjugated TexasRed (red) (**E** & **I**). An avidin/biotin blocking step was performed, followed by labeling and detection of SPDS probe using an avidin conjugated Fluorescein (green) (**F** & **J**). Sections were counter stained with DAPI (blue) (**G** & **K**). Centrin (red), SPDS (green), and DAPI (blue) labeling were merged in **H** and **L**, insets show foci of short probe hybridization. Bar = 25 µm. Boothby and Wolniak, 2010.

RNA and protein components of the nuclear speckle aggregate are asymmetrically redistributed to the cytoplasm of spermatogenous cells

We were interested in what happens to nuclear speckle components as they exit the nucleus, so we tracked the abundance and distribution of specific transcripts at various time points during spermatogenesis, starting as the gametophytes progressed through their first division cycle (Figure IV-10). Using short FISH probes specific for masked SPDS mRNA, we found that prior to the first division, masked SPDS transcripts were centrally localized within the nucleus (Figure IV-10A). As the first (prothallial) division approached, the nucleus became repositioned near the periphery of the microspore with the chromosomes to be apportioned to the prothallial cell arranged in a spherical array (Figure IV-10B). At this stage, masked SPDS transcripts were still encircled by the chromosomes that would later be segregated to the antheridial mother cell. As the prothallial division proceeded, masked SPDS transcripts relocated to a site in the cytosol adjacent to the antheridial mother cell nucleus (Figure IV-10C). Sections from the same sample blocks were made and assayed by FISH for SPDS qc-mRNA transcripts by using longer probes (Figure IV-11). At all time points before, during, and after the prothallial division, cytoplasmic SPDS transcripts were abundant throughout the cytoplasm but not detectable in the nuclei (Figure IV-10, IV-11). Consistent with DAPI/PY staining and poly(A)+ FISH assays, it is evident that the subnuclear stores of masked mRNA exit the nucleus at the time of nuclear envelope breakdown during the first division, and are specifically passed on to the antheridial mother cell.

By the end of all nine mitotic division cycles in the gametophyte, small foci of both SPDS and centrin masked transcripts were apparent in the cytosol, adjacent to the

nucleus of each spermatid (Figure IV-10). These fluorescent particles were absent in the adjacent jacket cells of the gametophyte.

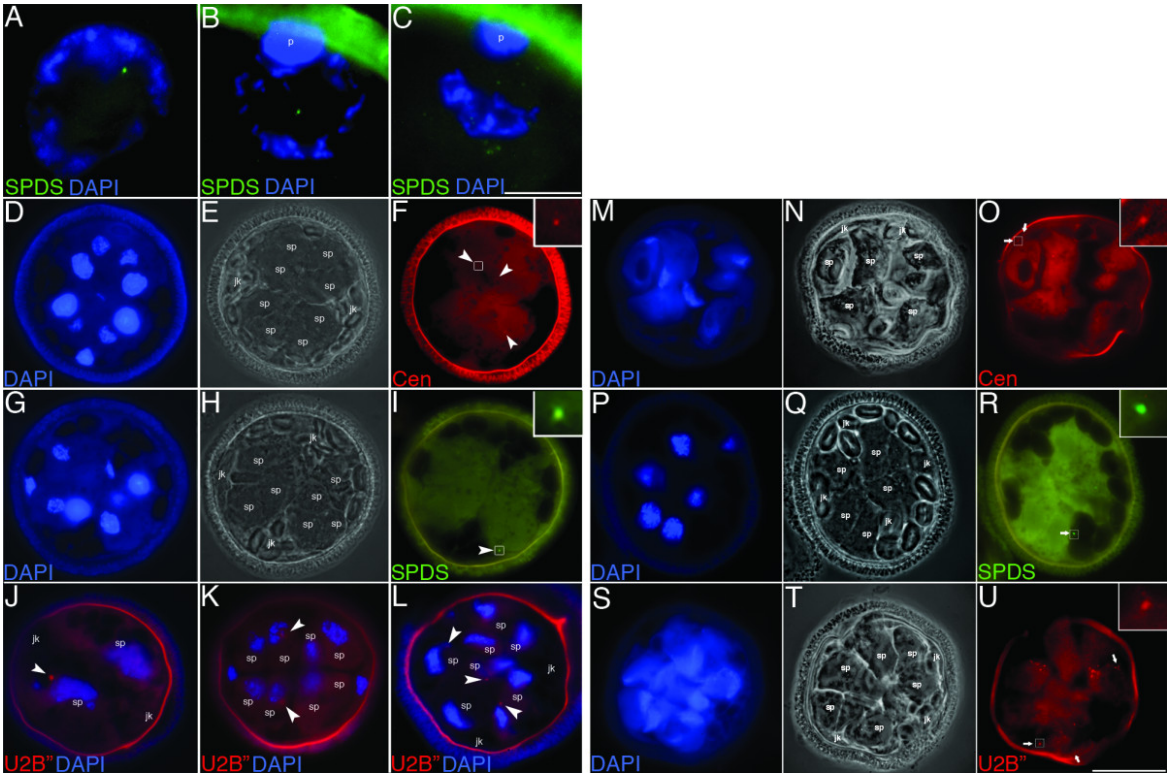


Figure IV-10. *Masked mRNAs and speckle markers enter the cytoplasm during the first division and Mago is required for their asymmetrically redistributed.* **A-C**, 25mer biotinylated probes directed against SPDS transcripts were hybridized and detected using avidin conjugated Fluorescein (green) and counter stained with DAPI (blue). Bar = 10 µm. **A**, Microspore prior to the prothallial division. **B**, Microspore during the prothallial division. **C**, Microspore just after completion of the prothallial division. The prothallial nucleus is denoted by "p" in **B** and **C**. **D-U**, "sp" denotes spermatogenous cells, "jk" denotes jacket cells, arrowheads mark foci of labeling within spermatogenous cells and arrows mark foci of labeling within jacket cells. **D-F**, Microspore fixed at 4 hours of development and probed with short FISH probes for masked centrin mRNA (red). **D**, DAPI (blue), **E** phase contrast, and **F** centrin FISH probes detected with TexasRed conjugated antibody (red). **G-I**, microspore fixed at 4 hours of development and probed with 25mer FISH probes for masked SPDS mRNA (green). **G**, DAPI (blue), **H** phase contrast, and **I** SPDS FISH probes detected with Fluorescein conjugated antibody (green). **J-K**, 4G3 labeling of U2B" protein (red) at (**J**) 2 hours, (**K**) 4 hours, and (**L**) 5 hours of development. Arrowheads denote cytoplasmic masked mRNA (**F** and **I**) and U2B" (**J-L**). **M-O**, Microspore subjected to Mv-Mago RNAi and fixed after 4 hours of development. (**M**) DAPI (blue), (**N**) phase contrast, (**O**) masked centrin transcripts (red) detected with 25mer biotinylated FISH probes. Arrows denote masked transcripts within jacket cells. Boxed regions enlarged in insets. Spermatogenous (sp) and jacket cells (jk) labeled in **N, Q and T. P-R**, Microspore subjected to Mv-Mago RNAi and fixed after 4 hours of development. (**P**) DAPI (blue), (**Q**) phase contrast, (**R**) masked SPDS transcripts (green) detected with 25mer biotinylated FISH probes. Arrows denote masked transcripts within jacket cells. Boxed regions enlarged in insets. **S-U**, Microspores subjected to Mv-Mago RNAi and fixed after 5 hours of development. (**S**) DAPI (blue), (**T**) phase contrast, (**U**) 4G3 labeling of U2B". Boothby and Wolniak, 2010.

Similar to masked mRNA, U2B” protein exhibited localized immunolabeling in aggregates within the antheridial initials (FigureIV-10). By 4 hours of development, immunolabeling revealed that U2B” protein was dispersed as clusters of punctate foci adjacent to the nuclei of the spermatogenous cells (Figure IV-10). During nuclear elongation of the maturing spermatid, anti-U2B” antibody labeled the spermatids, predominantly colocalizing with the ends of the elongating gamete nuclei (Figure IV-10). Both masked mRNA and protein were initially localized within the nucleus of the microspore, and they assumed precise distributions only in maturing spermatogenous cells. Thus, nuclear speckle components are distributed asymmetrically to the cytoplasm of spermatogenous, but not, sterile cells. This asymmetry underlies cell fate determination in the gametophyte where a single cell in the microspore gives rise to two distinct cell types, sterile cells and spermatogenous cells, in a precise a temporal and spatial framework.

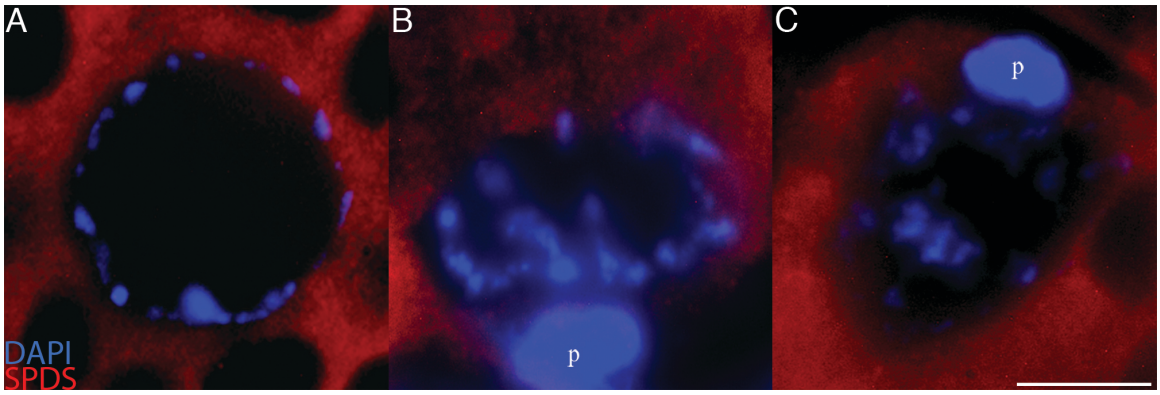


Figure IV-11. *Traditional FISH probes fail to detect masked SPDS transcripts within the nuclei of maturing microspores.* **A-C**, Traditional biotinylated probes directed against SPDS transcript (red). Pre-prothallial (**A**), mid-prothallial (**B**), and late-prothallial (**C**) division microspore. The prothallial nucleus denoted by "p" in **B** and **C**. Bar = 5 µm (red). Arrow denotes U2B" within jacket cell. Bar = 25 µm. Boothby and Wolniak, 2010.

Mv-Mago is required for the asymmetric distribution of components of the nuclear speckle aggregate

Previously, the RNAi-induced silencing of Mv-Mago, a homolog of the EJC component Mago nashi, was shown to deplete Mv-Mago levels within the spore as well as disrupt the endogenous asymmetry between spermatogenous and sterile cells in the developing male gametophyte of *M. vestita* (van der Weele *et al.,* 2007). A primary effect of depleting Mv-Mago mRNA was the loss of asymmetric centrin translation (van der Weele *et al.,* 2007), where centrin protein was no longer exclusively translated and assembled into basal bodies within spermatogenous cells, but instead, was synthesized and observed to aggregate into blepharoplast-like particles both in sterile jacket cells and spermatogenous cells within the spore wall (van der Weele *et al.,* 2007).

Since Mv-Mago may also play a role in the asymmetric distribution of nuclear speckle components, including masked centrin mRNA, the effect of Mv-Mago silencing on the distribution of both masked mRNA and associated proteins was assessed. RNAi has previously been demonstrated to abolish detectable translation of Mv-Mago (van der Weele *et al.,* 2007). Gametophytes treated with dsRNA targeting Mv-Mago transcripts for RNAi were labeled with anti-U2B”antibody as a way to determine whether Mv-Mago and the EJC play a role in the asymmetric distribution of protein associated with masked mRNA in the nucleus. RNAi silencing of Mv-Mago resulted in a range of division anomalies described previously (van der Weele *et al.,* 2007) and caused the symmetric distribution of U2B” protein to spermatogenous and jacket cells, as well as the disorganization of U2B” in spermatogenous cells (Figure IV-10S-U: compare with Figure IV-10J-L). We performed FISH on masked centrin transcripts after Mv-Mago silencing

(Figure IV-10M-O: compare with Figure IV-10D-I). Like U2B" protein, masked centrin mRNA localized symmetrically to both spermatogenous and jacket cells (Figure IV-10M-5O). The effect of Mv-Mago silencing on the asymmetric distribution of masked SDPS transcripts was also assessed (Figure IV-10P-R). Like with masked centrin transcripts (Figure IV-10M-O) and U2B" protein (Figure IV-10S-U), loss of Mv-Mago resulted in the symmetric distribution of masked SPDS transcripts (Figure IV-10P-R). These differences in transcript and protein distribution were also apparent in gametophytes that exhibited only minor anomalies in their cell division patterns after Mv-Mago silencing (Figure IV-12; see van der Weele *et al.,* 2007). Together, these results indicate that components associated with nuclear speckles become asymmetrically redistributed to the cytoplasm of spermatogenous cells by a mechanism dependent on Mv-Mago and likely to involve the EJC (compare localization of foci in spermatogenous cells (arrowheads) to localization of foci in jacket cells (arrows) in Figure IV-10D-U). In addition, the subcellular localization of U2B" protein to the ends of elongating chromosomes appears to be mediated by Mv-Mago.

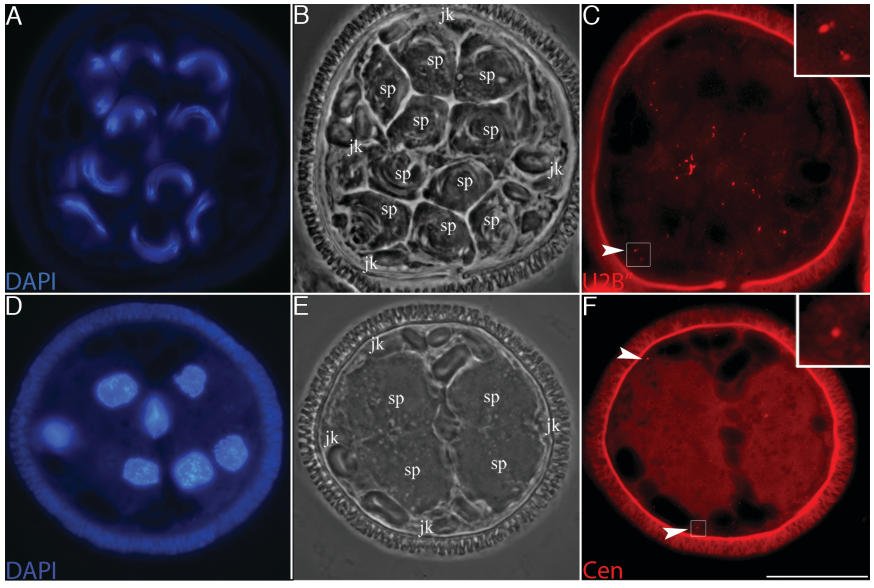


Figure IV-12. *Defects in asymmetric division are not requisite for symmetric distribution of subnuclear material in Mv-Mago knockdowns.* **A-C**, microspore subjected to Mv-Mago RNAi and fixed after 5 hours of development. (**A**) DAPI (blue), (**B**) phase contrast, (**C**) 4G3 labeling of U2B" (red). **D-F**, representative microspore subjected to Mv-Mago RNAi and fixed after 4 hours of development. (**D**) DAPI (blue), (**E**) phase contrast, (**F**) masked centrin transcripts (red) detected with 25mer biotinylated FISH probes. Spermatogenous cells denoted by "sp" and jacket cells denoted by "jk." Bar = 25 µm. Boothby and Wolniak, 2010.

Since nuclear speckles in *M. vestita* appear to contain a subset of stored RNA, and since it appears that post-transcriptional splicing (a process known to occur within nuclear speckles; Girard *et al.*, 2012) of IRTs is essential for differentiation of spermatids, we investigated the localization of IRTs within developing microspores. Using *in situ* probes directed against the intronic portion of IRTs known to be spliced during development we were able to localize IRTs to the nuclei of microspores prior to the first division. IRT dynamics during development were identical to those of masked speckle-associated RNAs (Figure IV-13), except that after the time when a particular IRT

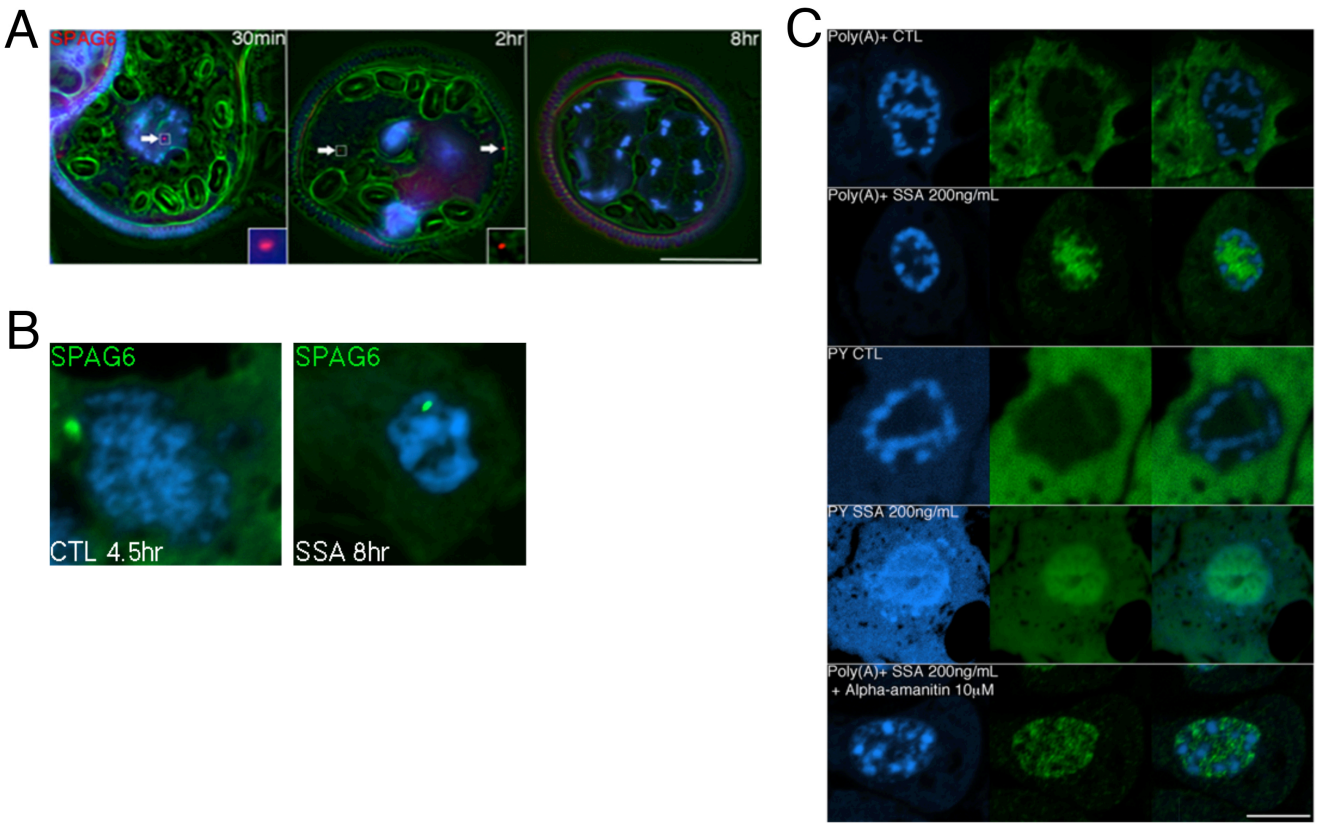


Figure IV-13. *IRT dynamics mirror those of speckle associated RNA and inhibition of splicing leads to their nuclear accumulation.* (**A**), FISH probes (red) directed against the intron of IRT SPAG6 were used to label control microspores fixed after 30 minutes (before the first division), 2 h (after the first division), and 8 h (after SPAG6 IRT is spliced). Bar = 25 mm. T.C.C. Boothby, unpublished. (**B**), Cropped images of spermatogenous nuclei. Labeling of 4.5 h control microspores or 8 h SSA (200ng/mL) treated spores (development arrested ~4.5 h) with SPAG6 intronic probes (green). T.C.C. Boothby, unpublished. (**C**), Cropped images of spermatogenous nuclei. Label of total (PY) or poly(A)+ RNA (both green) in control and SSA or SSA+amanitin treated spores. Boothby *et al*., 2013. (**B** & **C**), Bar = 5 mm. (**A-C**), counter staining with DAPI (blue).

is known to be spliced, it can no longer be detected in via *in situ* hybridization. This is presumably because after an intron is spliced out it is degraded (Clement *et al*., 1999).

We reasoned that since inhibition of splicing blocks the maturation of IRTs that it should also stabilize the labeling of these transcripts intronic portions via *in situ* hybridization. We treated microspores with Spliceostatin A to inhibit translation after which we could label intronic portions of IRTs with *in* situ probes, even after the time when these IRTs should have been spliced (Figure IV-13). However, it was observed that during splicing inhibition, these stabilized IRTs were localized to spermatogenous nuclei as opposed to the cytoplasm of spermatids. Labeling of poly(A)+ and total RNA in cells treated with Spliceostatin A revealed that these species of RNA also appear to localize within the nuclei of spermatogenous cells when splicing is inhibited (Figure IV-13). This phenomenon persists even in the presence of transcriptional inhibitors, suggesting that the RNA that is sequestered within the nuclei of spermatogenous cells following splicing inhibition is not newly transcribed but rather results from the import of stored RNA from the cytoplasm (Figure IV-13).

Conclusions/Discussion

Transcriptional silencing and dormancy triggers the coalescence of nuclear speckles and the storage of masked mRNA in the drying microspore of M. vestita

The male gametophyte of *M. vestita* relies on the regulated translation of stored mRNA for the rapid production of motile spermatozoids. The packaging of these transcripts for long-term storage in the spore is an essential mechanism required for rapid development leading to male gamete formation. In this chapter, we have shown that

nuclear speckles enlarge and aggregate as transcription is silenced in the desiccating microspore. In other systems when transcription is blocked, nuclear speckles often enlarge and serve as sites of storage for pre-mRNA processing machinery and under these conditions, a subset of poly(A)+ RNA is retained within the nucleus associated with nuclear speckles (Huang *et al.,* 1994).

Transcription in the microspore remains silent even upon hydration and recovery from dormancy. We have shown that during spermatogenesis, nuclear speckles are maintained as a single coalescence in the nucleoplasm of the newly hydrated microspore. Morphologically, the nuclear speckle aggregate superficially resembles a nucleolus. However, the lack of colocalization with traditional nucleolar markers (Fibrillarin and rRNA) in the subnuclear aggregate reduces the likelihood of this possibility. Moreover, our studies of the desiccating microspore demonstrate a marked loss of nucleolar markers during microspore entry into quiescence. Like many dormant systems that stockpile polysomes (Bewley, 1973; Bachvarova and Leon, 1977; Tate and Marshall, 1991), our findings show that a traditional nucleolus is absent from the desiccated microspore and suggest that little, if any, new ribosome biogenesis occurs during spermatogenesis in *M. vestita*.

U2B", which is present in the subnuclear aggregation, is commonly used as a marker of Cajal bodies (CBs) in both animals and plants. Is the aggregate a CB? Coilin is widely considered to be a diagnostic marker of the CB in animals (Andrade *et al*., 1991), where it functions to concentrate and facilitate snRNP formation (Strzelecka *et al*., 2010). Unfortunately, coilin proteins in animals and plants are highly divergent, and they are sufficiently distinct so that anti-coilin antibodies used to identify CBs in animal cells

show no specific affinity for plant coilins (Tucker and Matera, 2005). Despite the presence of U2B" within the subnuclear aggregate of *M. vestita*, no CBs examined have been shown to be associated with poly(A)+ RNA (review: Gall, 2000). In addition to poly(A)+ RNA, CBs are known to lack SC35 (Gama-Carvalho *et al*., 1997). Using the absence of poly(A)+ RNA and SC35 as criteria for CB identity, it does not appear that the subnuclear aggregation present in the microspore should be designated as a CB since it contains both polyadenylated RNA and the essential splicing factor SC35.

Nuclear speckles contain U2B" and SC35, and a subset of poly(A)+ RNA. In addition to similarities in composition, the behavior of the aggregated nuclear material in the microspore of *M. vestita* resembles traditional nuclear speckles more closely than other kinds of nuclear inclusions. Typically, nuclear speckles consist of small, interchromatin aggregations of pre-mRNA splicing proteins, but have been shown to enlarge upon transcriptional inhibition (Huang *et al.,* 1994). This enlargement apparently results from an accumulation of pre-mRNA splicing machinery in the speckles and these stores are utilized once transcription is reinitiated. In addition to the accumulation of pre-mRNA splicing machinery, previous work has shown that transcriptional inhibition causes a subset of poly(A)+ RNA to be sequestered within nuclear speckles (Huang *et al.,* 1994). Our observations of desiccating microspores suggest that a similar pattern occurs; TBO and PY staining as well as SC35 antibody labeling reveal that small aggregations coalesce into larger accumulations, finally resulting in a single large aggregation concurrent with the onset of transcriptional quiescence. The similarities between aggregated SC35, U2B" and poly(A)+ RNA in the microspore of *M. vestita* and traditional speckles are inescapable. In addition, during the onset of mitotic divisions

during spermatogenesis, nuclear speckle material enters the cytoplasm as is typical of the speckle cycle observed in other systems.

Beyond the storage of pre-mRNA processing machinery, nuclear speckles could serve as sites for the storage of mRNA during periods of inhibited transcription and/or splicing. Intriguingly, a subset of speckle associated poly(A)+ RNA is masked mRNA (SPDS and centrin) that is known to be essential for gametophyte development in *M. vestita*. It is important to point out that with the required use of Proteinase K in our *in situ* hybridization protocols, FISH and immunofluorescence labeling had to be performed sequentially as described in the Methods section. Immunofluorescence labeling had to be performed first, followed by imaging and removal of coverslips, prior to treating the cells for FISH. We attempted a number of methods that would allow simultaneous immuno- and *in situ* hybridization labeling on the same sections of gametophytes, but we had to resort to sequential labeling in order to obtain reproducible patterns of antibody and RNA distributions in the cells. The sequential method used has the unfortunate drawback that in between imaging sessions, specimen sections can become distorted because of the harsh incubation conditions for FISH or ISH. We feel that this is why in all such experiments (Figure IV-2) subnuclear poly(A)+ RNA appears to be more diffuse than immunofluorescence labeling or nuclear poly(A)+ labeling alone.

Asymmetric distribution of nuclear speckle components to the cytoplasm of spermatogenous cells

Nuclear speckles remain aggregated within the nucleus of the microspore until the first division. During the first division, both protein and RNA associated with nuclear speckles enter the cytosol adjacent to the nucleus of the antheridial initial. As additional division cycles progress, this material is asymmetrically distributed to spermatogenous cells, but not to jacket cells. During spermatogenesis, foci of masked transcripts were visible in the cytoplasm directly adjacent to the nuclei of spermatogenous cells, but were not seen in jacket cells (Figure IV-14).

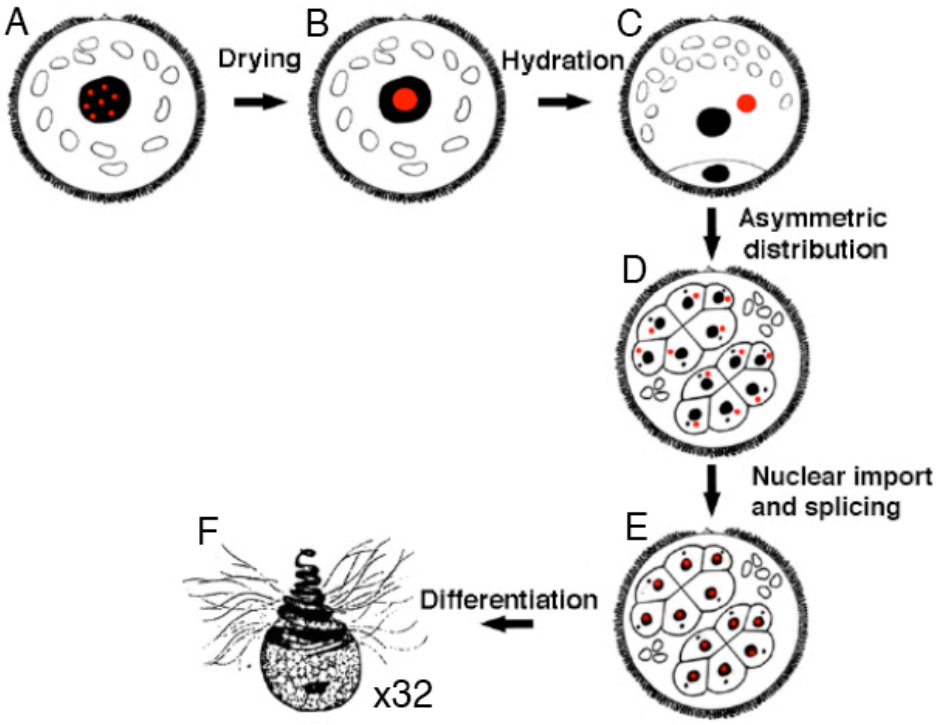


Figure IV-14. *Working model*. (**A**), Dispersed nuclear speckles and associated RNA (red) aggregate during drying and entry into transcriptional quiescence (**B**). Aggregated speckles persist within the nuclei of desiccated microspores until rehydration and then enter the cytoplasm of the antheridial initial during the first mitotic division (**C**). (**D**), speckles and their associated RNA are asymmetrically distributed to the cytoplasm of spermatogenous but not sterile cells. (**E**), Speckles must reenter the nucleus to mediate post-transcriptional splicing of associated IRTs. (**F**), Once spliced, former IRTs can be translated and their translation mediates differentiation of spermatozoa. Adapted from drawing by S.M. Wolniak; Sharp, 1914.

Sectioned material was used for the analysis of this asymmetric masked transcript distribution, because FISH and immunofluorescence assays cannot be performed together on intact, but fixed spores, even if they are imaged by confocal microscopy. Thus, while the foci of masked centrin and SPDS RNAs should not be expected to be visible in all of the spermatogenous cells in a single section of a single gametophyte, observations of many sections of many gametophytes provide strong indications that all of the spermatogenous cells contain these foci of masked transcripts.

While the ISH and FISH assays carried out in this study were not directed against RNA contained exclusively within jacket cells, previous studies (Deeb *et al*., 2010) have demonstrated that the techniques used here are capable of labeling transcripts contained exclusively within jacket cells. In addition Figure IV-10 shows the labeling of jackets cells, confirming their accessibility to our probes and hybridization techniques.

There are two critical covariants in gametophyte development: a) the patterns of divisions leading to the formation of sterile and spermatogenous cells are precise inside the spore wall, and b) there is no cell movement during development. These factors highlight an essential process in cell fate determination in the gametophyte: the fragmentation of aggregated nuclear speckle-associated masked transcripts must be orderly to ensure that appropriate complements of mRNAs are distributed among all of the spermatids. The movements of masked transcripts must be under strict control during the successive division cycles, so that appropriate parcels of transcripts are ultimately allocated among the 32 spermatids in the gametophyte. Thus, masked mRNA, originally

stored in a single coalescence of nuclear speckles within the nucleus of the desiccated microspore, is dispersed as a set of masked RNA-containing particles that are uniformly distributed among spermatogenous cells, but not to sterile jacket cells, which arise from the same progenitors.

The EJC exerts profound effects on the symmetry of divisions in the developing gametophyte of *M. vestita* (van der Weele *et al*., 2007). By silencing EJC components, we induced perturbations of division plane locations that effectively disrupted spermatid-specific events. Here, we show that dsRNA-mediated silencing of Mv-Mago also disrupted the asymmetric distribution of centrin masked mRNA and U2B" protein (Figure IV-10, IV-12). Mv-Mago clearly mediates the cell type-specific distribution of stored nuclear transcripts.

In other plants, the EJC, and in particular Mago nashi, have been shown to be essential for development in nearly all tissues and organs. The silencing of atMago in *Arabidopsis* (Park *et al*., 2009) results in aberrant microspore tetrad arrangement. eIF4A-III, another core component of the EJC relocalizes during stress (Koroleva *et al*., 2009a; Koroleva *et al*., 2009b) in nucleoli and nuclear speckles. When present in nuclear speckles, eIF4A-III displays lowered mobility suggesting that it is specifically retained within these subdomains (Koroleva *et al*., 2009a). The presumed consequence of this relocalization is that mRNA associated with eIF4A-III would not exit the nucleus and thus its translation would be inhibited (Koroleva *et al*., 2009a; Koroleva *et al*., 2009b). We believe a similar mechanism could be at work in the microspores of *M. vestita* where mRNA is stored during desiccation in association with nuclear speckles. We propose that

the association of this RNA with these speckle components ensures nuclear retention that in turn, forestalls its translation.

A mechanism for cell type-specific translation of centrin

Mv-Mago silencing results in the symmetric distribution of masked centrin mRNA as well as the symmetric translation of centrin in both spermatogenous and jacket cells at 4 hours of development. Masked centrin mRNA initially stored in the nucleus of the microspore is clearly required for centrin protein production in the spermatogenous cells of the gametophyte. It follows that additional levels of translational regulation must be in place to inhibit the translation of cytoplasmic centrin mRNA, which is ubiquitously present in all cells at all times during spermatogenesis (Tsai *et al*., 2004). Although qc-centrin transcripts are distributed throughout the cytosol of all cells in the gametophyte from the onset of development, translation of centrin protein occurs only after 4 hours of development, and only within spermatogenous cells. In addition, previous research from our laboratory (Deeb *et al*., 2010) has shown that masked subnuclear RNA that is initially undetectable using long *in situ* hybridization probes later becomes detectable with these long probes in the cytoplasm of spermatogenous cells at time points when the corresponding proteins become abundant (Deeb *et al*., 2010). When combined with information from *in vitro* translation assays (Hart and Wolniak, 1998) that show RNA isolated from these spores cannot be translated prior to the first division of the gametophyte when subnuclear transcripts are apparently released from the nucleus, these data suggest not only that masked subnuclear mRNA plays an important role in spermatogenesis, but also that cytoplasmic stores of RNA remain translationally

quiescent during gametophyte development. The differential regulation of translation for masked RNA and quiescent cytoplasmic transcripts is not fully apparent at this time, but an association with nuclear speckles appears to be essential for the translation of centrin mRNA, and probably other proteins as well. Further elucidation of this mechanism(s) will be essential for our understanding fate determination in the highly ordered and transcriptionally quiescent gametophyte.

IRT association with nuclear speckles

Since nuclear speckles are the site of post-transcriptional splicing (Girard *et al*., 2012) and post-transcriptional splicing mediates the temporal use of IRTs during spermatid differentiation in *M. vestita* (Boothby *et al*., 2013), it is not surprising to see that IRTs make up at least a subset of speckle associated RNA.

In Chapter IV and our recently published work (Boothby *et al*., 2013) Spliceostatin A (SSA) was used to block splicing and this treatment resulted in a stabilization of IRT RT-PCR products. This stabilization of IRT isoforms can also be seen using *in situ* probes directed against the intronic portion of IRTs. Interestingly, in spores treated with SSA and examined after the time when a particular IRT is supposed to be spliced, the stabilized isoform is now localized within the nucleus, as opposed to the cytoplasm. The likely explanation for this is that IRTs are reimported back into the nucleoplasm (where splicing is known to occur) as part of the speckle cycle, but the inhibitor prevents splicing from occurring. Since divisions have been completed, the nuclear envelope will not break down again and pre-mRNAs are typically not exported from intact nuclei. Thus, these transcripts have no way of escaping back to the cytoplasm.

Not only does the inhibition of splicing block pre-mRNA maturation, which is requisite for translation (Boothby *et al*., 2013), but it also appears to cause nuclear sequestration, which represents a physical separation of IRTs and cytoplasmic ribosomes (translational machinery).

It is of interest to note that masked centrin transcripts associate with nuclear speckles despite the fact that we have not been successful in identifying any centrin IRT isoforms. This suggests IRTs make up only a subset of speckle associated RNA. How mRNA associates with nuclear speckles, and how its temporal use is regulated remains a question under investigation. However, it is intriguing that inhibition of splicing does not affect the translation of non-IRT centrin (Figure III-3), while it simultaneously inhibits the translation of IRT SPDS (Figure III-3).

Future perspectives

The discovery of a coalesced nuclear speckle aggregate, whose functions include the storage and masking of developmentally important transcripts, reveals an important level of post-transcriptional regulation affecting rapid development of spermatozoids in *M. vestita*. We believe this coalescence contributes substantively to the long-term storage of stable, masked RNAs that may be present as fully or partially processed transcripts, which are essential for the formation of spermatozoids within hours after the dry microspore is hydrated. Furthermore, it is apparent that the mode of storage of these transcripts may play a role in their translational regulation and their asymmetric distributions during spermatogenesis.

Paraspeckles, subnuclear aggregates closely associated with nuclear speckles (Fox *et al*., 2002) retain CTN-RNA, which is a non-coding mCAT2 transcript (Parsanth *et al*., 2005). Under stress, CTN-RNA is cleaved, thereby releasing protein-encoding mCAT2 mRNA, which is quickly localized to the cytosol for translation (Parsanth *et al*., 2005). Our finding that nuclear speckles participate in the storage of masked mRNA builds on the paradigm that the nuclear retention and post-transcriptional maturation of transcripts plays an important role in gene expression.

Many subnuclear bodies including nuclear speckles, have a cytoplasmic phase in their cycles (Spector and Smith, 1986; Reuter *et al*., 1985; Ferreira *et al*., 1994; Thiry, 1995; Alliegro *et al*., 2010) and there is mounting evidence that the displacement of some bodies to the cytoplasm plays a key role in cellular function. For example, the nucleolinus of *Spisula* oocytes enters the cytoplasm following activation, and centrosomes form within the diffusing nucleolinus. In parthenogenetically-activated oocytes, laser ablation of the nucleolinus results in a failed meiotic division and microtubule disorganization, demonstrating a clear functional role for the nucleolinus in spindle formation and cell division (Alliegro *et al*., 2010). We suspect that in addition to the storage of masked mRNA, nuclear speckles could serve to regulate the asymmetric distribution and translation of these messages in the cytoplasm of spermatogenous cells.

Ongoing and future characterizations of the subnuclear masked mRNA and associated processing machinery should provide greater insights into regulatory mechanisms that underlie this rapid developmental process. We believe that traditional 'long' ISH or FISH probes could not detect masked mRNA because masking agents most likely proteins) obscure sites of hybridization. We believe because of their small size, our

25mer probes can be fit in between their masking agents and thus robustly label masked mRNA. We would like to identify these masking agents and further assess their roles in development and RNA regulation. An obvious area of interest centers on the fragmentation process and the distribution of masked transcripts among the spermatids. Since both centrin and SPDS masked transcripts become localized exclusively in foci within spermatogenous cells, these transcripts can reveal patterns of storage, movement and unmasking for particular mRNAs during development. Our FISH assays show that transcript pools may be stored discretely, so how and when are these particles fragmented and passed on to spermatogenous cells? If masked transcripts are stored in multiple foci, how does each spermatid receive a full complement of masked transcripts? How is the timing of translation controlled for specific transcripts during development? The mechanisms that control how the subnuclear RNA and pre-mRNA processing machinery becomes asymmetrically distributed between spermatogenous and sterile cells remains unclear. Positioning of the subnuclear material in the cytoplasm during the prothallial division clearly underlies this process, and identifying the components and factors that affect movements of these molecules will be important in understanding cell fate determination mechanisms in the highly ordered gametophyte.

The regulated reimport of nuclear speckles and associated IRTs into the nucleus where splicing occurs could be a step regulating the differential timing of IRT maturation. If pre-mRNAs are imported back to the nucleoplasm at different times, this process could account for differences in splicing patterns seen between IRTs (Boothby *et al*., 2013). It has previously been observed that MIG reentry into the nucleus is conducted

in a stepwise fashion (Ferreira *et al*., 1994; Prasanth *et al*., 2003) and this could account for differences in timing of IRT import.

Conclusions

We show here that translationally masked RNA is stored within aggregated nuclear speckles during desiccation and dormancy in the microspore of *M. vestita* and that a subset of this RNA consists of IRTs. In addition, we show that both protein and nucleic acid components of this speckle aggregate are asymmetrically localized to spermatogenous cells during development. This localization requires the EJC core component Mago nashi. Asymmetric localization of masked mRNA mirrors both the temporal and spatial patterns of their corresponding proteins. The inhibition of splicing results in the stabilization of speckle associated IRTs within the nuclei of spermatogenous cells. Thus, it appears that nuclear speckles play a role in the spatial regulation of IRT utilization through their asymmetric distribution to spermatogenous but not sterile cells.

Chapter V – Conclusions and Future Directions

Intron retention and speckle mediated post-transcriptional splicing regulate the use of stored RNA in the development male gametophyte of *M. vestita*

The aim of this dissertation was to gain a deeper understanding of how the use of stored RNA is regulated in the developing male gametophyte of *M. vestita*. Our transcriptome analysis (Chapter II) revealed that many post-transcriptional mechanisms involved in the modification of transcripts are present during spermatogenesis. In particular, transcripts encoding factors involved in post-transcriptional splicing were enriched. Additionally, our transcriptome was found to contain a subset of pre-mRNAs that harbor retained introns.

Experimental evidence presented here (Chapter III) suggests that intron retention is a mechanism forestalling the translation of stored pre-mRNAs. At distinct time points during development, post-transcriptional splicing occurs, serving as a temporal regulator of transcript utilization that results in fully mature and translatable mRNA.

In Chapter IV, data are presented that support the recent finding (Girard *et al*., 2012) that post-transcriptional splicing occurs in nuclear speckles. We also identified nuclear speckles as mediators of spatial regulation of stored transcripts, as they participate in the asymmetric distribution of pre-mRNA to spermatogenous, but not sterile cells.

In conclusion, intron retention and speckle mediated post-transcriptional splicing are mechanisms mediating both the temporal and spatial use of stored RNA in the developing male gametophyte of *Marsilea vestita*. IRTs are associated with nuclear

speckles during desiccation and the retention of an intron blocks their translation. Speckles mediate the asymmetric distribution of these masked IRTs to spermatogenous cells but not sterile cells in the gametophyte. Nuclear speckles reenter the nucleus in order to mediate post-transcriptional splicing of IRTs, resulting in the controlled release of full mature and translatable transcripts at the appropriate times during rapid development.

The potential for pre-association of IRTs with spliceosomal complexes, the speckle cycle, and the temporal regulation of splicing

The identification of speckle-associated IRTs and the observation that nuclear speckle components are asymmetrically distributed presents an obvious theoretical mechanism for controlling the spatial use of stored RNA. The discovery that different IRTs are spliced at distinct times requires that there are additional levels of regulation mediating the temporal maturation of these transcripts.

We reasoned that since the translation of stored transcripts encoding splicing factors would likely be essential for the temporal maturation of IRTs, we should identify conserved spliceosomal components and analyze their abundance and distribution patterns during development. From our RNAseq G.O. enrichment studies we knew that components of the catalytic spliceosome are enriched early in development. To identify potential splicing factors, a refseq_protein database was established using components associated with the NCBI BioSystems entries for yeast, *Arabidopsis*, and human spliceosomes. We conducted reciprocal best match BLAST analyses (BLASTx and tBLASTx) with our reference transcriptome against these databases.

Sequences of putative, non-redundant splicing factors identified through reciprocal best match BLAST searches were assembled into a reference for RNAseq read mapping. Reads from 1-2 h, 3-5 h, and 6-8 h were mapped to this reference and FPKM values estimated. Using FPKM values from different time ranges, transcripts encoding putative splicing factors were ranked on the basis of their temporal abundance during spermatogenesis. These transcripts were cross-referenced with the NCBI BioSystems entries for both precatalytic and catalytic spliceosomes to determine what transcript encoded proteins were associated with particular spliceosome complexes. For those transcripts whose reciprocal best match BLAST identities were not annotated in either of these BioSystem databases, literature searches were used to establish what complex of the spliceosome they are associated with. Since the spliceosome is known to assemble in a stepwise and highly stereotyped fashion, we suspected that transcripts encoding proteins involved in the early spliceosome would be abundant early, and transcripts encoding proteins involved in late spliceosome maturation would be abundant later in development. To our surprise, the gametophytes displayed the opposite pattern. We found that the vast majority (82.76%) of transcripts that have their peak abundance early in development are unique to the catalytic spliceosome, where as only (3.45%) splicing components with peak abundances early in development are from the precatalytic spliceosome (the remaining 13.79% of early abundance splicing factors are associated with both complexes). Alternatively, relatively fewer transcripts (20.93%) that are abundant late in development are associated with the catalytic spliceosome, and more (32.56%) splicing factors abundant late in development are associated with the

precatalytic spliceosome (the remaining 46.51% of late abundant splicing factors are found in both complexes; Figure V-1).

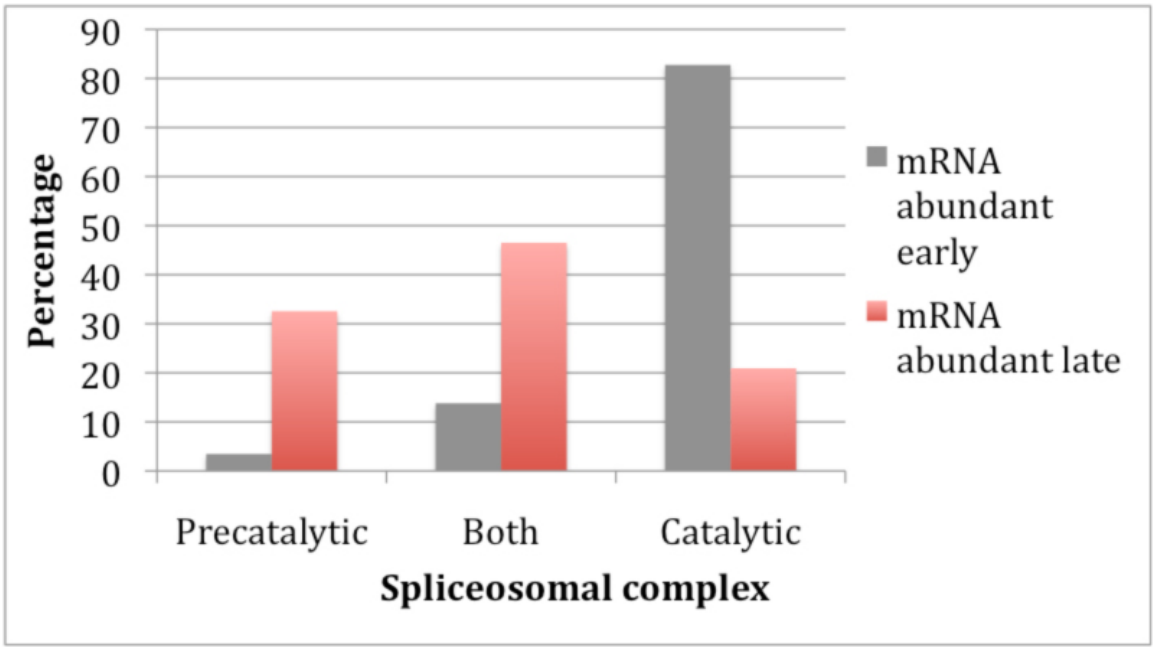


Figure V-1 Unmasking of transcripts encoding *catalytic spliceosome components precede pre-catalytic spliceosome component during spermatogenesis in M. vestita*. Non-redundant putative spliceosomal components were cross-referenced with NCBI BioSystems entries for pre-catalytic and catalytic spliceosomes. Using FPKM values, these transcripts were assigned as most abundant "early" (1-2 h) or "late" (6-8 h). T.C.C. Boothby, unpublished.

Since the temporal abundance of splicing factors involved in early and late spliceosome maturation is the opposite of the sequential order of spliceosome assembly, we conducted RNAi experiments to deplete transcripts encoding components associated with pre-catalytic, catalytic, or both spliceosomal complexes (Figure V-2). RNAi experiments allow us to establish the time of development when stored transcripts are essential. As expected, the timing of RNAi-mediated perturbation mirrored the temporal abundance of splicing factors. This was evidenced by the lack of elongated nuclei in

catalytic spliceosome knockdowns and the elongation (and sometimes coiling) of nuclei in knockdowns of splicing components associated with the pre-catalytic or both spliceosomal complexes. Our previous findings showed that masked RNA is associated with splicing factor components of nuclear speckles (U2B" and SC35; Boothby and Wolniak, 2011) and that IRTs are spliced at different times (Boothby *et al*., 2013). In finding that splicing factors unique to catalytic spliceosomes are abundant and essential early in development, before splicing factors associated with the pre-catalytic spliceosome become abundant leads us to suspect that IRTs are stored in association with spliceosomes in various stages of maturation (Figure V-3).

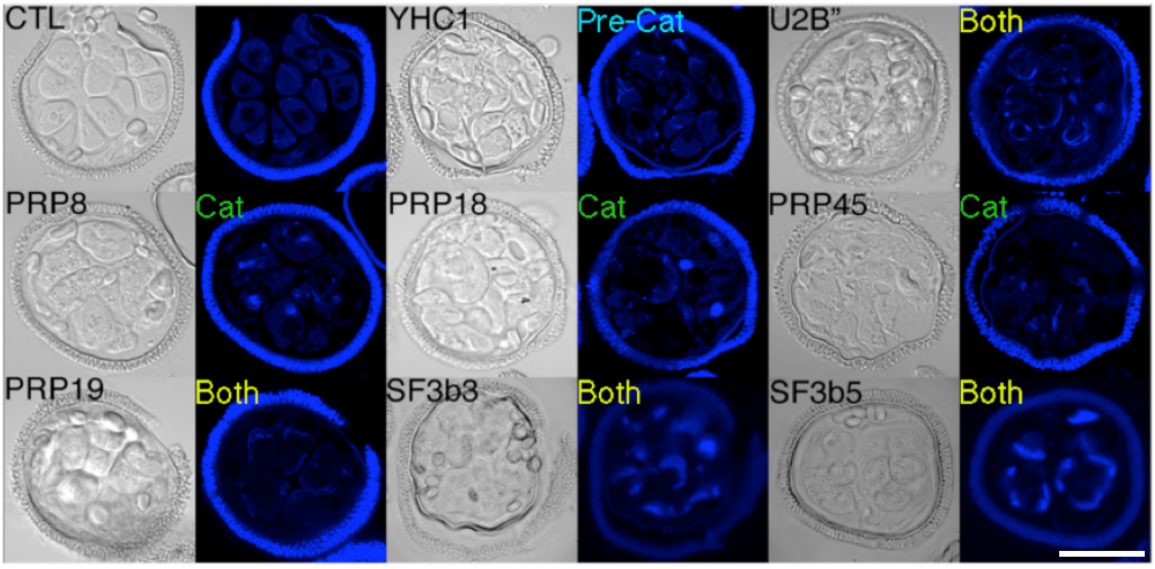


Figure V-2 *Splicing factors are essential during their period of highest abundance.* RNAi constructs targeting the indicated spliceosomal component were introduced to microspores at the time of imbibition. Microspores were allowed to develop for 8 h and then fixed, sectioned, and stained with DAPI (blue). T.C.C. Boothby, unpublished.

The maturation of the spliceosome takes place through a stepwise assembly of different spliceosomal complexes. The E complex forms first, which is followed by the A complex. Maturation of the A complex to the B complex is followed by activation of the B complex to form the catalytic C complex. Theoretically, then, IRTs could be

differentially associated with E, A, and B complex splicesomes in the dormant microspore. Transcripts essential to the formation of the C complex become unmasked early and are then translated, allowing B, but not E and A complexes to mature. Later, recycled C and B complex proteins plus the translation of unmasked pre-catalytic spliceosome components would then allow for the maturation of E and A complex spliceosomes (Figure V-3). In this way, differential maturation of spliceosomal complexes could regulate the timing of splicing for their associated IRTs. The further elucidation of mechanisms involved in mediating the timing of splicing of different transcripts will be essential in understanding post-transcriptional regulation in this system.

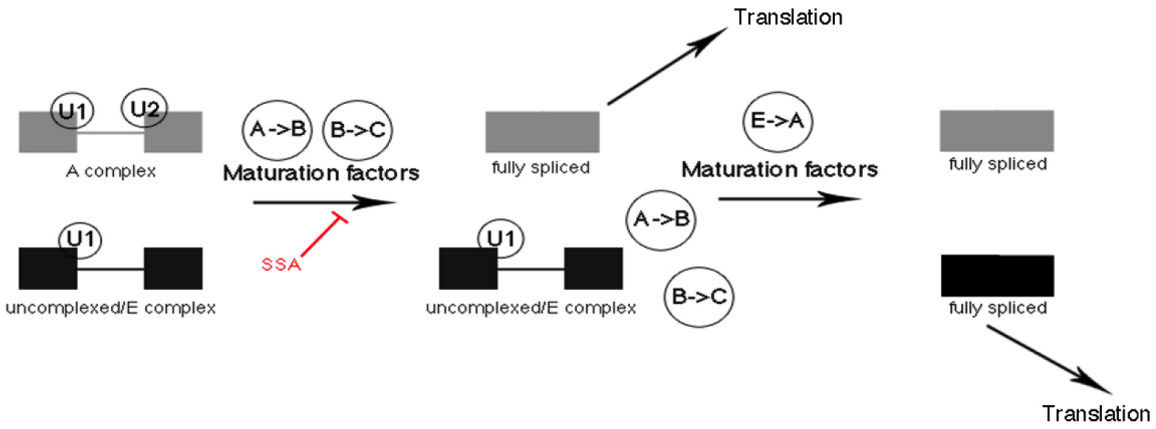


Figure V-3 *Model for differential timing of splicing through pre-association of IRTs with different spliceosomal complexes.* Exons are represented by boxes and introns by intervening lines. Early in development, transcripts encoding factors required for maturation of the catalytic spliceosome are unmasked and translated. These factors allow for the maturation of A and B spliceosomal complexes, but not E complexes, leading to the splicing of some IRTs. Unmasking of factors required for early spliceosome maturation, coupled with recycling of later maturation factors allows for the production of catalytic spliceosomes from E complexes. This results in a second wave of IRT splicing and translation. T.C.C. Boothby, unpublished.

As a next step, RNA immunoprecipitation RNA sequencing (RIPseq) could be performed by isolating RNA from pull-downs of splicing factors unique to different spliceosomal complexes. This would allow for the detection of differential spliceosomal

complex/IRTs associations. Additionally, SSA treatments could be performed and RNA isolates used for RNA sequencing. Since SSA blocks maturation of the spliceosome from the A to B complex, any IRTs associated with B or C complex spliceosomes should still be spliced, while those associated with E or A complexes would not be. RNAi mediated knockdowns could also be performed on transcripts encoding splicing factors unique to different complexes; however, the presence and recycling of preexisting proteins might obscure these results. It is interesting to note that the literature contains seemingly contradictory evidence as to the ability of the spliceosome to form prolonged stable complexes (Denis *et al*., 2005; Hoskins *et al*., 2011). The developing male gametophyte of *M. vestita* presents a simplified developmental system in which to address questions surrounding the possibility of spliceosome formation associated with retained introns for the purpose of priming them for post-transcriptional splicing.

Stored RNA is subject to multiple levels of post-transcriptional regulation

From the experiments presented here, as well as previously published research, it is clear that additional forms of post-transcriptional regulation participate in mediating the use of stored transcripts in these cells (review: Wolniak *et al*., 2011).

For example, mRNA of Mv-Cen1 is present in microspores from the onset of development, though the transcript is only translated after 4 hours of development. Although masked Mv-Cen1 mRNA associates with nuclear speckles, to date we have not identified an IRT isoform for Mv-Cen1. The translation of Mv-Cen1 appears to be independent of splicing; however, only cells that receive speckle associated Mv-Cen1 produce Centrin protein. This suggests that nuclear speckles could also be essential for

the asymmetric distribution of non-IRT transcripts and that this regulation is independent of spliceosome maturation.

Unpublished results suggest that Mv-Cen1 transcripts may undergo differential polyadenylation (Figure V-4). Elongation of poly(A) tails has been observed in other systems to increase the translational rates of transcripts (Gorgoni and Gray, 2004).

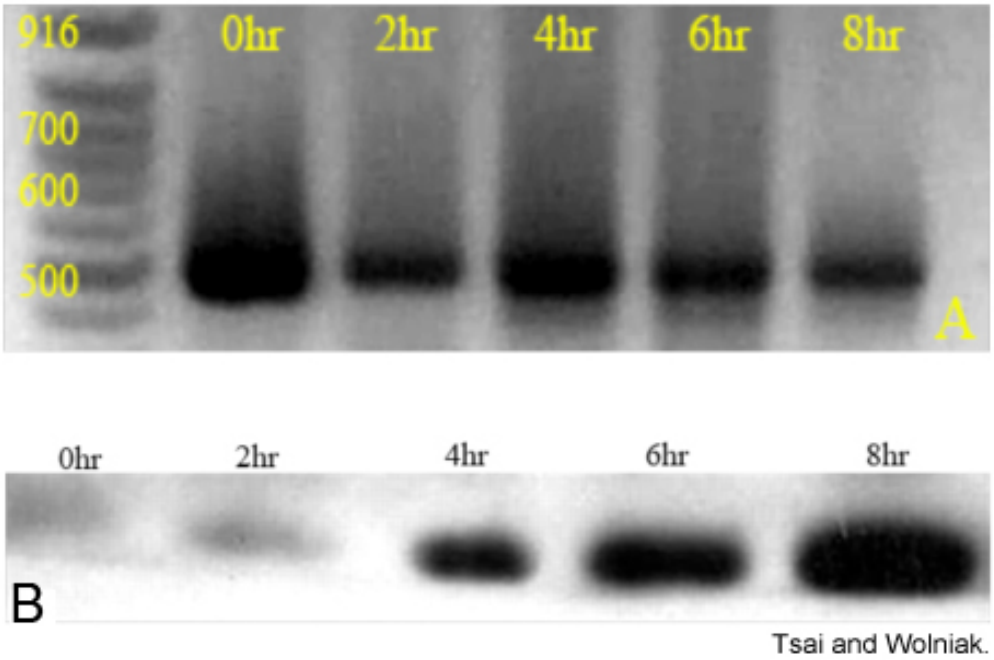
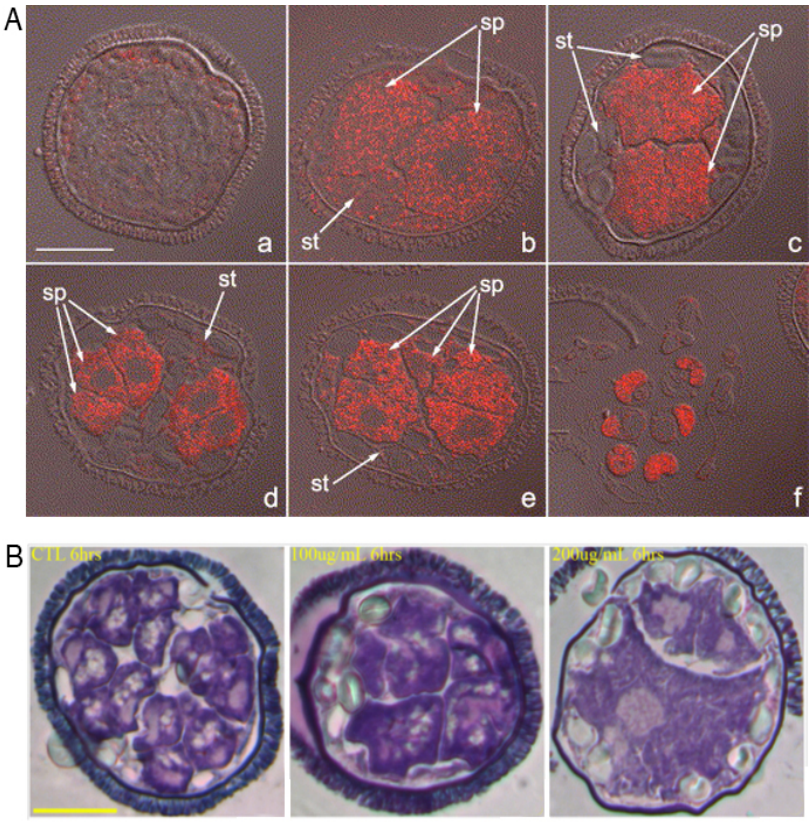


Figure V-4 *Mv-Cen1 RNA is present from the onset of development and its differential polyadenylation mirrors translation of Mv-Cen1 protein.* (**A**), Rapid Amplification of cDNA Ends Polyadenylation Test (RACE-PAT) for Mv-Cen1 using a unique gene specific 3' UTR primer was performed to assess the distribution of poly(A) tail lengths associated with this transcript at different times. (**B**), Protein blots using a monoclonal antibody that uniquely labels Mv-Cen1. Tsai and Wolniak, 2001.

Components involved in differential cytoplasmic polyadenylation are known to localize predominantly in spermatogenous cells during male gametophyte development in *M. vestita* (Figure V-5; Tsai *et al*., 2004). RNAi mediated depletion of transcripts encoding components involved in differential polyadenylation result in perturbed divisions in the

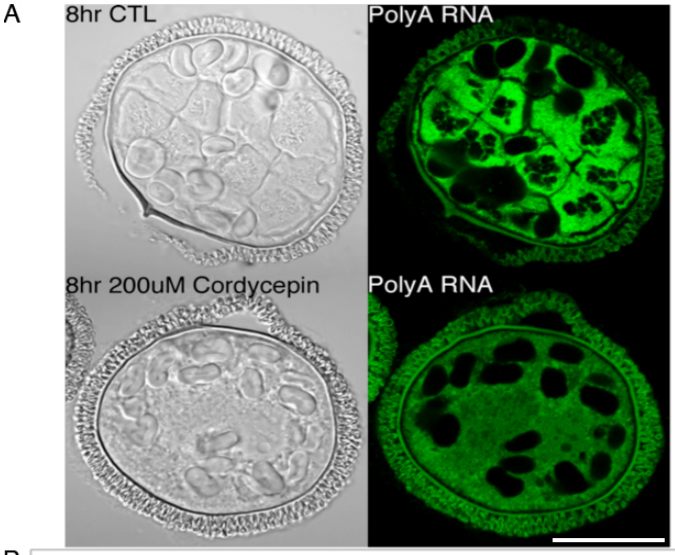
majority of microspores observed (Figure V-5). The inhibition of polyadenylation via cordycepin (3'-deoxyadenosine) treatment, exerts more severe effects on development and results in restricted divisions and reduced poly(A)+ labeling of microspores (Figure V-6). Our *in silico* analysis of *Marsilea* transcriptomes reveals that in addition to splicing factors, transcripts encoding products involved in differential polyadenylation are enriched (Figure V-7). As in several other systems (Wilt, 1973; Simon *et al*., 1992; Wu *et al*., 1998), these observations and experimental results indicate that differential cytoplasmic polyadenylation likely plays a role in modulating the translational activity of some stored transcripts during spermatogenesis in *M. vestita*. Methods do not currently exist for large-scale identification of differential cytoplasmic polyadenylation, but total RNA isolations from different time points, followed by isolation using poly(T) capture probes of differing lengths coupled with RNAseq could prove useful.



In addition to splicing and polyadenylation, the phenomenon of transcript masking is clearly central to post-transcriptional control of stored RNA in this system. Binding of RNA by Y-box proteins has been found both *in vivo* and *in vitro* to repress or 'mask' the translation of transcripts (Sommerville and Ladomery, 1996). The identification of masking agents in *M. vestita* is undoubtedly the obvious and essential next step in understanding how post-transcriptional regulation mediates processes in these cells. With our initial transcriptome sequencing complete, identifying known masking proteins and analyzing their role in development should be straightforward. The interplay between these three (unmasking, splicing, and polyadenylation) modes of regulation is likely to be complex and may differ on a transcript-by-transcript basis.

Figure V-5 *RNAi knockdown of conserved polyadenylation machinery leads to disrupted divisions.* (**A**), Immunolocalizations of gametophytes using anticytoplasmic poly-A RNA polymerase antibody (anti-PAP) during spermiogenesis. (a) Just after spore hydration, there is a weak antibody label scattered throughout the cytoplasm of the gametophyte. (b) At 2 h of development, the anti-PAP antibody label is more abundant in the spermatogenous cells (sp) of the gametophyte, and less abundant in the sterile cells (st) of the gametophyte. The open area in the spermatogenous cell on the right is occupied by a nucleus. (c) At 4 h of development, the spermatogenous cells (sp) exhibit high levels of anti-PAP antibody labeling, while the sterile hacket cells (st) exhibit almost no detectable labeling with this antibody. (d, e) At 6 h of development, the spermatogenous cells (sp) are heavily labeled (the nuclei of these cells do not label with the antibody) and the sterile cells (st) exhibit almost no staining. (f) At 8 h of development, anti-PAP antibody labeling is particularly intense in the cytoplasmic vesicle of the maturing spermatids. Scale bar = 25 μm. Tsai *et al*., 2004. (**B**), Microspores were treated with RNAi constructs targeting the 100kd subunit of CPSF (Cytoplasmic Polyadenylation Specificity Factor) at the concentrations indicated. Microspores were allowed to develop to 6 hours and then fixed, sectioned, and stained with TBO. T.C.C. Boothby, unpublished. Bar = 25 μm.

Identification of *cis*-acting regulatory elements that are likely to be involved in controlling these processes will be required for their large-scale study.



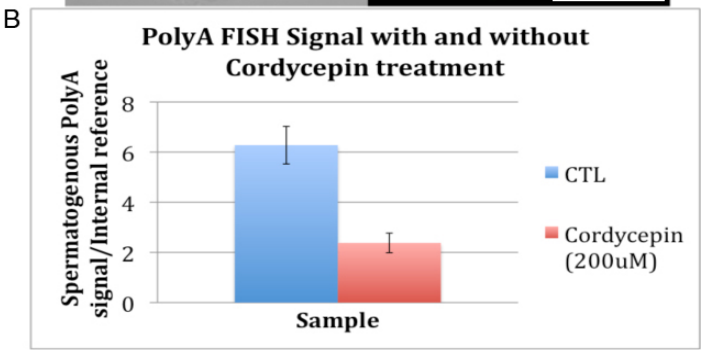


Figure V-6 (left). *Inhibition of polyadenylation results in perturbed divisions and decreased poly(T) probe labeling.* (**A**), microspores were treated with 200 μM cordycepin during rehydration. Microspores were allowed to develop for 8 h, fixed, sectioned, and labeled with biotinylated poly(T) probes. Poly(T) probes were detected using avidin conjugated Fluoroscene. Bar = 25 μm. (**B**), Intensities of poly(T) signals within spermatogenous cells were calculated using ImageJ's ROI selection tool using intene autofluorescence as an internal control. For CTL samples n = 120. For cordycepin treated samples n = 35. Error bars = 2 StDev. T.C.C. Boothby, unpublished.

G.O. term	Description	P-value
GO:0043631	RNA polyadenylation	2.61E-06
GO:0006378	mRNA polyadenylation	7.70E-06
GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	3.74E-08

Figure V-7 *Enriched GO terms related to polyadenylation of RNA*. T.C.C. Boothby, unpublished.

Marsilea as an experimental system for studying alternative splicing

In 3 of the 4 eukaryotic kingdoms, intron retention is the predominant form of alternative splicing (Campbell *et al*., 2006, McGuire *et al*., 2008). Alternative splicing is a major mechanism regulating gene expression and has been implicated in generating phenotypic differences among species (Barbosa-Morais *et al*., 2012; Merkin *et al.,* 2012), such that different species that share tissue specific transcriptional programs differ in their processing of transcriptional products (Merkin *et al*., 2012). From an experimental standpoint, the generation of diversity and complexity of RNA isoforms through

alternative splicing greatly complicates the process of RNA isoform analysis. One only has to consider the human genome, with its ~24,000 protein coding genes (Pennisi, 2003), each with an average of ~8.4 introns (Mourier and Jeffares, 2003) along with the fact that ~95% of multi-exonic human genes are predicted to undergo alternative splicing (Pan *et al*., 2008), to realize that differential modification of transcripts can greatly expand extents of gene expression. Couple this with the fact that different tissue types, organs, developmental periods, and sexes display differential processing of transcripts and the problem of understanding alternative splicing in a single (let alone multiple) model species grows even greater.

As a model system for studying intron retention and alternative splicing, the microspore of *M. vestita* offers several advantages. First and foremost, it is transcriptionally silent, so no new input of RNA should confound experimental results. Secondly, the single cell initially housed within the microspore gives rise to only two different cell types, minimizing cell-type specific isoform generation. Since the cells do not move, position and size define cell fates. Finally, development of these cells is specialized, rapid, and synchronous at 20 ^{0}C, which simplify analyses of perturbations, derived from experimental manipulations of alternative splicing.

Findings generated from the study of alternative splicing in *M. vestita* are likely to be applicable to a wide variety of eukaryotes. In this study, we have identified a large subset of conserved spliceosomal factors encoded by stored transcripts in the microspore of *M. vestita*, many of whose translational products can be labeled using antibodies made using mammalian antigens. In addition to splicing machinery, we have identified a subset of IRTs that contain conserved U2 splice signals, the splicing of which can be inhibited

by treatment with a splicing inhibitor (SSA) that was originally tested on metazoans. We have characterized the developmental dynamics of nuclear speckles during spermatogenesis in *M. vestita* and found that speckles in this species behave similarly to those in other organisms (coalescing during transcriptional inhibition, undergoing the speckle cycle, *etc.*). The apparent degree of conservation in splicing machinery and substrate specificity in *M. vestita* compared with more mainstream models leads us to suspect that mechanisms identified in this work will extend to the development of many other organisms.

Methods

Microspores harvesting and isolation

Sporocarps containing microspores were obtained from mature sporophytes grown in artificial ponds (University of Maryland, Research Greenhouse Complex). Microspores were isolated as previously described (Klink and Wolniak, 2001).

Microspore incubations

4 mg of microspores were placed in clear 2 mL tubes. A pushpin was used to make 5 small holes in the top and 3 small holes on the side of the tube. Microspores were hydrated with 1 mL of commercial spring water and agitated on an orbitron device for 1 h at 20 ^{0}C. Subsequently, hydrated microspores were transferred to 50 mL Erlenmeyer flasks and an additional 9 mL of water was added (for a total of 10 mL/flask). Flask openings were covered with aluminum foil and were placed in a shaking water bath (20 ^{0}C) for the remainder of their incubation time.

Poly(A)+ RNA isolation

After microspore incubation (see above) the contents of each flask (10 mL spring water and 4 mg microspores) was decanted into a 50 mL conical tube. Conical tubes were spun at maximum RPMs in a clinical centrifuge for ~5 minutes to pellet microspores. Microspore pellets were transferred to 2 mL tubes and spun again to pellet. As much spring water as possible was removed from the tubes with a micropipette. Microspores

were used with NEB's Magnetic mRNA Isolation Kit (#S1550S, New England Biolabs) according to the manufacturer's instructions. Isolated RNA was used immediately in downstream applications.

Preparation of cDNA for RNAseq

Poly(A)+ RNA was isolated directly (see method above), then concentrated and cleaned using the RNeasy MiniElute kit (#74204, Qiagen) according to manufacturer's instructions and eluted in a final volume of 10µL. 8µL of RNA was used for as input and library preparation conducted according to manufacturer's instructions ("Low-Throughput (LT) Protocol" TruSeq RNA Sample Preparation Guide , Illumina). Resulting cDNA library preps were assessed for quality and concentration using the Agilent 2100 Bioanalyzer (Agilent Technologies) as well as qPCR by the IBBR sequencing core at the University of Maryland, College Park.

RNAseq

100 base pair paired-end read sequencing was conducted by the IBBR sequencing core (http://www.ibbr.umd.edu/facilities/sequencing) using the HiSeq1000 (Illumina).

De novo Transcriptome assembly

For each *de novo* transcriptome assembled, the appropriate raw sequenced fragments were combined into left and right read files. These raw sequences were concatenated, filtered, and then subsequently used with the 2011-05-19 release of Trinity (Grabherr *et al*., 2011) to assemble *de novo* transcriptomes.

‘Unigene’ construction

Trinity output was parsed to retain only the longest isoform from each predicted isoform group. Longest isoforms were subsequently assembled into contigs using CAP3 with default settings (Huang and Madan, 1999). Resulting contigs and singletons were concatenated into a single file.

FPKM abundance estimations

To estimate abundances for unigenes, fragments per kilobase of exon per million fragments mapped (FPKM) values were calculated using the tuxedo suite (Trapnell *et al*., 2012). First, RNAseq reads were mapped back to our reference transcriptome using Bowtie (Langmead *et al*., 2009). Read mapping was then analyzed using Cufflinks to obtain FPKM values (Trapnell *et al*., 2010).

IGV

Mapped reads were subsequently viewed and analyzed using IGV (http://www.broadinstitute.org/igv/ Robinson *et al*., 2011; Thorvaldsdottir *et al*., 2012)

Identity assignments

Our reference or time range transcriptomes were used as queries in Stand Alone BLAST (BLASTx for protein databases and tBLASTx for nucleotide databases) searches using the ncbi-blast-2.2.25+ suite. *A. thaliana* and *H. sapiens* NCBI refseq proteins databases (downloaded on 2/12/2012) and the *P. aquilinum* gametophyte transcriptome (Der *et al*., 2011) were used. For assigning putative identities to our reference transcripts only top hits with E-values below 1E-10 were kept.

Background subtracted gene ontology over-representation analysis.

To determine over and under represented G.O. terms in our reference transcriptome we used putative gene identities (see above). These gene identities were subtracted from their corresponding database to generate "background" lists. To identify over-represented G.O. terms, our transcriptome gene identities were used as input into GOrilla (Eden *et al*., 2009) along with their corresponding background list. To identify under-represented G.O. terms the inverse was performed, with background lists being used as input and transcriptome identity using used as a subtractive background. When provided with an input and background file, GOrilla analysizes G.O. terms associated

with each list of identities and calculates over-represented G.O. terms in the input with respect to the background list. A P-value of 1E-4 was used as a threshold for detecting over-representation.

Ranked gene ontology enrichment analysis

To determine enriched G.O. terms associated with our transcriptomes, transcript abundance estimations were cross-references with transcript identity. These identities were then ranked in order of decreasing abundance. These ranked lists were used as input for GOrilla ranked G.O. enrichment analysis. A P-value of 1E-4 was used as a threshold for detecting G.O. enrichment.

IRT identification

MegaBLAST was conducted using our transcriptome both as a query and as a database with a maximum e-value of $1E^{-10}$. To identify potential IRT pairs, MegaBLAST results were parsed to select hits with a gap (intron) of at least 15 bases and gap starting position unambiguously known to within 5 bases (as some sequences make the exact starting position of a gap ambiguous). Sequences were allowed to align in any orientation that was consistent for the entirety of the alignment. Duplicate introns were removed on the basis of size, position, and sequence identity of the 20 bases surrounding the 5' and 3' splice sites.

Gene Ontology analysis of IRTs

The sequences of fully spliced isoforms from IRT pairs were used as queries in a Stand Alone BLASTx using the ncbi-blast-2.2.25+ suite against the NCBI refseq proteins database (downloaded on 2/12/2012). Our search was tailored to return the best 50 hits for each sequence in XML format. These hits were used as input into the Blast2GO (Conesa *et al*., 2005) program (http://www.blast2go.com/b2ghome). Mapping, annotation, and analysis were carried out using B2G's online database (b2g_jun11).

Genomic DNA isolation and sequencing

Leaflets from mature *Marsilea* vestita sporophytes were harvested and stored frozen at -80 ^{0}C until use. A single frozen leaflet was placed in a 1.5mL centrifuge tube and the tube immersed in liquid nitrogen. With the tube still immersed in liquid nitrogen, the leaflet was ground with a sterile plastic pestle. Grinding was continued until the leaflet was reduced to powder. 300µl of 2X CTAB (2% weight to volume cetyl-trimethyl-ammonium bromide, 1.4M NaCl, 100mM Tris-HCL pH 8.0, 20mM EDTA) was added to the tube. The contents for the tube were mixed and dissolved using the plastic pestle. The tube was capped and placed on a heat block at 65 ^{0}C for 10 minutes. After removing the tube from the heat block, 300µl of chloroform was added. The tube was vortexed for 30 seconds and then centrifuged for 5 minutes at 13,000 RPM. The translucent water phase was removed (top layer) and placed in a new clean tube. 300µl of 2-propanol was added and mixed by inversion. The sample was centrifuged again for 5 minutes at 13,000 RPM. The supernatant was removed. 500µl of 70% ethanol was added to the tube and the pellet

was loosened with gentle tapping of the tube. The tube was left to stand for 5 minutes, and then centrifuged again for 5 minutes at 13,000 RPM. The supernatant was again removed and the pellet was air dried for 10 minutes. 100µl of TE buffer (10mM Tris-HCL pH 8.0, 1mM EDTA) was added to resuspend the pellet. The tube was placed in a 37 ^{0}C heat block for 10 minutes to evaporate any remaining ethanol.

Isolated DNA was used along with gene specific primers (Integrated DNA Technologies) to perform sequencing. Sequencing was conducted by GeneWiz Inc. (www.genewiz.com) with DNA and primers being supplied according to GeneWiz's sample preparation guidelines.

Splice site sequence logos

Sequence logos were generating using WebLogo available at http://weblogo.berkeley.edu (Crooks *et al*., 2004).

Total RNA isolations

RNA isolations were conducted on 4 mg of microspores using the RNeasy Plant Mini Kit (Qiagen catalog number: 74904) according to manufacturers instructions.

Reverse Transcription Polymerase Chain Reaction

RT-PCR was conducted using AMV RT enzyme and *Taq* polymerase according to manufacturers instructions with 30 PCR cycles (New England BioLabs).

Quantification of RT-PCR product relative gel intensities

RT-PCR products were run on 2% TAE agarose gels stained with ethidium bromide. Gel images were analyzed using ImageJ software. Background subtraction was conducted using a rolling ball radius of 50 pixels. Intensities for each lane were measured using the built in gel analysis tool. Raw intensities from each time point were expressed as a ratio of the lowest intensity measured for each gene product and graphed.

Spliceostatin A and α-amanitin drug treatments

For SSA, α-amanitin, and control samples, 4mg of microspores were measured into 2mL epitubes. Spores were treated with 100ng/mL (or 200ng/mL as indicated) and 100mM of SSA (100ng/μL in methanol) and α-amanitin (1mg/mL) respectively and the total volume of each epitube increased to 1mL with spring water. Control spores were imbibed with 1mL of methanol:spring water (ratio 1:1000) or 1mL of pure spring water. SSA, α-amanitin, and controls microspores were grown on an Orbitron rotating shaker at 20 ^{0}C for the entirety of incubation. Subsequent RNA isolations or fixation was conducted as described above and below respectively.

dsRNA production for RNAi

dsRNA construct generation was performed as previously described (Klink and Wolniak, 2001). Alternatively, starting material for dsRNA construct generation were made via RT-PCR amplification of specific targets using gene specific primers with T7 promoter sequences at their extreme 5' ends. After RT-PCR dsRNA constructs were made as in Klink and Wolniak, 2001.

Fixation, embedding, and sectioning

Marsilea vestita microspores were obtained, fixed, embedded, and sectioned as previously described (Hepler, 1976; Klink and Wolniak, 2001).

DAPI staining, Immunofluorescence, and fluorescence microscopy

Microscopy was conducted as previously described (Boothby and Wolniak, 2011). Primary antibodies used in this study were as follows: rabbit anti-human Spermidine (Arcis) 1:200 in PBS, mouse anti-alpha-tubulin (Calbiochem Cat# CP06) 1:200, and mouse anti-Centrin clone 20H5 (Millipore 04-1624) 1:200. Secondary antibodies used in this study were Alexa Fluor goat anti-rabbit 594 and Alexa Fluor goat anti-mouse 594 (Molecular Probes cat# A11012 and A11005 respectively) both diluted 1:1000 in PBST. Incident light fluorescence imaging was conducted on a Zeiss Axio microscope and fluorescence confocal imaging was conducted on a Zeiss LSM 700

microspore. For all imaged material, thousands of microspores are simultaneously treated and subsequently viewed. Approximately 100 microspores were photographed for each treatment. Each treatment (SSA, α-amanitin, and RNAi) experiments were conducted at minimum in duplicate.

Widefield microscopy

All widefield fluorescence microscopy was performed with a Zeiss Axioscope equipped with standard Fluorescein, TexasRed and UV filter sets. Confocal microscopy was performed on a Zeiss LSM700 using Zen 2009 software. Subsequently, .lsm stacks were exported into ImageJ 1.44k and rendered as 3D models using the ImageJ 3D Viewer plugin.

DAPI Pyronin Y double staining

DAPI and pyronin Y are available from a number of commercial sources. DAPI was diluted in PBS and used at a final concentration of 2.5 µg/ml. Pyronin Y was used at 5% in deionized water. Pyronin Y 5%) was added 1:1 with 0.2M pH 4 acetate buffer on the day of use to make PY staining solution. Slides with sectioned material were incubated in acetone for 15 minutes on a rocker. Slides were next incubated in PBS for five minutes, this was repeated twice. Samples were placed in a humid chamber and incubated with DAPI for 10 minutes followed by incubation in PBS-T for five minutes and then PBS for five minutes on a rocker. Slides were incubated in a humid chamber for

1.5 minutes with PY staining solution, rinsed with deionized water and dehydrated in an ethanol series 1 minute incubations at 25%, 50%, 75%, 90%, 100%, and 100%). Slides were dipped briefly in xylene.

ISH and FISH probe production

Probes for 5S and 25S rRNA were constructed through digoxigenin-11-dUTP incorporation as previously described (Tsai and Wolniak, 2001). ISH probes against cytoplasmic qc centrin and SPDS transcripts were made as follows. cDNA constructs encoding both centrin and SPDS were obtained from our cDNA library (Hart and Wolniak, 1998; Hart and Wolniak, 1999). Single stranded antisense probes were transcribed using T7 RNA polymerase. Probes were biotinylated using the PHOTOPROBE biotin kit for nucleic acid labeling Vector Laboratories http://www.vectorlabs.com/). Biotinylated polyT) probe was purchased from GibcoBRL. 25mer biotinylated DNA probes for the detection of masked centrin and SPDS transcripts were obtained through the custom DNA oligo service at Integrated DNA Technologies http://www.idtdna.com).

ISH and FISH detection

The ISH protocol used here has been previously described (11). Detection of biotinylated FISH probes was achieved through the use of a goat-anti-biotin antibody, followed by a rabbit-anti-goat antibody conjugated with AlexaFluor 594. Alternatively,

biotinylated probes requiring signal amplification were detected using either TexasRed or Fluorescein labeled Avidin Vector Laboratories A-2016 and A20-11) followed by biotinylated anti-Avidin D Vector Laboratories BA-0300) labeling and a second round of TexasRed or Fluorescein labeled Avidin labeling following the manufacturers instructions.

Fluorescence *in situ* hybridization of masked transcripts

FISH on masked transcripts was performed as detailed previously (Tsai and Wolniak, 2001) with the following modifications. Hybridization was carried out for 16 hours at 30 ^{0}C. Detection was carried out using avidin/biotin signal amplification reagents Vector Laboratories) as described above. In double labeling experiments, hybridization and detection of one probe was carried out, followed by avidin/biotin blocking Vector Laboratories SP-2001), and subsequent hybridization and detection of the second probe.

Cytology and Immunocytochemistry

Toluidine Blue O staining (O'Brien and McCully, 1981) was performed on sectioned material and observed via bright-field microscopy. DAPI staining was performed as described previously (van der Weele *et al*., 2007). Immunolabeling of sectioned material was carried out was described by Baskin and Wilson (Baskin and Wilson, 1997) with modifications described by van der Weele *et al*. (van der Weele *et al*.,

2007). Antibodies used in this study were as follows: 4G3 mouse monoclonal antibody against U2B" and 72B9 mouse monoclonal antibody against Fibrillarin generously provided by Dr. Joseph Gall Carnegie Institution of Washington, Baltimore, MD) and were used at dilutions of 1:5, and 1:10 respectively. Mouse monoclonal antibody against SC35 was obtained commercially from Abcam ab11826 http://www.abcam.com) and used at a dilution of 1:200. Goat-anti-biotin, AlexaFluor 594 conjugated goat-anti-mouse and AlexaFluor 594 conjugated rabbit-anti-goat antibodies were obtained from Molecular Probes Molecular Probes, Invitrogen Detection Technologies http://www.invitrogen.com) and were used at 1:200, 1:1000, and 1:1000 dilutions respectively. Additional primary antibodies used were polyclonal anti-spermidine (1:100) (Abcam), anti-centrin monoclonal 20H5 directed against *Chlamydomonas reinhardtii* (1:100) (kind gift from Dr. J. Salisbury, Mayo Clinic, Rochester, MN) and monoclonal anti α-tubulin (1:100) (CalBiolChem, DM1A). Histochemical detection of antibodies was carried out using an alkaline phosphatase conjugated anti-mouse antibody, followed by visualization with nitro-blue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate.

Immunocytochemistry/FISH double labeling

In double labeling experiments, antibody labeling was performed as described above, coverglasses were mounted but not sealed. Following imaging, coverglasses were removed by incubating slides in PBS-T, 3 times for 5 minutes on a rocker. Following coverglass removal, our standard FISH protocol was followed as detailed above.

RNAi

RNA interference was performed as described previously Klink and Wolniak, 2001). Briefly, Microspores were placed 1 mL of spring water with 200 mg/mL double stranded RNA and allowed to develop for desired periods of time, fixed, and sectioned as described above and previously.

Polyamine treatment

10 mM Spermidine treatments were performed as previously described (Deeb *et al*., 2010).

Poly(T) signal quantification

For quantification of poly(T) signals, images of representative spores were loaded into ImageJ. Individual spermatogenous cells were selected using the free hand selection and ROI tools. For each sets of spermatogenous cells from the same microspore, the intene (inner spore wall) autofluorescence signal was used as an internal control.

RACE-PAT

RACE-PAT was performed as previously described (Salles *et al*., 1999). Briefly, poly(A)+ RNA was isolated from microspores grown for different time intervals. An anchored 1st strand primer was used for reverse transcription. RT was followed by PCR

amplification using a gene specific 2nd strand primer. Products of varying sizes were analyzed via gel electrophoresis.

Appendices

Appendix xxvi - Overrepresented 3-5 h G.O. Cellular Component terms (*A. thaliana*)

Appendix xxvii – Overrepresented 6-8 h G.O. Process terms (*A. thaliana*)

Appendix xxviii - Overrepresented 6-8 h G.O. Function terms (*A. thaliana*)

Appendix xxix - Overrepresented 6-8 h G.O. Cellular Component terms (*A. thaliana*)

Appendix i – Overrepresented G.O. Process (*A. thaliana*)

GO Term	Description	P-value
GO:0071704	organic substance metabolic process	0.00E+00
GO:0008152	metabolic process	0.00E+00
GO:0044710	single-organism metabolic process	0.00E+00
GO:0009987	cellular process	0.00E+00
GO:0044237	cellular metabolic process	0.00E+00
GO:0044238	primary metabolic process	0.00E+00
GO:0043170	macromolecule metabolic process	3.34E-265
GO:0044260	cellular macromolecule metabolic process	1.26E-248
GO:0006807	nitrogen compound metabolic process	3.77E-205
GO:0034641	cellular nitrogen compound metabolic process	5.18E-184
GO:1901360	organic cyclic compound metabolic process	1.10E-176
GO:0046483	heterocycle metabolic process	2.49E-174
GO:0006139	nucleobase-containing compound metabolic process	1.87E-167
GO:0006725	cellular aromatic compound metabolic process	7.44E-166
GO:0009058	biosynthetic process	1.32E-148
GO:0019538	protein metabolic process	3.15E-147
GO:1901576	organic substance biosynthetic process	3.37E-135
GO:0043412	macromolecule modification	4.70E-132

GO:0006793	phosphorus metabolic process	2.53E-130
GO:0044249	cellular biosynthetic process	5.63E-129
GO:0006796	phosphate-containing compound metabolic process	1.93E-126
GO:0090304	nucleic acid metabolic process	3.63E-125
GO:0071840	cellular component organization or biogenesis	1.24E-123
GO:0044267	cellular protein metabolic process	1.65E-117
GO:0016043	cellular component organization	7.28E-117
GO:0044281	small molecule metabolic process	1.25E-114
GO:0051704	multi-organism process	1.14E-107
GO:0006464	cellular protein modification process	4.03E-104
GO:0036211	protein modification process	4.03E-104
GO:0006996	organelle organization	1.31E-100
GO:1901564	organonitrogen compound metabolic process	9.11E-94
GO:0044723	single-organism carbohydrate metabolic process	7.06E-93
GO:0016070	RNA metabolic process	7.83E-87
GO:0005975	carbohydrate metabolic process	1.09E-84
GO:1901575	organic substance catabolic process	1.12E-84
GO:0009056	catabolic process	1.82E-84
GO:0019637	organophosphate metabolic process	5.43E-83
GO:0050896	response to stimulus	1.05E-75
GO:0032502	developmental process	2.54E-70
GO:0044699	single-organism process	1.80E-63
GO:0044271	cellular nitrogen compound biosynthetic process	2.43E-63

GO:0055086	nucleobase-containing small molecule metabolic process	1.35E-62
GO:1901566	organonitrogen compound biosynthetic process	2.29E-62
GO:0009059	macromolecule biosynthetic process	2.79E-62
GO:0006753	nucleoside phosphate metabolic process	3.31E-62
GO:0009117	nucleotide metabolic process	3.31E-62
GO:0019752	carboxylic acid metabolic process	2.46E-61
GO:1901135	carbohydrate derivative metabolic process	1.00E-60
GO:0006082	organic acid metabolic process	1.32E-60
GO:0043436	oxoacid metabolic process	1.83E-60
GO:0034645	cellular macromolecule biosynthetic process	3.33E-59
GO:0090407	organophosphate biosynthetic process	7.17E-59
GO:0009628	response to abiotic stimulus	8.36E-59
GO:1901362	organic cyclic compound biosynthetic process	4.22E-58
GO:0044702	single organism reproductive process	1.10E-56
GO:0051234	establishment of localization	3.11E-56
GO:0018130	heterocycle biosynthetic process	4.04E-56
GO:0044255	cellular lipid metabolic process	5.57E-56
GO:0006396	RNA processing	7.01E-56
GO:0022414	reproductive process	1.33E-55
GO:0006629	lipid metabolic process	2.32E-55
GO:0051649	establishment of localization in cell	6.92E-54
GO:0016051	carbohydrate biosynthetic process	6.21E-51
GO:0006810	transport	1.45E-50

GO:0005996	monosaccharide metabolic process	1.45E-49
GO:0065007	biological regulation	2.44E-49
GO:0019438	aromatic compound biosynthetic process	2.78E-49
GO:0044248	cellular catabolic process	2.79E-49
GO:0003006	developmental process involved in reproduction	2.71E-48
GO:0022402	cell cycle process	6.04E-48
GO:0051186	cofactor metabolic process	1.26E-47
GO:0034654	nucleobase-containing compound biosynthetic process	3.34E-46
GO:0046907	intracellular transport	4.20E-46
GO:0019318	hexose metabolic process	8.07E-46
GO:0006950	response to stress	4.16E-45
GO:0006259	DNA metabolic process	1.14E-44
GO:0006006	glucose metabolic process	2.60E-44
GO:0022607	cellular component assembly	3.57E-44
GO:0009314	response to radiation	3.07E-42
GO:0016052	carbohydrate catabolic process	3.73E-42
GO:0044724	single-organism carbohydrate catabolic process	3.73E-42
GO:0048856	anatomical structure development	2.25E-41
GO:0016310	phosphorylation	4.34E-41
GO:0050789	regulation of biological process	8.01E-41
GO:0032259	methylation	3.53E-40
GO:0005976	polysaccharide metabolic process	3.58E-40
GO:0044262	cellular carbohydrate metabolic process	4.42E-40

GO:0008610	lipid biosynthetic process	6.00E-40
GO:0006468	protein phosphorylation	6.62E-40
GO:0071822	protein complex subunit organization	1.48E-39
GO:0043414	macromolecule methylation	1.00E-38
GO:0032787	monocarboxylic acid metabolic process	2.27E-38
GO:0006461	protein complex assembly	3.18E-38
GO:0008150	biological_process	7.12E-38
GO:0055114	oxidation-reduction process	1.63E-37
GO:0051716	cellular response to stimulus	1.97E-37
GO:0034660	ncRNA metabolic process	2.38E-37
GO:0043623	cellular protein complex assembly	2.43E-37
GO:0055085	transmembrane transport	6.69E-37
GO:0044283	small molecule biosynthetic process	2.27E-36
GO:0009416	response to light stimulus	2.47E-36
GO:0044711	single-organism biosynthetic process	2.75E-36
GO:1901293	nucleoside phosphate biosynthetic process	3.79E-36
GO:0009165	nucleotide biosynthetic process	3.79E-36
GO:0044264	cellular polysaccharide metabolic process	7.33E-36
GO:0006732	coenzyme metabolic process	1.30E-35
GO:0019320	hexose catabolic process	1.90E-35
GO:0046365	monosaccharide catabolic process	1.90E-35
GO:0006007	glucose catabolic process	2.88E-35
GO:0043933	macromolecular complex subunit organization	5.04E-35

GO:0044042	glucan metabolic process	7.19E-35
GO:0000271	polysaccharide biosynthetic process	8.28E-35
GO:0051276	chromosome organization	9.38E-35
GO:0006073	cellular glucan metabolic process	1.72E-34
GO:0009790	embryo development	2.47E-34
GO:0050793	regulation of developmental process	1.42E-33
GO:0065003	macromolecular complex assembly	1.91E-33
GO:0042221	response to chemical stimulus	2.50E-33
GO:1901137	carbohydrate derivative biosynthetic process	8.04E-33
GO:0065008	regulation of biological quality	1.00E-32
GO:0007010	cytoskeleton organization	1.04E-32
GO:0034622	cellular macromolecular complex assembly	2.01E-32
GO:0009259	ribonucleotide metabolic process	2.11E-32
GO:0019693	ribose phosphate metabolic process	2.11E-32
GO:0009250	glucan biosynthetic process	7.21E-32
GO:0034637	cellular carbohydrate biosynthetic process	1.16E-31
GO:0009057	macromolecule catabolic process	2.08E-31
GO:0016568	chromatin modification	2.70E-31
GO:0033692	cellular polysaccharide biosynthetic process	4.80E-31
GO:0034470	ncRNA processing	7.93E-31
GO:0006520	cellular amino acid metabolic process	1.17E-30
GO:0032501	multicellular organismal process	4.17E-30
GO:0048518	positive regulation of biological process	2.11E-29

GO:0048646	anatomical structure formation involved in morphogenesis	3.76E-29
GO:0006886	intracellular protein transport	4.48E-29
GO:0009793	embryo development ending in seed dormancy	6.66E-29
GO:0016569	covalent chromatin modification	1.28E-28
GO:0009657	plastid organization	1.67E-28
GO:0006508	proteolysis	2.06E-28
GO:0072594	establishment of protein localization to organelle	3.57E-28
GO:0015031	protein transport	6.53E-28
GO:0045184	establishment of protein localization	6.53E-28
GO:0006631	fatty acid metabolic process	1.49E-27
GO:0009451	RNA modification	1.84E-27
GO:0016053	organic acid biosynthetic process	2.73E-27
GO:0046394	carboxylic acid biosynthetic process	2.73E-27
GO:0051239	regulation of multicellular organismal process	3.24E-27
GO:0044707	single-multicellular organism process	3.36E-27
GO:2000026	regulation of multicellular organismal development	8.70E-27
GO:0048519	negative regulation of biological process	1.66E-26
GO:0006081	cellular aldehyde metabolic process	1.89E-26
GO:0046390	ribose phosphate biosynthetic process	2.18E-26
GO:0009260	ribonucleotide biosynthetic process	2.18E-26
GO:0006644	phospholipid metabolic process	2.41E-26
GO:0051188	cofactor biosynthetic process	3.47E-26
GO:0006974	response to DNA damage stimulus	3.66E-26

GO:0072524	pyridine-containing compound metabolic process	7.36E-26
GO:0006364	rRNA processing	7.40E-26
GO:0019362	pyridine nucleotide metabolic process	2.45E-25
GO:0050794	regulation of cellular process	2.50E-25
GO:0016072	rRNA metabolic process	3.02E-25
GO:0044265	cellular macromolecule catabolic process	3.40E-25
GO:0040029	regulation of gene expression, epigenetic	3.48E-25
GO:0048522	positive regulation of cellular process	4.92E-25
GO:0016570	histone modification	5.87E-25
GO:0046496	nicotinamide nucleotide metabolic process	6.37E-25
GO:0006733	oxidoreduction coenzyme metabolic process	7.67E-25
GO:0006260	DNA replication	7.67E-25
GO:0072521	purine-containing compound metabolic process	1.85E-24
GO:0048193	Golgi vesicle transport	2.12E-24
GO:0009658	chloroplast organization	2.25E-24
GO:0016044	cellular membrane organization	2.43E-24
GO:0061024	membrane organization	2.43E-24
GO:0010027	thylakoid membrane organization	2.67E-24
GO:0009668	plastid membrane organization	2.67E-24
GO:0016192	vesicle-mediated transport	3.39E-24
GO:0071702	organic substance transport	4.19E-24
GO:0009653	anatomical structure morphogenesis	1.18E-23
GO:0006479	protein methylation	2.21E-23

GO:0008213	protein alkylation	2.21E-23
GO:0006090	pyruvate metabolic process	2.57E-23
GO:0006739	NADP metabolic process	2.86E-23
GO:0009240	isopentenyl diphosphate biosynthetic process	4.02E-23
GO:0046490	isopentenyl diphosphate metabolic process	4.02E-23
GO:0008654	phospholipid biosynthetic process	4.39E-23
GO:0048580	regulation of post-embryonic development	5.40E-23
GO:1901657	glycosyl compound metabolic process	6.07E-23
GO:0008652	cellular amino acid biosynthetic process	7.17E-23
GO:0006325	chromatin organization	8.16E-23
GO:0022403	cell cycle phase	9.25E-23
GO:0006740	NADPH regeneration	1.18E-22
GO:0006098	pentose-phosphate shunt	1.18E-22
GO:0006163	purine nucleotide metabolic process	1.49E-22
GO:0019682	glyceraldehyde-3-phosphate metabolic process	1.52E-22
GO:0019288	isopentenyl diphosphate biosynthetic process, mevalonate-independent pathway	1.52E-22
GO:0016571	histone methylation	1.99E-22
GO:0033554	cellular response to stress	2.23E-22
GO:0051128	regulation of cellular component organization	3.59E-22
GO:0042440	pigment metabolic process	8.83E-22
GO:2000241	regulation of reproductive process	9.05E-22
GO:0006281	DNA repair	1.03E-21
GO:0006605	protein targeting	2.18E-21

GO:0051603	proteolysis involved in cellular protein catabolic process	2.74E-21
GO:0016458	gene silencing	3.11E-21
GO:0009909	regulation of flower development	3.93E-21
GO:0016071	mRNA metabolic process	6.10E-21
GO:0006486	protein glycosylation	7.86E-21
GO:0043413	macromolecule glycosylation	7.86E-21
GO:0010035	response to inorganic substance	8.11E-21
GO:0010038	response to metal ion	8.47E-21
GO:0007166	cell surface receptor signaling pathway	1.11E-20
GO:0070085	glycosylation	1.70E-20
GO:0033013	tetrapyrrole metabolic process	1.95E-20
GO:0046364	monosaccharide biosynthetic process	2.32E-20
GO:0009311	oligosaccharide metabolic process	3.03E-20
GO:0031047	gene silencing by RNA	3.03E-20
GO:0016108	tetraterpenoid metabolic process	3.18E-20
GO:0016116	carotenoid metabolic process	3.18E-20
GO:1901361	organic cyclic compound catabolic process	3.64E-20
GO:0042278	purine nucleoside metabolic process	4.03E-20
GO:0006778	porphyrin-containing compound metabolic process	4.69E-20
GO:1901605	alpha-amino acid metabolic process	5.63E-20
GO:0006310	DNA recombination	6.12E-20
GO:0046128	purine ribonucleoside metabolic process	6.36E-20
GO:0010228	vegetative to reproductive phase transition of meristem	7.91E-20

GO:0048523	negative regulation of cellular process	8.97E-20
GO:0006091	generation of precursor metabolites and energy	8.98E-20
GO:0009119	ribonucleoside metabolic process	1.13E-19
GO:0016109	tetraterpenoid biosynthetic process	1.54E-19
GO:0016117	carotenoid biosynthetic process	1.54E-19
GO:0007017	microtubule-based process	1.64E-19
GO:0070647	protein modification by small protein conjugation or removal	1.71E-19
GO:0015994	chlorophyll metabolic process	2.30E-19
GO:0017038	protein import	2.34E-19
GO:0033043	regulation of organelle organization	3.33E-19
GO:0071806	protein transmembrane transport	3.50E-19
GO:0065002	intracellular protein transmembrane transport	3.50E-19
GO:0044743	intracellular protein transmembrane import	5.24E-19
GO:0009150	purine ribonucleotide metabolic process	6.58E-19
GO:0046700	heterocycle catabolic process	7.06E-19
GO:0034968	histone lysine methylation	8.93E-19
GO:0005984	disaccharide metabolic process	9.63E-19
GO:0019439	aromatic compound catabolic process	1.26E-18
GO:0009116	nucleoside metabolic process	1.94E-18
GO:0010498	proteasomal protein catabolic process	2.22E-18
GO:0044270	cellular nitrogen compound catabolic process	2.28E-18
GO:0048869	cellular developmental process	2.33E-18
GO:0051052	regulation of DNA metabolic process	2.47E-18

GO:0045814	negative regulation of gene expression, epigenetic	2.96E-18
GO:0006790	sulfur compound metabolic process	3.23E-18
GO:0051273	beta-glucan metabolic process	3.55E-18
GO:0019252	starch biosynthetic process	3.71E-18
GO:0000023	maltose metabolic process	4.55E-18
GO:0043900	regulation of multi-organism process	4.77E-18
GO:0019319	hexose biosynthetic process	5.54E-18
GO:0006342	chromatin silencing	6.86E-18
GO:0005982	starch metabolic process	6.86E-18
GO:0072528	pyrimidine-containing compound biosynthetic process	7.39E-18
GO:0010638	positive regulation of organelle organization	8.02E-18
GO:1901565	organonitrogen compound catabolic process	9.60E-18
GO:0043902	positive regulation of multi-organism process	1.13E-17
GO:0051130	positive regulation of cellular component organization	1.13E-17
GO:0006094	gluconeogenesis	1.36E-17
GO:0072527	pyrimidine-containing compound metabolic process	1.42E-17
GO:0006811	ion transport	1.42E-17
GO:0009892	negative regulation of metabolic process	1.53E-17
GO:0006635	fatty acid beta-oxidation	2.14E-17
GO:0034440	lipid oxidation	2.25E-17
GO:0016054	organic acid catabolic process	2.25E-17
GO:0046395	carboxylic acid catabolic process	2.25E-17
GO:0007062	sister chromatid cohesion	2.43E-17

GO:0046686	response to cadmium ion	3.63E-17
GO:1901659	glycosyl compound biosynthetic process	3.69E-17
GO:0044712	single-organism catabolic process	3.93E-17
GO:0044282	small molecule catabolic process	3.93E-17
GO:0009266	response to temperature stimulus	4.25E-17
GO:0009640	photomorphogenesis	4.34E-17
GO:0010605	negative regulation of macromolecule metabolic process	6.17E-17
GO:0006220	pyrimidine nucleotide metabolic process	1.05E-16
GO:0006261	DNA-dependent DNA replication	1.09E-16
GO:0051726	regulation of cell cycle	1.39E-16
GO:0072522	purine-containing compound biosynthetic process	1.47E-16
GO:0044272	sulfur compound biosynthetic process	1.61E-16
GO:0006221	pyrimidine nucleotide biosynthetic process	1.66E-16
GO:0046148	pigment biosynthetic process	1.71E-16
GO:0015672	monovalent inorganic cation transport	1.97E-16
GO:0009560	embryo sac egg cell differentiation	2.00E-16
GO:0019395	fatty acid oxidation	2.02E-16
GO:0006164	purine nucleotide biosynthetic process	2.07E-16
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	2.36E-16
GO:0007167	enzyme linked receptor protein signaling pathway	2.36E-16
GO:0030243	cellulose metabolic process	2.64E-16
GO:0043248	proteasome assembly	3.03E-16
GO:0051788	response to misfolded protein	3.03E-16

GO:0009220	pyrimidine ribonucleotide biosynthetic process	3.72E-16
GO:0009218	pyrimidine ribonucleotide metabolic process	3.72E-16
GO:0030258	lipid modification	3.80E-16
GO:0009791	post-embryonic development	4.40E-16
GO:0000226	microtubule cytoskeleton organization	6.56E-16
GO:0070646	protein modification by small protein removal	1.59E-15
GO:0010629	negative regulation of gene expression	1.70E-15
GO:0048610	cellular process involved in reproduction	2.23E-15
GO:0010090	trichome morphogenesis	2.29E-15
GO:0009639	response to red or far red light	2.32E-15
GO:0001510	RNA methylation	2.51E-15
GO:0010207	photosystem II assembly	2.72E-15
GO:0006663	fatty acid biosynthetic process	2.73E-15
GO:0042592	homeostatic process	2.80E-15
GO:0034655	nucleobase-containing compound catabolic process	2.97E-15
GO:0043094	cellular metabolic compound salvage	3.12E-15
GO:0044242	cellular lipid catabolic process	3.59E-15
GO:0046034	ATP metabolic process	3.64E-15
GO:0016042	lipid catabolic process	5.13E-15
GO:0010033	response to organic substance	5.70E-15
GO:0010564	regulation of cell cycle process	5.92E-15
GO:0009141	nucleoside triphosphate metabolic process	6.85E-15
GO:0090305	nucleic acid phosphodiester bond hydrolysis	9.57E-15

GO:0032535	regulation of cellular component size	1.05E-14
GO:0090066	regulation of anatomical structure size	1.05E-14
GO:0009062	fatty acid catabolic process	1.11E-14
GO:0072329	monocarboxylic acid catabolic process	1.36E-14
GO:0009108	coenzyme biosynthetic process	1.42E-14
GO:0000910	cytokinesis	1.44E-14
GO:0009205	purine ribonucleoside triphosphate metabolic process	1.64E-14
GO:0009199	ribonucleoside triphosphate metabolic process	1.64E-14
GO:0009144	purine nucleoside triphosphate metabolic process	1.64E-14
GO:0009605	response to external stimulus	1.84E-14
GO:0006304	DNA modification	2.00E-14
GO:0080129	proteasome core complex assembly	2.12E-14
GO:0031324	negative regulation of cellular metabolic process	2.18E-14
GO:0032774	RNA biosynthetic process	2.31E-14
GO:0009853	photorespiration	2.32E-14
GO:0042455	ribonucleoside biosynthetic process	2.76E-14
GO:0009163	nucleoside biosynthetic process	2.76E-14
GO:0044728	DNA methylation or demethylation	3.03E-14
GO:0006511	ubiquitin-dependent protein catabolic process	3.17E-14
GO:0010604	positive regulation of macromolecule metabolic process	3.21E-14
GO:1901607	alpha-amino acid biosynthetic process	3.74E-14
GO:0033014	tetrapyrrole biosynthetic process	3.76E-14
GO:0006306	DNA methylation	4.60E-14

GO:0006305	DNA alkylation	4.60E-14
GO:0043632	modification-dependent macromolecule catabolic process	4.73E-14
GO:0019941	modification-dependent protein catabolic process	4.73E-14
GO:0007059	chromosome segregation	5.78E-14
GO:0032273	positive regulation of protein polymerization	7.04E-14
GO:0031334	positive regulation of protein complex assembly	7.04E-14
GO:0043085	positive regulation of catalytic activity	7.49E-14
GO:0044093	positive regulation of molecular function	7.49E-14
GO:0040007	growth	7.69E-14
GO:0006096	glycolysis	8.90E-14
GO:0051274	beta-glucan biosynthetic process	1.02E-13
GO:0045010	actin nucleation	1.18E-13
GO:0030838	positive regulation of actin filament polymerization	1.18E-13
GO:0000902	cell morphogenesis	1.21E-13
GO:0032989	cellular component morphogenesis	1.21E-13
GO:0006302	double-strand break repair	1.34E-13
GO:0006779	porphyrin-containing compound biosynthetic process	1.36E-13
GO:0010557	positive regulation of macromolecule biosynthetic process	1.71E-13
GO:0030029	actin filament-based process	1.94E-13
GO:0042451	purine nucleoside biosynthetic process	2.08E-13
GO:0046129	purine ribonucleoside biosynthetic process	2.08E-13
GO:0051495	positive regulation of cytoskeleton organization	2.08E-13
GO:0051567	histone H3-K9 methylation	2.67E-13

GO:0051173	positive regulation of nitrogen compound metabolic process	3.04E-13
GO:0008380	RNA splicing	3.75E-13
GO:0009629	response to gravity	4.05E-13
GO:0048285	organelle fission	4.78E-13
GO:0072330	monocarboxylic acid biosynthetic process	6.10E-13
GO:0031327	negative regulation of cellular biosynthetic process	6.17E-13
GO:0016114	terpenoid biosynthetic process	6.88E-13
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	6.92E-13
GO:0051254	positive regulation of RNA metabolic process	7.03E-13
GO:0051172	negative regulation of nitrogen compound metabolic process	7.96E-13
GO:0009890	negative regulation of biosynthetic process	8.48E-13
GO:0006839	mitochondrial transport	9.25E-13
GO:0022610	biological adhesion	9.36E-13
GO:0007155	cell adhesion	9.36E-13
GO:0009902	chloroplast relocation	1.07E-12
GO:0051667	establishment of plastid localization	1.07E-12
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	1.07E-12
GO:0010558	negative regulation of macromolecule biosynthetic process	1.10E-12
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	1.10E-12
GO:0015931	nucleobase-containing compound transport	1.13E-12
GO:0030833	regulation of actin filament polymerization	1.15E-12
GO:0006066	alcohol metabolic process	1.16E-12
GO:0008299	isoprenoid biosynthetic process	1.22E-12

GO:0042793	transcription from plastid promoter	1.28E-12
GO:0015995	chlorophyll biosynthetic process	1.35E-12
GO:0006351	transcription, DNA-dependent	1.56E-12
GO:0009630	gravitropism	1.56E-12
GO:0045893	positive regulation of transcription, DNA-dependent	1.60E-12
GO:0009152	purine ribonucleotide biosynthetic process	1.87E-12
GO:0010628	positive regulation of gene expression	2.05E-12
GO:0048638	regulation of developmental growth	2.23E-12
GO:0065009	regulation of molecular function	2.30E-12
GO:0032271	regulation of protein polymerization	2.32E-12
GO:0043254	regulation of protein complex assembly	2.32E-12
GO:0044085	cellular component biogenesis	2.40E-12
GO:0006406	mRNA export from nucleus	2.41E-12
GO:0051028	mRNA transport	2.41E-12
GO:0006812	cation transport	2.54E-12
GO:0031325	positive regulation of cellular metabolic process	2.55E-12
GO:0006399	tRNA metabolic process	2.60E-12
GO:0031048	chromatin silencing by small RNA	2.70E-12
GO:1901615	organic hydroxy compound metabolic process	2.85E-12
GO:0070887	cellular response to chemical stimulus	2.93E-12
GO:0033044	regulation of chromosome organization	3.03E-12
GO:0009893	positive regulation of metabolic process	3.05E-12
GO:0009606	tropism	3.62E-12

GO:0006721	terpenoid metabolic process	3.66E-12
GO:0033205	cell cycle cytokinesis	3.93E-12
GO:0051168	nuclear export	4.25E-12
GO:0006720	isoprenoid metabolic process	4.88E-12
GO:0040008	regulation of growth	5.13E-12
GO:0009743	response to carbohydrate stimulus	5.16E-12
GO:0007165	signal transduction	5.82E-12
GO:0030832	regulation of actin filament length	5.94E-12
GO:0008064	regulation of actin polymerization or depolymerization	5.94E-12
GO:0000338	protein deneddylation	6.82E-12
GO:0010388	cullin deneddylation	6.82E-12
GO:0006346	methylation-dependent chromatin silencing	6.82E-12
GO:0051319	G2 phase	6.82E-12
GO:0000085	G2 phase of mitotic cell cycle	6.82E-12
GO:0030244	cellulose biosynthetic process	7.86E-12
GO:0034220	ion transmembrane transport	8.03E-12
GO:0045892	negative regulation of transcription, DNA-dependent	8.43E-12
GO:0051253	negative regulation of RNA metabolic process	8.43E-12
GO:0006275	regulation of DNA replication	1.09E-11
GO:0030036	actin cytoskeleton organization	1.19E-11
GO:0051656	establishment of organelle localization	1.22E-11
GO:0000911	cytokinesis by cell plate formation	1.27E-11
GO:0016049	cell growth	1.28E-11

GO:0050790	regulation of catalytic activity	1.33E-11
GO:0048229	gametophyte development	1.41E-11
GO:0010103	stomatal complex morphogenesis	1.52E-11
GO:0032446	protein modification by small protein conjugation	1.62E-11
GO:0006623	protein targeting to vacuole	1.78E-11
GO:0072666	establishment of protein localization to vacuole	1.78E-11
GO:0045132	meiotic chromosome segregation	1.91E-11
GO:0006405	RNA export from nucleus	1.95E-11
GO:0050658	RNA transport	1.95E-11
GO:0050657	nucleic acid transport	1.95E-11
GO:0051236	establishment of RNA localization	1.95E-11
GO:0006970	response to osmotic stress	2.19E-11
GO:0009607	response to biotic stimulus	2.27E-11
GO:0007034	vacuolar transport	2.30E-11
GO:0009066	aspartate family amino acid metabolic process	2.40E-11
GO:0032956	regulation of actin cytoskeleton organization	2.71E-11
GO:0032970	regulation of actin filament-based process	2.71E-11
GO:0035825	reciprocal DNA recombination	2.72E-11
GO:0007131	reciprocal meiotic recombination	4.16E-11
GO:0009891	positive regulation of biosynthetic process	4.53E-11
GO:0031328	positive regulation of cellular biosynthetic process	4.53E-11
GO:0000302	response to reactive oxygen species	4.91E-11
GO:0048645	organ formation	5.34E-11

GO:0019222	regulation of metabolic process	5.83E-11
GO:0051707	response to other organism	6.16E-11
GO:0006457	protein folding	6.33E-11
GO:0006979	response to oxidative stress	6.85E-11
GO:0051493	regulation of cytoskeleton organization	7.22E-11
GO:0007346	regulation of mitotic cell cycle	1.07E-10
GO:0000096	sulfur amino acid metabolic process	1.14E-10
GO:0009409	response to cold	1.14E-10
GO:0000394	RNA splicing, via endonucleolytic cleavage and ligation	1.18E-10
GO:0046486	glycerolipid metabolic process	1.65E-10
GO:0009651	response to salt stress	1.91E-10
GO:0009642	response to light intensity	2.37E-10
GO:0000724	double-strand break repair via homologous recombination	2.44E-10
GO:0000725	recombinational repair	2.44E-10
GO:0006650	glycerophospholipid metabolic process	2.53E-10
GO:0016567	protein ubiquitination	2.67E-10
GO:0006888	ER to Golgi vesicle-mediated transport	2.70E-10
GO:0006487	protein N-linked glycosylation	2.72E-10
GO:0007015	actin filament organization	3.06E-10
GO:0006497	protein lipidation	3.55E-10
GO:0009142	nucleoside triphosphate biosynthetic process	4.04E-10
GO:0043161	proteasomal ubiquitin-dependent protein catabolic process	4.08E-10
GO:0051301	cell division	4.15E-10

GO:0006913	nucleocytoplasmic transport	4.40E-10
GO:0051169	nuclear transport	4.40E-10
GO:0046165	alcohol biosynthetic process	4.43E-10
GO:0006606	protein import into nucleus	4.54E-10
GO:0032268	regulation of cellular protein metabolic process	4.93E-10
GO:0043687	post-translational protein modification	6.13E-10
GO:0009201	ribonucleoside triphosphate biosynthetic process	6.67E-10
GO:0009206	purine ribonucleoside triphosphate biosynthetic process	6.67E-10
GO:0009145	purine nucleoside triphosphate biosynthetic process	6.67E-10
GO:0048449	floral organ formation	6.98E-10
GO:0070192	chromosome organization involved in meiosis	7.51E-10
GO:0044087	regulation of cellular component biogenesis	7.84E-10
GO:0071310	cellular response to organic substance	8.22E-10
GO:0051170	nuclear import	8.41E-10
GO:0010608	posttranscriptional regulation of gene expression	9.70E-10
GO:0042138	meiotic DNA double-strand break formation	9.73E-10
GO:0016226	iron-sulfur cluster assembly	9.94E-10
GO:0031163	metallo-sulfur cluster assembly	9.94E-10
GO:0035966	response to topologically incorrect protein	1.00E-09
GO:0010212	response to ionizing radiation	1.02E-09
GO:0019748	secondary metabolic process	1.05E-09
GO:1901617	organic hydroxy compound biosynthetic process	1.07E-09
GO:0006754	ATP biosynthetic process	1.10E-09

GO:0007129	synapsis	1.21E-09
GO:0051246	regulation of protein metabolic process	1.35E-09
GO:0000280	nuclear division	1.54E-09
GO:0048827	phyllome development	1.81E-09
GO:0046471	phosphatidylglycerol metabolic process	1.81E-09
GO:0006626	protein targeting to mitochondrion	2.00E-09
GO:0072655	establishment of protein localization to mitochondrion	2.00E-09
GO:0006643	membrane lipid metabolic process	2.03E-09
GO:0072511	divalent inorganic cation transport	2.04E-09
GO:0071407	cellular response to organic cyclic compound	2.31E-09
GO:0045595	regulation of cell differentiation	2.36E-09
GO:0009411	response to UV	2.64E-09
GO:0070838	divalent metal ion transport	2.81E-09
GO:0006665	phosphatidylglycerol biosynthetic process	2.99E-09
GO:0006401	RNA catabolic process	3.04E-09
GO:0071322	cellular response to carbohydrate stimulus	3.04E-09
GO:0006200	ATP catabolic process	3.08E-09
GO:0043574	peroxisomal transport	3.34E-09
GO:0048878	chemical homeostasis	3.80E-09
GO:0043543	protein acylation	4.71E-09
GO:0006625	protein targeting to peroxisome	5.09E-09
GO:0072663	establishment of protein localization to peroxisome	5.09E-09
GO:0030154	cell differentiation	6.02E-09

GO:0016556	mRNA modification	6.09E-09
GO:0009110	vitamin biosynthetic process	6.63E-09
GO:0030422	production of siRNA involved in RNA interference	6.86E-09
GO:0009644	response to high light intensity	7.27E-09
GO:0006402	mRNA catabolic process	7.88E-09
GO:0043269	regulation of ion transport	8.07E-09
GO:0051604	protein maturation	9.37E-09
GO:0016558	protein import into peroxisome matrix	1.17E-08
GO:0050801	ion homeostasis	1.31E-08
GO:0006928	cellular component movement	1.49E-08
GO:0009067	aspartate family amino acid biosynthetic process	1.49E-08
GO:0006152	purine nucleoside catabolic process	1.57E-08
GO:0072523	purine-containing compound catabolic process	1.57E-08
GO:0000097	sulfur amino acid biosynthetic process	1.61E-08
GO:0043331	response to dsRNA	1.68E-08
GO:0071359	cellular response to dsRNA	1.68E-08
GO:0031050	dsRNA fragmentation	1.68E-08
GO:0070918	production of small RNA involved in gene silencing by RNA	1.68E-08
GO:0010182	sugar mediated signaling pathway	1.74E-08
GO:0009756	carbohydrate mediated signaling	1.74E-08
GO:0007389	pattern specification process	1.75E-08
GO:0009203	ribonucleoside triphosphate catabolic process	1.78E-08
GO:0009207	purine ribonucleoside triphosphate catabolic process	1.78E-08

GO:0009146	purine nucleoside triphosphate catabolic process	1.78E-08
GO:0009143	nucleoside triphosphate catabolic process	1.78E-08
GO:0009965	leaf morphogenesis	1.91E-08
GO:0032200	telomere organization	2.04E-08
GO:0000723	telomere maintenance	2.04E-08
GO:0060249	anatomical structure homeostasis	2.04E-08
GO:0007030	Golgi organization	2.14E-08
GO:0048609	multicellular organismal reproductive process	2.14E-08
GO:0006270	DNA replication initiation	2.22E-08
GO:0009308	amine metabolic process	2.33E-08
GO:0046130	purine ribonucleoside catabolic process	2.36E-08
GO:0048767	root hair elongation	2.38E-08
GO:0000956	nuclear-transcribed mRNA catabolic process	2.59E-08
GO:0007020	microtubule nucleation	2.80E-08
GO:0035196	production of miRNAs involved in gene silencing by miRNA	3.18E-08
GO:0009073	aromatic amino acid family biosynthetic process	3.54E-08
GO:0055080	cation homeostasis	3.85E-08
GO:0010817	regulation of hormone levels	4.04E-08
GO:0046459	short-chain fatty acid metabolic process	4.71E-08
GO:0007033	vacuole organization	5.17E-08
GO:0006195	purine nucleotide catabolic process	5.30E-08
GO:0009261	ribonucleotide catabolic process	5.30E-08
GO:0009154	purine ribonucleotide catabolic process	5.30E-08

GO:0006816	calcium ion transport	5.43E-08
GO:0006766	vitamin metabolic process	5.66E-08
GO:0048366	leaf development	6.73E-08
GO:0046474	glycerophospholipid biosynthetic process	6.93E-08
GO:0009072	aromatic amino acid family metabolic process	7.02E-08
GO:0016125	sterol metabolic process	7.10E-08
GO:0042542	response to hydrogen peroxide	7.28E-08
GO:0032879	regulation of localization	8.12E-08
GO:0006665	sphingolipid metabolic process	8.21E-08
GO:0010332	response to gamma radiation	8.25E-08
GO:0006498	N-terminal protein lipidation	8.29E-08
GO:0006499	N-terminal protein myristoylation	8.29E-08
GO:0018377	protein myristoylation	8.29E-08
GO:0009887	organ morphogenesis	8.91E-08
GO:0009309	amine biosynthetic process	9.03E-08
GO:0008202	steroid metabolic process	9.14E-08
GO:0000272	polysaccharide catabolic process	9.45E-08
GO:0031365	N-terminal protein amino acid modification	9.76E-08
GO:0009845	seed germination	9.99E-08
GO:0009637	response to blue light	1.01E-07
GO:0045017	glycerolipid biosynthetic process	1.10E-07
GO:0009886	post-embryonic morphogenesis	1.16E-07
GO:0043247	telomere maintenance in response to DNA damage	1.17E-07

GO:0009069	serine family amino acid metabolic process	1.18E-07
GO:0015748	organophosphate ester transport	1.34E-07
GO:1901292	nucleoside phosphate catabolic process	1.48E-07
GO:0009166	nucleotide catabolic process	1.48E-07
GO:0016579	protein deubiquitination	1.54E-07
GO:0051049	regulation of transport	1.55E-07
GO:0031399	regulation of protein modification process	1.60E-07
GO:0010383	cell wall polysaccharide metabolic process	1.68E-07
GO:0009880	embryonic pattern specification	1.74E-07
GO:0016126	sterol biosynthetic process	1.79E-07
GO:2000602	regulation of interphase of mitotic cell cycle	1.87E-07
GO:0009555	pollen development	1.91E-07
GO:0006312	mitotic recombination	2.06E-07
GO:0006555	methionine metabolic process	2.16E-07
GO:0042454	ribonucleoside catabolic process	2.16E-07
GO:0009164	nucleoside catabolic process	2.16E-07
GO:0048583	regulation of response to stimulus	2.18E-07
GO:0010075	regulation of meristem growth	2.49E-07
GO:0015711	organic anion transport	2.55E-07
GO:0016441	posttranscriptional gene silencing	2.69E-07
GO:0006694	steroid biosynthetic process	2.74E-07
GO:0044036	cell wall macromolecule metabolic process	2.86E-07
GO:0006767	water-soluble vitamin metabolic process	2.99E-07

GO:0042364	water-soluble vitamin biosynthetic process	2.99E-07
GO:0030001	metal ion transport	3.18E-07
GO:0032504	multicellular organism reproduction	3.21E-07
GO:0032204	regulation of telomere maintenance	3.25E-07
GO:0010155	regulation of proton transport	3.31E-07
GO:0010014	meristem initiation	3.59E-07
GO:0015979	photosynthesis	3.66E-07
GO:0042546	cell wall biogenesis	3.66E-07
GO:0022603	regulation of anatomical structure morphogenesis	3.67E-07
GO:0072593	reactive oxygen species metabolic process	3.86E-07
GO:0006862	nucleotide transport	4.07E-07
GO:0009553	embryo sac development	4.07E-07
GO:0051187	cofactor catabolic process	4.07E-07
GO:0035194	posttranscriptional gene silencing by RNA	4.54E-07
GO:0022900	electron transport chain	4.57E-07
GO:0010389	regulation of G2/M transition of mitotic cell cycle	4.78E-07
GO:0019220	regulation of phosphate metabolic process	4.88E-07
GO:0051174	regulation of phosphorus metabolic process	4.88E-07
GO:0051179	localization	5.01E-07
GO:0006397	mRNA processing	5.30E-07
GO:0001558	regulation of cell growth	5.35E-07
GO:0071705	nitrogen compound transport	5.71E-07
GO:0046434	organophosphate catabolic process	6.51E-07

GO:0048481	ovule development	6.61E-07
GO:0034285	response to disaccharide stimulus	6.87E-07
GO:0046467	membrane lipid biosynthetic process	6.95E-07
GO:0030163	protein catabolic process	7.57E-07
GO:0042743	hydrogen peroxide metabolic process	7.62E-07
GO:0006787	porphyrin-containing compound catabolic process	7.63E-07
GO:0015996	chlorophyll catabolic process	7.63E-07
GO:0033015	tetrapyrrole catabolic process	7.63E-07
GO:0046149	pigment catabolic process	7.63E-07
GO:0009617	response to bacterium	7.86E-07
GO:0022613	ribonucleoprotein complex biogenesis	9.11E-07
GO:0048509	regulation of meristem development	9.15E-07
GO:0009744	response to sucrose stimulus	9.15E-07
GO:0010267	production of ta-siRNAs involved in RNA interference	9.57E-07
GO:0002376	immune system process	1.01E-06
GO:0007126	meiosis	1.01E-06
GO:0019684	photosynthesis, light reaction	1.18E-06
GO:0042254	ribosome biogenesis	1.19E-06
GO:0032844	regulation of homeostatic process	1.24E-06
GO:0015849	organic acid transport	1.24E-06
GO:0046942	carboxylic acid transport	1.24E-06
GO:0006955	immune response	1.28E-06
GO:0051790	short-chain fatty acid biosynthetic process	1.38E-06

GO:0009086	methionine biosynthetic process	1.38E-06
GO:1901658	glycosyl compound catabolic process	1.38E-06
GO:0016144	S-glycoside biosynthetic process	1.45E-06
GO:0019758	glycosinolate biosynthetic process	1.45E-06
GO:0019761	glucosinolate biosynthetic process	1.45E-06
GO:1901136	carbohydrate derivative catabolic process	1.47E-06
GO:0009933	meristem structural organization	1.48E-06
GO:0009825	multidimensional cell growth	1.49E-06
GO:0045087	innate immune response	1.65E-06
GO:0048589	developmental growth	1.65E-06
GO:0006820	anion transport	1.66E-06
GO:0014070	response to organic cyclic compound	1.67E-06
GO:0009408	response to heat	1.71E-06
GO:0019725	cellular homeostasis	1.75E-06
GO:0000904	cell morphogenesis involved in differentiation	1.76E-06
GO:0006366	transcription from RNA polymerase II promoter	1.93E-06
GO:0006637	acyl-CoA metabolic process	1.96E-06
GO:0043478	pigment accumulation in response to UV light	1.96E-06
GO:0043479	pigment accumulation in tissues in response to UV light	1.96E-06
GO:0043476	pigment accumulation	1.96E-06
GO:0043481	anthocyanin accumulation in tissues in response to UV light	1.96E-06
GO:0043480	pigment accumulation in tissues	1.96E-06
GO:0043473	pigmentation	1.96E-06

GO:0035383	thioester metabolic process	1.96E-06
GO:0010073	meristem maintenance	1.98E-06
GO:0051094	positive regulation of developmental process	1.98E-06
GO:0048585	negative regulation of response to stimulus	2.10E-06
GO:0051510	regulation of unidimensional cell growth	2.41E-06
GO:0007018	microtubule-based movement	2.71E-06
GO:0016143	S-glycoside metabolic process	2.82E-06
GO:0019760	glucosinolate metabolic process	2.82E-06
GO:0019757	glycosinolate metabolic process	2.82E-06
GO:0046185	aldehyde catabolic process	2.85E-06
GO:0009832	plant-type cell wall biogenesis	3.02E-06
GO:0000067	mitosis	3.06E-06
GO:0006084	acetyl-CoA metabolic process	3.56E-06
GO:0042742	defense response to bacterium	3.58E-06
GO:0006865	amino acid transport	3.90E-06
GO:0051596	methylglyoxal catabolic process	4.08E-06
GO:0009438	methylglyoxal metabolic process	4.08E-06
GO:0035303	regulation of dephosphorylation	4.25E-06
GO:0071554	cell wall organization or biogenesis	4.81E-06
GO:0006412	translation	4.96E-06
GO:0006944	cellular membrane fusion	4.96E-06
GO:0061025	membrane fusion	4.96E-06
GO:0018193	peptidyl-amino acid modification	5.79E-06

GO:0022604	regulation of cell morphogenesis	5.79E-06
GO:0006089	lactate metabolic process	5.83E-06
GO:0019243	methylglyoxal catabolic process to D-lactate	5.83E-06
GO:0046777	protein autophosphorylation	5.92E-06
GO:0048532	anatomical structure arrangement	6.47E-06
GO:0009063	cellular amino acid catabolic process	6.62E-06
GO:0043648	dicarboxylic acid metabolic process	7.00E-06
GO:0030148	sphingolipid biosynthetic process	7.17E-06
GO:0009914	hormone transport	7.42E-06
GO:0009767	photosynthetic electron transport chain	7.51E-06
GO:0006813	potassium ion transport	7.63E-06
GO:0048582	positive regulation of post-embryonic development	7.63E-06
GO:0051607	defense response to virus	7.66E-06
GO:0016246	RNA interference	8.22E-06
GO:0048825	cotyledon development	8.22E-06
GO:0051093	negative regulation of developmental process	8.59E-06
GO:0048437	floral organ development	8.59E-06
GO:0006873	cellular ion homeostasis	8.68E-06
GO:0008283	cell proliferation	8.75E-06
GO:0007275	multicellular organismal development	1.03E-05
GO:0060918	auxin transport	1.06E-05
GO:0008033	tRNA processing	1.07E-05
GO:0035304	regulation of protein dephosphorylation	1.08E-05

GO:0010118	stomatal movement	1.15E-05
GO:0030003	cellular cation homeostasis	1.15E-05
GO:0090421	embryonic meristem initiation	1.20E-05
GO:0009648	photoperiodism	1.55E-05
GO:0031323	regulation of cellular metabolic process	1.64E-05
GO:0009064	glutamine family amino acid metabolic process	1.64E-05
GO:0050826	response to freezing	1.67E-05
GO:0060560	developmental growth involved in morphogenesis	1.76E-05
GO:0042023	DNA endoreduplication	1.82E-05
GO:0042180	cellular ketone metabolic process	1.84E-05
GO:0042181	ketone biosynthetic process	1.84E-05
GO:0009773	photosynthetic electron transport in photosystem I	1.88E-05
GO:0009615	response to virus	2.02E-05
GO:0009926	auxin polar transport	2.15E-05
GO:0048513	organ development	2.23E-05
GO:0051322	anaphase	2.73E-05
GO:0008219	cell death	2.90E-05
GO:0016265	death	2.90E-05
GO:0010431	seed maturation	2.93E-05
GO:0045426	quinone cofactor biosynthetic process	2.93E-05
GO:1901661	quinone metabolic process	2.93E-05
GO:1901663	quinone biosynthetic process	2.93E-05
GO:0042375	quinone cofactor metabolic process	2.93E-05

GO:0015696	ammonium transport	3.18E-05
GO:0035725	sodium ion transmembrane transport	3.18E-05
GO:0055082	cellular chemical homeostasis	3.47E-05
GO:1901606	alpha-amino acid catabolic process	3.52E-05
GO:0043038	amino acid activation	4.05E-05
GO:0043039	tRNA aminoacylation	4.05E-05
GO:0060255	regulation of macromolecule metabolic process	4.14E-05
GO:0044275	cellular carbohydrate catabolic process	4.23E-05
GO:0010119	regulation of stomatal movement	4.26E-05
GO:0052249	modulation of RNA levels in other organism involved in symbiotic interaction	4.66E-05
GO:0052018	modulation by symbiont of RNA levels in host	4.66E-05
GO:0009616	virus induced gene silencing	4.66E-05
GO:0006505	GPI anchor metabolic process	4.71E-05
GO:0022611	dormancy process	4.78E-05
GO:0019321	pentose metabolic process	4.86E-05
GO:0048443	stamen development	5.05E-05
GO:0015802	basic amino acid transport	5.32E-05
GO:0051817	modification of morphology or physiology of other organism involved in symbiotic interaction	5.53E-05
GO:0044003	modification by symbiont of host morphology or physiology	5.53E-05
GO:0035821	modification of morphology or physiology of other organism	5.53E-05
GO:0048608	reproductive structure development	5.53E-05
GO:0033036	macromolecule localization	5.76E-05
GO:0080090	regulation of primary metabolic process	5.94E-05

GO:0006418	tRNA aminoacylation for protein translation	5.98E-05
GO:0048573	photoperiodism, flowering	6.22E-05
GO:0016926	protein desumoylation	6.45E-05
GO:0009070	serine family amino acid biosynthetic process	6.76E-05
GO:0010410	hemicellulose metabolic process	6.76E-05
GO:0006897	endocytosis	6.86E-05
GO:0010072	primary shoot apical meristem specification	7.16E-05
GO:0010876	lipid localization	7.60E-05
GO:0019915	lipid storage	7.60E-05
GO:0009888	tissue development	7.61E-05
GO:0044550	secondary metabolite biosynthetic process	7.96E-05
GO:0045036	protein targeting to chloroplast	8.20E-05
GO:0072596	establishment of protein localization to chloroplast	8.20E-05
GO:0010048	vernalization response	8.26E-05
GO:0048364	root development	8.42E-05
GO:0002252	immune effector process	8.44E-05
GO:0043603	cellular amide metabolic process	8.62E-05
GO:0006534	cysteine metabolic process	8.86E-05
GO:0006636	unsaturated fatty acid biosynthetic process	1.02E-04
GO:0033559	unsaturated fatty acid metabolic process	1.02E-04
GO:0010218	response to far red light	1.04E-04
GO:0010114	response to red light	1.10E-04
GO:0010162	seed dormancy process	1.10E-04

GO:0006814	sodium ion transport	1.11E-04
GO:0016036	cellular response to phosphate starvation	1.12E-04
GO:0016444	somatic cell DNA recombination	1.16E-04
GO:0009911	positive regulation of flower development	1.16E-04
GO:0046519	sphingoid metabolic process	1.16E-04
GO:0006417	regulation of translation	1.18E-04
GO:0050665	hydrogen peroxide biosynthetic process	1.18E-04
GO:0010039	response to iron ion	1.30E-04
GO:0071495	cellular response to endogenous stimulus	1.37E-04
GO:0032870	cellular response to hormone stimulus	1.37E-04
GO:0019344	cysteine biosynthetic process	1.39E-04
GO:0034614	cellular response to reactive oxygen species	1.42E-04
GO:0007186	G-protein coupled receptor signaling pathway	1.43E-04
GO:0021700	developmental maturation	1.46E-04
GO:0006664	glycolipid metabolic process	1.51E-04
GO:0051701	interaction with host	1.51E-04
GO:0044419	interspecies interaction between organisms	1.52E-04
GO:0051302	regulation of cell division	1.60E-04
GO:0046520	sphingoid biosynthetic process	1.83E-04
GO:0070592	cell wall polysaccharide biosynthetic process	1.88E-04
GO:0070589	cellular component macromolecule biosynthetic process	1.88E-04
GO:0044038	cell wall macromolecule biosynthetic process	1.88E-04
GO:0009407	toxin catabolic process	1.89E-04

GO:0009404	toxin metabolic process	1.89E-04
GO:0006354	DNA-dependent transcription, elongation	1.89E-04
GO:0006566	threonine metabolic process	2.48E-04
GO:0008356	asymmetric cell division	2.48E-04
GO:0042559	pteridine-containing compound biosynthetic process	2.48E-04
GO:0031347	regulation of defense response	2.53E-04
GO:0009966	regulation of signal transduction	2.60E-04
GO:0007051	spindle organization	2.64E-04
GO:0034599	cellular response to oxidative stress	2.70E-04
GO:0019751	polyol metabolic process	2.70E-04
GO:0048569	post-embryonic organ development	2.87E-04
GO:0009060	aerobic respiration	2.88E-04
GO:0010351	lithium ion transport	2.88E-04
GO:0010304	PSII associated light-harvesting complex II catabolic process	2.88E-04
GO:0006338	chromatin remodeling	2.88E-04
GO:0044257	cellular protein catabolic process	2.88E-04
GO:0009123	nucleoside monophosphate metabolic process	2.88E-04
GO:0010393	galacturonan metabolic process	2.89E-04
GO:0008361	regulation of cell size	2.89E-04
GO:0045488	pectin metabolic process	2.89E-04
GO:0045229	external encapsulating structure organization	3.00E-04
GO:0048284	organelle fusion	3.12E-04
GO:0046488	phosphatidylinositol metabolic process	3.28E-04

GO:0009749	response to glucose stimulus	3.43E-04
GO:1901068	guanosine-containing compound metabolic process	3.46E-04
GO:0009846	pollen germination	3.48E-04
GO:0009799	specification of symmetry	3.56E-04
GO:0031348	negative regulation of defense response	3.88E-04
GO:0045491	xylan metabolic process	4.02E-04
GO:0071241	cellular response to inorganic substance	4.05E-04
GO:0072525	pyridine-containing compound biosynthetic process	4.07E-04
GO:0031408	oxylipin biosynthetic process	4.07E-04
GO:0000279	M phase	4.31E-04
GO:0010315	auxin efflux	4.31E-04
GO:0000087	M phase of mitotic cell cycle	4.31E-04
GO:0023051	regulation of signaling	4.35E-04
GO:0048469	cell maturation	4.45E-04
GO:0048765	root hair cell differentiation	4.45E-04
GO:0048764	trichoblast maturation	4.45E-04
GO:0000741	karyogamy	4.48E-04
GO:0007031	peroxisome organization	4.62E-04
GO:0009124	nucleoside monophosphate biosynthetic process	4.62E-04
GO:0009161	ribonucleoside monophosphate metabolic process	4.62E-04
GO:0009750	response to fructose stimulus	4.67E-04
GO:0009855	determination of bilateral symmetry	4.79E-04
GO:0051641	cellular localization	4.79E-04

GO:0016129	phytosteroid biosynthetic process	4.82E-04
GO:0006818	hydrogen transport	4.91E-04
GO:0015992	proton transport	4.91E-04
GO:0009624	response to nematode	4.91E-04
GO:0080134	regulation of response to stress	4.91E-04
GO:0019216	regulation of lipid metabolic process	5.23E-04
GO:0071446	cellular response to salicylic acid stimulus	5.27E-04
GO:0003002	regionalization	5.30E-04
GO:0010646	regulation of cell communication	5.49E-04
GO:0048588	developmental cell growth	5.84E-04
GO:0052546	cell wall pectin metabolic process	6.19E-04
GO:0051225	spindle assembly	6.19E-04
GO:0009863	salicylic acid mediated signaling pathway	6.30E-04
GO:0002237	response to molecule of bacterial origin	6.34E-04
GO:0042335	cuticle development	6.35E-04
GO:0070925	organelle assembly	6.35E-04
GO:0016128	phytosteroid metabolic process	6.40E-04
GO:0045492	xylan biosynthetic process	6.42E-04
GO:0016132	brassinosteroid biosynthetic process	6.44E-04
GO:0006400	tRNA modification	6.74E-04
GO:0043604	amide biosynthetic process	6.74E-04
GO:0010540	basipetal auxin transport	6.74E-04
GO:0071555	cell wall organization	7.35E-04

GO:0009156	ribonucleoside monophosphate biosynthetic process	7.38E-04
GO:0009084	glutamine family amino acid biosynthetic process	7.38E-04
GO:0006833	water transport	7.39E-04
GO:0042044	fluid transport	7.39E-04
GO:0006506	GPI anchor biosynthetic process	7.51E-04
GO:0006525	arginine metabolic process	7.51E-04
GO:0006367	transcription initiation from RNA polymerase II promoter	7.51E-04
GO:0019400	alditol metabolic process	7.51E-04
GO:0007140	male meiosis	7.51E-04
GO:0051026	chiasma assembly	7.51E-04
GO:0009396	folic acid-containing compound biosynthetic process	7.51E-04
GO:0019852	L-ascorbic acid metabolic process	7.51E-04
GO:0019853	L-ascorbic acid biosynthetic process	7.51E-04
GO:0051258	protein polymerization	7.83E-04
GO:0043090	amino acid import	8.08E-04
GO:0010413	glucuronoxylan metabolic process	8.09E-04
GO:0016131	brassinosteroid metabolic process	8.49E-04
GO:0048439	flower morphogenesis	9.10E-04
GO:0009755	hormone-mediated signaling pathway	9.37E-04
GO:0015988	energy coupled proton transmembrane transport, against electrochemical gradient	9.41E-04
GO:0015991	ATP hydrolysis coupled proton transport	9.41E-04
GO:0009225	nucleotide-sugar metabolic process	9.41E-04
GO:0030048	actin filament-based movement	9.77E-04

GO:0019827	stem cell maintenance	9.77E-04
GO:0034284	response to monosaccharide stimulus	9.95E-04

Appendix ii – Overrepresented G.O. Functions (*A. thaliana*)

GO Term	Description	P-value
GO:0003824	catalytic activity	9.26E-265
GO:0036094	small molecule binding	4.05E-133
GO:1901265	nucleoside phosphate binding	5.97E-131
GO:0000166	nucleotide binding	5.97E-131
GO:0005488	binding	1.45E-129
GO:0043168	anion binding	6.08E-126
GO:0005524	ATP binding	1.99E-121
GO:0032559	adenyl ribonucleotide binding	5.21E-116
GO:0030554	adenyl nucleotide binding	6.46E-116
GO:0035639	purine ribonucleoside triphosphate binding	1.82E-115
GO:0032553	ribonucleotide binding	3.38E-112
GO:0001882	nucleoside binding	4.66E-112
GO:0043167	ion binding	5.35E-112
GO:0032549	ribonucleoside binding	6.98E-112
GO:0017076	purine nucleotide binding	2.67E-111
GO:0032555	purine ribonucleotide binding	5.26E-111
GO:0032550	purine ribonucleoside binding	7.88E-111
GO:0001883	purine nucleoside binding	7.88E-111
GO:0016740	transferase activity	1.29E-103
GO:0097159	organic cyclic compound binding	1.96E-86
GO:1901363	heterocyclic compound binding	4.57E-

3		86
GO:0016787	hydrolase activity	2.38E-70
GO:0016772	transferase activity, transferring phosphorus-containing groups	3.25E-64
GO:0016817	hydrolase activity, acting on acid anhydrides	3.78E-63
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	1.66E-62
GO:0016462	pyrophosphatase activity	1.47E-61
GO:0017111	nucleoside-triphosphatase activity	2.33E-61
GO:0016887	ATPase activity	1.84E-52
GO:0016301	kinase activity	1.28E-49
GO:0016773	phosphotransferase activity, alcohol group as acceptor	1.20E-47
GO:0042623	ATPase activity, coupled	1.13E-42
GO:0004672	protein kinase activity	1.77E-36
GO:0005215	transporter activity	3.51E-35
GO:0004674	protein serine/threonine kinase activity	7.27E-34
GO:0005515	protein binding	1.29E-31
GO:0043492	ATPase activity, coupled to movement of substances	1.10E-26
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	1.10E-26
GO:0022857	transmembrane transporter activity	4.87E-26
GO:0022892	substrate-specific transporter activity	2.74E-25
GO:0016820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	4.55E-25
GO:0022804	active transmembrane transporter activity	8.43E-25
GO:0015075	ion transmembrane transporter activity	9.83E-25
GO:0048037	cofactor binding	4.83E-24

GO:0015399	primary active transmembrane transporter activity	7.16E-24
GO:0015405	P-P-bond-hydrolysis-driven transmembrane transporter activity	1.14E-23
GO:0022891	substrate-specific transmembrane transporter activity	3.79E-22
GO:0043169	cation binding	2.23E-21
GO:0046872	metal ion binding	2.06E-20
GO:0004386	helicase activity	3.26E-19
GO:0050662	coenzyme binding	7.60E-19
GO:0016874	ligase activity	9.13E-17
GO:0008324	cation transmembrane transporter activity	1.48E-16
GO:0016491	oxidoreductase activity	4.41E-16
GO:0003723	RNA binding	1.02E-15
GO:0016853	isomerase activity	4.07E-15
GO:0016757	transferase activity, transferring glycosyl groups	4.13E-14
GO:0008026	ATP-dependent helicase activity	7.49E-14
GO:0070035	purine NTP-dependent helicase activity	7.49E-14
GO:0042625	ATPase activity, coupled to transmembrane movement of ions	7.61E-13
GO:0008233	peptidase activity	9.54E-13
GO:0046914	transition metal ion binding	1.19E-12
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	2.24E-12
GO:0022890	inorganic cation transmembrane transporter activity	2.50E-12
GO:0070011	peptidase activity, acting on L-amino acid peptides	2.95E-12
GO:0016779	nucleotidyltransferase activity	4.62E-12
GO:0008270	zinc ion binding	1.08E-11

GO:0008237	metallopeptidase activity	3.24E-11
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	4.01E-11
GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	4.50E-11
GO:0008509	anion transmembrane transporter activity	9.71E-11
GO:0016746	transferase activity, transferring acyl groups	2.22E-10
GO:0016829	lyase activity	2.38E-10
GO:0004175	endopeptidase activity	5.08E-10
GO:0046873	metal ion transmembrane transporter activity	6.71E-10
GO:0016758	transferase activity, transferring hexosyl groups	7.63E-10
GO:0016879	ligase activity, forming carbon-nitrogen bonds	7.89E-10
GO:0051540	metal cluster binding	8.83E-10
GO:0051536	iron-sulfur cluster binding	8.83E-10
GO:0003674	molecular_function	1.23E-09
GO:0016741	transferase activity, transferring one-carbon groups	1.76E-09
GO:0004222	metalloendopeptidase activity	7.29E-09
GO:0015077	monovalent inorganic cation transmembrane transporter activity	7.60E-09
GO:0008168	methyltransferase activity	9.91E-09
GO:0004713	protein tyrosine kinase activity	1.01E-08
GO:0008094	DNA-dependent ATPase activity	1.07E-08
GO:0042578	phosphoric ester hydrolase activity	1.11E-08
GO:0008514	organic anion transmembrane transporter activity	1.81E-08
GO:0016788	hydrolase activity, acting on ester bonds	2.20E-08
GO:0016747	transferase activity, transferring acyl groups other than amino-acyl groups	6.93E-08

GO:0016407	acetyltransferase activity	7.26E-08
GO:0050660	flavin adenine dinucleotide binding	9.03E-08
GO:0000287	magnesium ion binding	9.16E-08
GO:0016860	intramolecular oxidoreductase activity	1.03E-07
GO:0004518	nuclease activity	1.19E-07
GO:0019829	cation-transporting ATPase activity	1.81E-07
GO:0019899	enzyme binding	2.08E-07
GO:0016881	acid-amino acid ligase activity	3.77E-07
GO:0016830	carbon-carbon lyase activity	5.30E-07
GO:0046943	carboxylic acid transmembrane transporter activity	8.39E-07
GO:0005342	organic acid transmembrane transporter activity	1.01E-06
GO:0004872	receptor activity	1.20E-06
GO:0034061	DNA polymerase activity	1.69E-06
GO:0003676	nucleic acid binding	1.77E-06
GO:0038023	signaling receptor activity	1.98E-06
GO:0016791	phosphatase activity	2.57E-06
GO:0008173	RNA methyltransferase activity	2.95E-06
GO:0051287	NAD binding	2.96E-06
GO:0030170	pyridoxal phosphate binding	4.08E-06
GO:0015291	secondary active transmembrane transporter activity	4.20E-06
GO:0042802	identical protein binding	4.99E-06
GO:0005509	calcium ion binding	8.25E-06
GO:0003887	DNA-directed DNA polymerase activity	8.93E-06

GO:0050661	NADP binding	9.51E-06
GO:0015079	potassium ion transmembrane transporter activity	1.20E-05
GO:0015171	amino acid transmembrane transporter activity	1.23E-05
GO:0003684	damaged DNA binding	1.60E-05
GO:0003777	microtubule motor activity	1.64E-05
GO:0008080	N-acetyltransferase activity	1.88E-05
GO:0019787	small conjugating protein ligase activity	1.93E-05
GO:0008135	translation factor activity, nucleic acid binding	2.03E-05
GO:0008194	UDP-glycosyltransferase activity	2.61E-05
GO:0015081	sodium ion transmembrane transporter activity	3.18E-05
GO:0003678	DNA helicase activity	3.18E-05
GO:0051537	2 iron, 2 sulfur cluster binding	3.18E-05
GO:0016410	N-acyltransferase activity	3.78E-05
GO:0004812	aminoacyl-tRNA ligase activity	4.05E-05
GO:0016875	ligase activity, forming carbon-oxygen bonds	4.05E-05
GO:0016876	ligase activity, forming aminoacyl-tRNA and related compounds	4.05E-05
GO:0003735	structural constituent of ribosome	4.31E-05
GO:0004888	transmembrane signaling receptor activity	4.65E-05
GO:0004842	ubiquitin-protein ligase activity	4.73E-05
GO:0003774	motor activity	4.86E-05
GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	5.27E-05
GO:0004712	protein serine/threonine/tyrosine kinase activity	5.72E-05
GO:0003924	GTPase activity	6.38E-05

GO:0005198	structural molecule activity	6.56E-05
GO:0005507	copper ion binding	7.05E-05
GO:0016866	intramolecular transferase activity	8.62E-05
GO:0016831	carboxy-lyase activity	1.03E-04
GO:0019200	carbohydrate kinase activity	1.11E-04
GO:0051082	unfolded protein binding	1.12E-04
GO:0004003	ATP-dependent DNA helicase activity	1.43E-04
GO:0015078	hydrogen ion transmembrane transporter activity	1.50E-04
GO:0016903	oxidoreductase activity, acting on the aldehyde or oxo group of donors	1.89E-04
GO:0016645	oxidoreductase activity, acting on the CH-NH group of donors	2.46E-04
GO:0010329	auxin efflux transmembrane transporter activity	2.48E-04
GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	2.64E-04
GO:0005319	lipid transporter activity	2.64E-04
GO:0016836	hydro-lyase activity	2.66E-04
GO:0015562	efflux transmembrane transporter activity	2.88E-04
GO:0015085	calcium ion transmembrane transporter activity	2.88E-04
GO:0004519	endonuclease activity	3.05E-04
GO:0031072	heat shock protein binding	3.47E-04
GO:0016859	cis-trans isomerase activity	3.76E-04
GO:0005388	calcium-transporting ATPase activity	4.31E-04
GO:0005337	nucleoside transmembrane transporter activity	4.31E-04
GO:0004702	receptor signaling protein serine/threonine kinase activity	4.50E-04
GO:0005057	receptor signaling protein activity	4.50E-04

GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	5.86E-04
GO:0015293	symporter activity	6.17E-04
GO:0008170	N-methyltransferase activity	6.35E-04
GO:0004721	phosphoprotein phosphatase activity	6.42E-04
GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	6.44E-04
GO:0004252	serine-type endopeptidase activity	6.44E-04
GO:0003724	RNA helicase activity	6.74E-04
GO:0072509	divalent inorganic cation transmembrane transporter activity	6.94E-04
GO:0004683	calmodulin-dependent protein kinase activity	7.01E-04
GO:0003682	chromatin binding	7.01E-04
GO:0043566	structure-specific DNA binding	7.78E-04
GO:0015294	solute:cation symporter activity	8.09E-04
GO:0008081	phosphoric diester hydrolase activity	8.30E-04
GO:0004497	monooxygenase activity	8.61E-04
GO:0015103	inorganic anion transmembrane transporter activity	9.10E-04

Appendix iii – Overrepresented G.O. Cellular Component terms (*A. thaliana*)

GO Term	Description	P-value
GO:0044446	intracellular organelle part	2.08E-159
GO:0044422	organelle part	1.39E-158
GO:0044464	cell part	1.37E-149
GO:0016020	membrane	2.27E-116
GO:0009536	plastid	1.17E-106
GO:0044444	cytoplasmic part	1.92E-104
GO:0009507	chloroplast	1.10E-103
GO:0044434	chloroplast part	4.11E-95
GO:0044435	plastid part	7.17E-95
GO:0044424	intracellular part	6.52E-81
GO:0005829	cytosol	1.35E-78
GO:0005886	plasma membrane	3.51E-60
GO:0032991	macromolecular complex	1.96E-59
GO:0043234	protein complex	8.94E-59
GO:0009570	chloroplast stroma	1.19E-55
GO:0009532	plastid stroma	1.19E-55
GO:0031975	envelope	2.44E-47
GO:0031967	organelle envelope	2.44E-47
GO:0005737	cytoplasm	3.36E-46
GO:0009526	plastid envelope	3.09E-41
GO:0009941	chloroplast envelope	2.50E-40
GO:0044428	nuclear part	7.07E-35
GO:0031090	organelle membrane	1.07E-33
GO:0005794	Golgi apparatus	2.57E-29
GO:0044425	membrane part	3.87E-28
GO:0043226	organelle	1.26E-27
GO:0043229	intracellular organelle	1.71E-27
GO:0043227	membrane-bounded organelle	5.25E-25
GO:0043231	intracellular membrane-bounded organelle	6.98E-25
GO:0043228	non-membrane-bounded organelle	3.85E-24
GO:0043232	intracellular non-membrane-bounded organelle	3.85E-24
GO:0009579	thylakoid	8.20E-24
GO:0044436	thylakoid part	2.20E-22
GO:0016021	integral to membrane	1.49E-20
GO:0055035	plastid thylakoid membrane	5.36E-20
GO:0009535	chloroplast thylakoid membrane	7.92E-20
GO:0034357	photosynthetic membrane	1.22E-19

GO:0042651	thylakoid membrane	1.22E-19
GO:0005730	nucleolus	1.20E-18
GO:0005774	vacuolar membrane	4.47E-18
GO:0009506	plasmodesma	8.76E-18
GO:0030054	cell junction	1.38E-17
GO:0005911	cell-cell junction	1.42E-17
GO:0031984	organelle subcompartment	1.73E-17
GO:0044437	vacuolar part	4.54E-17
GO:0009534	chloroplast thylakoid	6.07E-17
GO:0031976	plastid thylakoid	6.07E-17
GO:0019866	organelle inner membrane	3.21E-16
GO:0000151	ubiquitin ligase complex	6.48E-15
GO:0005768	endosome	1.48E-14
GO:0044429	mitochondrial part	3.38E-13
GO:0005575	cellular_component	4.86E-13
GO:0042579	microbody	6.31E-13
GO:0031461	cullin-RING ubiquitin ligase complex	1.13E-12
GO:0005802	trans-Golgi network	1.36E-12
GO:0005783	endoplasmic reticulum	1.91E-12
GO:0080008	Cul4-RING ubiquitin ligase complex	3.19E-12
GO:0030312	external encapsulating structure	6.15E-12
GO:0005618	cell wall	6.15E-12
GO:0005777	peroxisome	1.02E-11
GO:0031224	intrinsic to membrane	2.17E-11
GO:0005773	vacuole	2.06E-10
GO:0044432	endoplasmic reticulum part	4.23E-10
GO:0005743	mitochondrial inner membrane	2.49E-09
GO:0030529	ribonucleoprotein complex	9.49E-09
GO:0031966	mitochondrial membrane	3.18E-08
GO:0009528	plastid inner membrane	3.96E-08
GO:0022626	cytosolic ribosome	6.80E-08
GO:0005840	ribosome	1.09E-07
GO:0009706	chloroplast inner membrane	1.09E-07
GO:0044451	nucleoplasm part	3.21E-07
GO:0042170	plastid membrane	4.57E-07
GO:0019898	extrinsic to membrane	5.69E-07
GO:0031969	chloroplast membrane	6.15E-07
GO:0031300	intrinsic to organelle membrane	7.57E-07
GO:0044431	Golgi apparatus part	1.20E-06
GO:0044430	cytoskeletal part	1.72E-06
GO:0005635	nuclear envelope	1.72E-06

GO:0044445	cytosolic part	2.39E-06
GO:0009505	plant-type cell wall	4.84E-06
GO:0005622	intracellular	5.85E-06
GO:0031234	extrinsic to internal side of plasma membrane	7.00E-06
GO:0044455	mitochondrial membrane part	8.59E-06
GO:0005643	nuclear pore	8.93E-06
GO:0046930	pore complex	8.93E-06
GO:0044459	plasma membrane part	9.75E-06
GO:0005834	heterotrimeric G-protein complex	1.07E-05
GO:0005789	endoplasmic reticulum membrane	1.64E-05
GO:0000325	plant-type vacuole	1.88E-05
GO:0019897	extrinsic to plasma membrane	2.16E-05
GO:0010319	stromule	2.60E-05
GO:0005694	chromosome	2.68E-05
GO:0010287	plastoglobule	2.73E-05
GO:0000502	proteasome complex	3.78E-05
GO:0031977	thylakoid lumen	4.66E-05
GO:0009543	chloroplast thylakoid lumen	1.02E-04
GO:0031978	plastid thylakoid lumen	1.02E-04
GO:0044454	nuclear chromosome part	1.16E-04
GO:0044427	chromosomal part	1.39E-04
GO:0000139	Golgi membrane	1.48E-04
GO:0044433	cytoplasmic vesicle part	1.52E-04
GO:0030120	vesicle coat	1.72E-04
GO:0005819	spindle	1.79E-04
GO:0031301	integral to organelle membrane	3.46E-04
GO:0030117	membrane coat	3.99E-04
GO:0031974	membrane-enclosed lumen	3.99E-04
GO:0044391	ribosomal subunit	5.27E-04
GO:0005874	microtubule	6.40E-04
GO:0009524	phragmoplast	7.38E-04

Appendix iv – Underrepresented G.O. Function terms (*A. thaliana*)

GO Term	Description	P-value
GO:0001071	nucleic acid binding transcription factor activity	3.86E-15
GO:0003700	sequence-specific DNA binding transcription factor activity	3.86E-15
GO:0046910	pectinesterase inhibitor activity	9.48E-06
GO:0004857	enzyme inhibitor activity	5.18E-05
GO:0046983	protein dimerization activity	2.44E-04
GO:0008083	growth factor activity	2.54E-04
GO:0005102	receptor binding	3.25E-04
GO:0043565	sequence-specific DNA binding	3.93E-04
GO:0045735	nutrient reservoir activity	5.28E-04

Appendix v – Underrepresented G.O. Cellular Component terms (*A. thaliana*)

GO Term	Description	P-value
GO:0005576	extracellular region	3.08E-14
GO:0005739	mitochondrion	1.94E-08
GO:0005634	nucleus	8.77E-06

Appendix vi – Overrepresented G.O. Process terms (*H. sapiens*)

GO Term	Description	P-value
GO:0008152	metabolic process	1.28E-182
GO:0044710	single-organism metabolic process	1.65E-181
GO:0071704	organic substance metabolic process	4.40E-177
GO:0044237	cellular metabolic process	1.80E-176
GO:0044238	primary metabolic process	2.46E-167
GO:0044267	cellular protein metabolic process	2.52E-99
GO:1901575	organic substance catabolic process	3.49E-89
GO:0016071	mRNA metabolic process	8.87E-88
GO:0044260	cellular macromolecule metabolic process	2.42E-87
GO:0009056	catabolic process	8.24E-86
GO:0019538	protein metabolic process	3.67E-83
GO:0006396	RNA processing	5.08E-81
GO:0043170	macromolecule metabolic process	1.78E-80
GO:0044248	cellular catabolic process	1.52E-78
GO:0044281	small molecule metabolic process	8.92E-78
GO:0010467	gene expression	7.23E-76
GO:0006807	nitrogen compound metabolic process	2.69E-71
GO:1901360	organic cyclic compound metabolic process	5.29E-70
GO:0034641	cellular nitrogen compound metabolic process	1.42E-69
GO:0044265	cellular macromolecule catabolic process	7.34E-68
GO:000905	macromolecule catabolic process	3.71E

7		-63
GO:0046483	heterocycle metabolic process	6.28E-63
GO:0006725	cellular aromatic compound metabolic process	4.92E-60
GO:0043412	macromolecule modification	1.20E-56
GO:0006139	nucleobase-containing compound metabolic process	3.61E-56
GO:0009987	cellular process	1.96E-55
GO:0000956	nuclear-transcribed mRNA catabolic process	2.30E-51
GO:0006402	mRNA catabolic process	3.13E-50
GO:0006401	RNA catabolic process	7.39E-50
GO:0006464	cellular protein modification process	3.32E-49
GO:0036211	protein modification process	3.32E-49
GO:0034660	ncRNA metabolic process	1.51E-48
GO:0046907	intracellular transport	1.14E-46
GO:1901361	organic cyclic compound catabolic process	1.15E-46
GO:0044270	cellular nitrogen compound catabolic process	1.52E-45
GO:0046700	heterocycle catabolic process	2.91E-45
GO:0019439	aromatic compound catabolic process	1.35E-44
GO:0034655	nucleobase-containing compound catabolic process	8.04E-44
GO:0006397	mRNA processing	3.06E-43
GO:0006793	phosphorus metabolic process	3.14E-42
GO:0006413	translational initiation	3.41E-41
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	8.71E-41
GO:0019752	carboxylic acid metabolic process	2.78E-40

GO:0006259	DNA metabolic process	7.78E-40
GO:0019637	organophosphate metabolic process	9.38E-39
GO:1901564	organonitrogen compound metabolic process	1.09E-38
GO:0006412	translation	1.91E-38
GO:0006796	phosphate-containing compound metabolic process	2.68E-38
GO:0034470	ncRNA processing	1.37E-36
GO:0016032	viral reproduction	2.13E-36
GO:0006082	organic acid metabolic process	3.91E-36
GO:0043436	oxoacid metabolic process	6.52E-36
GO:0008380	RNA splicing	8.88E-36
GO:0090304	nucleic acid metabolic process	3.57E-35
GO:0006613	cotranslational protein targeting to membrane	3.93E-35
GO:0045047	protein targeting to ER	3.93E-35
GO:0072599	establishment of protein localization to endoplasmic reticulum	3.93E-35
GO:0071702	organic substance transport	9.05E-35
GO:0006614	SRP-dependent cotranslational protein targeting to membrane	9.88E-35
GO:0006974	response to DNA damage stimulus	3.75E-34
GO:0006281	DNA repair	1.15E-33
GO:0000375	RNA splicing, via transesterification reactions	1.28E-33
GO:0006415	translational termination	2.33E-33
GO:0072594	establishment of protein localization to organelle	3.90E-33
GO:0043241	protein complex disassembly	1.07E-32
GO:0006414	translational elongation	1.90E-32

GO:0008150	biological_process	2.65E-32
GO:0043624	cellular protein complex disassembly	2.71E-32
GO:0032984	macromolecular complex disassembly	2.83E-32
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	3.10E-32
GO:0000398	mRNA splicing, via spliceosome	3.10E-32
GO:0019058	viral infectious cycle	6.83E-32
GO:0022402	cell cycle process	1.12E-31
GO:0009058	biosynthetic process	2.30E-31
GO:0019083	viral transcription	3.99E-31
GO:0015031	protein transport	1.50E-30
GO:0033554	cellular response to stress	2.21E-29
GO:0006886	intracellular protein transport	2.22E-29
GO:0045184	establishment of protein localization	2.76E-29
GO:0006612	protein targeting to membrane	1.74E-28
GO:0055086	nucleobase-containing small molecule metabolic process	2.61E-28
GO:0022403	cell cycle phase	3.56E-28
GO:1901576	organic substance biosynthetic process	3.75E-28
GO:0006605	protein targeting	1.16E-27
GO:0070647	protein modification by small protein conjugation or removal	1.67E-27
GO:0090407	organophosphate biosynthetic process	2.45E-27
GO:0051649	establishment of localization in cell	2.89E-27
GO:0007049	cell cycle	9.88E-27
GO:0000278	mitotic cell cycle	3.93E-26

GO:0000084	S phase of mitotic cell cycle	4.52E-26
GO:0051186	cofactor metabolic process	8.93E-26
GO:0016072	rRNA metabolic process	1.04E-25
GO:0006810	transport	1.04E-25
GO:0005975	carbohydrate metabolic process	2.05E-25
GO:0051320	S phase	2.27E-25
GO:0006753	nucleoside phosphate metabolic process	3.96E-25
GO:0006364	rRNA processing	5.68E-25
GO:0009117	nucleotide metabolic process	6.44E-25
GO:0051234	establishment of localization	6.81E-25
GO:0051603	proteolysis involved in cellular protein catabolic process	7.56E-25
GO:0043632	modification-dependent macromolecule catabolic process	8.69E-25
GO:0006457	protein folding	1.21E-24
GO:0032446	protein modification by small protein conjugation	1.75E-24
GO:0044249	cellular biosynthetic process	2.76E-24
GO:0019941	modification-dependent protein catabolic process	6.12E-24
GO:0006399	tRNA metabolic process	6.18E-24
GO:0006511	ubiquitin-dependent protein catabolic process	8.33E-24
GO:0006520	cellular amino acid metabolic process	1.65E-23
GO:0044283	small molecule biosynthetic process	1.10E-22
GO:0016567	protein ubiquitination	2.09E-22
GO:0022415	viral reproductive process	3.01E-22
GO:0008610	lipid biosynthetic process	3.71E-22

GO:0044255	cellular lipid metabolic process	4.35E-22
GO:0044711	single-organism biosynthetic process	8.38E-22
GO:0006732	coenzyme metabolic process	1.22E-21
GO:0000075	cell cycle checkpoint	1.97E-21
GO:0006629	lipid metabolic process	7.85E-21
GO:0071840	cellular component organization or biogenesis	2.65E-20
GO:0043933	macromolecular complex subunit organization	3.63E-20
GO:0044723	single-organism carbohydrate metabolic process	5.41E-20
GO:0071156	regulation of cell cycle arrest	1.30E-19
GO:1901657	glycosyl compound metabolic process	1.66E-19
GO:1901135	carbohydrate derivative metabolic process	4.75E-19
GO:0006996	organelle organization	6.30E-19
GO:0000209	protein polyubiquitination	9.11E-19
GO:0010498	proteasomal protein catabolic process	1.06E-18
GO:0043161	proteasomal ubiquitin-dependent protein catabolic process	1.06E-18
GO:0016053	organic acid biosynthetic process	1.08E-18
GO:0046394	carboxylic acid biosynthetic process	1.08E-18
GO:0055114	oxidation-reduction process	1.80E-18
GO:0051704	multi-organism process	4.33E-18
GO:0044712	single-organism catabolic process	6.06E-18
GO:0044282	small molecule catabolic process	6.06E-18
GO:0016043	cellular component organization	6.20E-18
GO:0006508	proteolysis	6.34E-18

GO:0000216	M/G1 transition of mitotic cell cycle	9.81E-18
GO:0015931	nucleobase-containing compound transport	1.42E-17
GO:1901566	organonitrogen compound biosynthetic process	3.03E-17
GO:0006310	DNA recombination	4.53E-17
GO:0051028	mRNA transport	4.68E-17
GO:0009116	nucleoside metabolic process	5.13E-17
GO:0000082	G1/S transition of mitotic cell cycle	6.11E-17
GO:0016070	RNA metabolic process	7.88E-17
GO:0050658	RNA transport	9.37E-17
GO:0050657	nucleic acid transport	9.37E-17
GO:0051236	establishment of RNA localization	9.37E-17
GO:0031145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	1.12E-16
GO:0051439	regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	1.12E-16
GO:0022411	cellular component disassembly	1.43E-16
GO:1901605	alpha-amino acid metabolic process	2.22E-16
GO:0016054	organic acid catabolic process	2.36E-16
GO:0046395	carboxylic acid catabolic process	2.36E-16
GO:0018193	peptidyl-amino acid modification	2.38E-16
GO:0055085	transmembrane transport	2.81E-16
GO:0051443	positive regulation of ubiquitin-protein ligase activity	4.62E-16
GO:0051276	chromosome organization	5.55E-16
GO:0022613	ribonucleoprotein complex biogenesis	6.30E-16
GO:1901565	organonitrogen compound catabolic process	7.47E-16

GO:0051438	regulation of ubiquitin-protein ligase activity	7.57E-16
GO:0031123	RNA 3'-end processing	8.11E-16
GO:0051437	positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	1.48E-15
GO:0051188	cofactor biosynthetic process	1.52E-15
GO:0044085	cellular component biogenesis	1.77E-15
GO:0071826	ribonucleoprotein complex subunit organization	2.20E-15
GO:0009119	ribonucleoside metabolic process	2.79E-15
GO:0009165	nucleotide biosynthetic process	2.80E-15
GO:1901293	nucleoside phosphate biosynthetic process	3.17E-15
GO:0006260	DNA replication	4.00E-15
GO:0044703	multi-organism reproductive process	4.61E-15
GO:0006354	DNA-dependent transcription, elongation	1.09E-14
GO:0071822	protein complex subunit organization	1.35E-14
GO:0022618	ribonucleoprotein complex assembly	1.37E-14
GO:0051436	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	1.67E-14
GO:0072521	purine-containing compound metabolic process	1.82E-14
GO:1901137	carbohydrate derivative biosynthetic process	1.83E-14
GO:0046128	purine ribonucleoside metabolic process	1.95E-14
GO:0042278	purine nucleoside metabolic process	1.97E-14
GO:0051351	positive regulation of ligase activity	2.15E-14
GO:0051340	regulation of ligase activity	2.28E-14
GO:0051168	nuclear export	2.94E-14
GO:0006289	nucleotide-excision repair	3.02E-14

GO:0008033	tRNA processing	3.02E-14
GO:0009063	cellular amino acid catabolic process	3.18E-14
GO:0032787	monocarboxylic acid metabolic process	3.19E-14
GO:0006283	transcription-coupled nucleotide-excision repair	3.93E-14
GO:0006487	protein N-linked glycosylation	9.67E-14
GO:0006353	DNA-dependent transcription, termination	1.13E-13
GO:0006405	RNA export from nucleus	1.14E-13
GO:0046034	ATP metabolic process	1.17E-13
GO:0006406	mRNA export from nucleus	1.26E-13
GO:0019693	ribose phosphate metabolic process	1.34E-13
GO:0018196	peptidyl-asparagine modification	1.34E-13
GO:0018279	protein N-linked glycosylation via asparagine	1.34E-13
GO:0031397	negative regulation of protein ubiquitination	1.98E-13
GO:0010564	regulation of cell cycle process	2.02E-13
GO:0006368	transcription elongation from RNA polymerase II promoter	2.59E-13
GO:0008652	cellular amino acid biosynthetic process	3.01E-13
GO:0031396	regulation of protein ubiquitination	3.21E-13
GO:0009259	ribonucleotide metabolic process	3.49E-13
GO:0051352	negative regulation of ligase activity	3.93E-13
GO:0051444	negative regulation of ubiquitin-protein ligase activity	3.93E-13
GO:0006163	purine nucleotide metabolic process	5.33E-13
GO:0009141	nucleoside triphosphate metabolic process	5.80E-13
GO:0000288	nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay	6.53E-13

GO:0043687	post-translational protein modification	6.72E-13
GO:0006418	tRNA aminoacylation for protein translation	9.66E-13
GO:0043038	amino acid activation	1.33E-12
GO:0043039	tRNA aminoacylation	1.33E-12
GO:0022616	DNA strand elongation	1.33E-12
GO:0031124	mRNA 3'-end processing	1.52E-12
GO:0051169	nuclear transport	2.24E-12
GO:0009199	ribonucleoside triphosphate metabolic process	2.47E-12
GO:0048610	cellular process involved in reproduction	2.60E-12
GO:0031398	positive regulation of protein ubiquitination	2.85E-12
GO:0046364	monosaccharide biosynthetic process	3.67E-12
GO:0009144	purine nucleoside triphosphate metabolic process	4.09E-12
GO:0051301	cell division	4.14E-12
GO:0009108	coenzyme biosynthetic process	4.42E-12
GO:1901659	glycosyl compound biosynthetic process	5.30E-12
GO:0009163	nucleoside biosynthetic process	5.30E-12
GO:0009150	purine ribonucleotide metabolic process	5.91E-12
GO:0006913	nucleocytoplasmic transport	7.09E-12
GO:0005996	monosaccharide metabolic process	7.54E-12
GO:0006271	DNA strand elongation involved in DNA replication	8.40E-12
GO:0009205	purine ribonucleoside triphosphate metabolic process	1.10E-11
GO:0016052	carbohydrate catabolic process	1.11E-11
GO:0042455	ribonucleoside biosynthetic process	1.11E-11

GO:0032200	telomere organization	1.76E-11
GO:0006644	phospholipid metabolic process	1.85E-11
GO:0036260	RNA capping	1.94E-11
GO:0009452	7-methylguanosine RNA capping	1.94E-11
GO:0044724	single-organism carbohydrate catabolic process	2.15E-11
GO:0000723	telomere maintenance	3.97E-11
GO:0019318	hexose metabolic process	4.83E-11
GO:0006370	7-methylguanosine mRNA capping	5.02E-11
GO:0006488	dolichol-linked oligosaccharide biosynthetic process	5.27E-11
GO:0019884	antigen processing and presentation of exogenous antigen	6.38E-11
GO:0048193	Golgi vesicle transport	7.91E-11
GO:0006006	glucose metabolic process	7.91E-11
GO:1901606	alpha-amino acid catabolic process	9.66E-11
GO:0006312	mitotic recombination	1.32E-10
GO:0071103	DNA conformation change	1.45E-10
GO:0006458	'de novo' protein folding	1.63E-10
GO:0008654	phospholipid biosynthetic process	1.84E-10
GO:0002478	antigen processing and presentation of exogenous peptide antigen	1.94E-10
GO:0046365	monosaccharide catabolic process	1.95E-10
GO:0046390	ribose phosphate biosynthetic process	2.42E-10
GO:1901607	alpha-amino acid biosynthetic process	2.76E-10
GO:0016569	covalent chromatin modification	3.22E-10
GO:0016570	histone modification	5.26E-10

GO:0016051	carbohydrate biosynthetic process	5.46E-10
GO:0009451	RNA modification	5.98E-10
GO:0006091	generation of precursor metabolites and energy	7.19E-10
GO:0034622	cellular macromolecular complex assembly	9.34E-10
GO:0042451	purine nucleoside biosynthetic process	1.19E-09
GO:0046129	purine ribonucleoside biosynthetic process	1.19E-09
GO:0048002	antigen processing and presentation of peptide antigen	1.26E-09
GO:0009260	ribonucleotide biosynthetic process	1.31E-09
GO:0006007	glucose catabolic process	1.40E-09
GO:0006200	ATP catabolic process	1.40E-09
GO:0051726	regulation of cell cycle	1.42E-09
GO:0019320	hexose catabolic process	1.57E-09
GO:0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	1.84E-09
GO:0072431	signal transduction involved in mitotic cell cycle G1/S transition DNA damage checkpoint	1.84E-09
GO:0072413	signal transduction involved in mitotic cell cycle checkpoint	1.84E-09
GO:0072474	signal transduction involved in mitotic cell cycle G1/S checkpoint	1.84E-09
GO:0072404	signal transduction involved in G1/S transition checkpoint	1.84E-09
GO:0000280	nuclear division	2.02E-09
GO:0007067	mitosis	2.02E-09
GO:0032392	DNA geometric change	2.89E-09
GO:0006733	oxidoreduction coenzyme metabolic process	2.93E-09
GO:0072330	monocarboxylic acid biosynthetic process	2.94E-09
GO:0071158	positive regulation of cell cycle arrest	3.11E-09

GO:0006302	double-strand break repair	3.65E-09
GO:0051084	'de novo' posttranslational protein folding	3.74E-09
GO:0048285	organelle fission	3.81E-09
GO:0030330	DNA damage response, signal transduction by p53 class mediator	4.33E-09
GO:0016568	chromatin modification	4.63E-09
GO:0072395	signal transduction involved in cell cycle checkpoint	4.83E-09
GO:0072422	signal transduction involved in DNA damage checkpoint	4.83E-09
GO:0072401	signal transduction involved in DNA integrity checkpoint	4.83E-09
GO:0072522	purine-containing compound biosynthetic process	4.95E-09
GO:0006631	fatty acid metabolic process	5.04E-09
GO:0045017	glycerolipid biosynthetic process	5.43E-09
GO:0019319	hexose biosynthetic process	6.07E-09
GO:0032508	DNA duplex unwinding	6.96E-09
GO:0010833	telomere maintenance via telomere lengthening	6.96E-09
GO:0006650	glycerophospholipid metabolic process	9.69E-09
GO:0006369	termination of RNA polymerase II transcription	1.07E-08
GO:0006766	vitamin metabolic process	1.12E-08
GO:0000722	telomere maintenance via recombination	1.26E-08
GO:0046486	glycerolipid metabolic process	1.28E-08
GO:0006096	glycolysis	1.75E-08
GO:0072331	signal transduction by p53 class mediator	2.04E-08
GO:0030258	lipid modification	2.28E-08
GO:0018205	peptidyl-lysine modification	2.53E-08

GO:0046434	organophosphate catabolic process	2.86E-08
GO:0006094	gluconeogenesis	3.02E-08
GO:0006081	cellular aldehyde metabolic process	3.02E-08
GO:0008334	histone mRNA metabolic process	3.12E-08
GO:0006066	alcohol metabolic process	3.16E-08
GO:0032259	methylation	3.36E-08
GO:0046474	glycerophospholipid biosynthetic process	3.52E-08
GO:0000718	nucleotide-excision repair, DNA damage removal	3.88E-08
GO:0044349	DNA excision	3.88E-08
GO:0051701	interaction with host	4.12E-08
GO:0006261	DNA-dependent DNA replication	5.11E-08
GO:0072524	pyridine-containing compound metabolic process	5.33E-08
GO:0009152	purine ribonucleotide biosynthetic process	5.99E-08
GO:0006164	purine nucleotide biosynthetic process	6.88E-08
GO:0032201	telomere maintenance via semi-conservative replication	7.70E-08
GO:0071705	nitrogen compound transport	7.96E-08
GO:0065003	macromolecular complex assembly	8.00E-08
GO:0035966	response to topologically incorrect protein	8.85E-08
GO:0019048	virus-host interaction	9.07E-08
GO:0070085	glycosylation	9.74E-08
GO:0006486	protein glycosylation	9.93E-08
GO:0043413	macromolecule glycosylation	9.93E-08
GO:0009083	branched-chain amino acid catabolic process	1.00E-07

GO:0042770	signal transduction in response to DNA damage	1.06E-07
GO:0006720	isoprenoid metabolic process	1.43E-07
GO:0019882	antigen processing and presentation	1.49E-07
GO:0043414	macromolecule methylation	1.50E-07
GO:0019362	pyridine nucleotide metabolic process	1.64E-07
GO:0046496	nicotinamide nucleotide metabolic process	1.64E-07
GO:0006084	acetyl-CoA metabolic process	1.69E-07
GO:0042787	protein ubiquitination involved in ubiquitin-dependent protein catabolic process	1.86E-07
GO:0007017	microtubule-based process	1.92E-07
GO:0006767	water-soluble vitamin metabolic process	2.25E-07
GO:0000291	nuclear-transcribed mRNA catabolic process, exonucleolytic	2.26E-07
GO:0043928	exonucleolytic nuclear-transcribed mRNA catabolic process involved in deadenylation-dependent decay	2.26E-07
GO:0018410	C-terminal protein amino acid modification	2.26E-07
GO:0009143	nucleoside triphosphate catabolic process	2.48E-07
GO:0050434	positive regulation of viral transcription	2.84E-07
GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	3.01E-07
GO:0006497	protein lipidation	3.35E-07
GO:0044419	interspecies interaction between organisms	3.53E-07
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	3.63E-07
GO:0006637	acyl-CoA metabolic process	3.72E-07
GO:0035383	thioester metabolic process	3.72E-07
GO:0006633	fatty acid biosynthetic process	3.79E-07
GO:0009146	purine nucleoside triphosphate catabolic process	3.95E-07

GO:0009203	ribonucleoside triphosphate catabolic process	4.03E-07
GO:0009207	purine ribonucleoside triphosphate catabolic process	4.03E-07
GO:0042254	ribosome biogenesis	4.17E-07
GO:0006270	DNA replication initiation	4.67E-07
GO:0043648	dicarboxylic acid metabolic process	4.94E-07
GO:0009225	nucleotide-sugar metabolic process	5.01E-07
GO:0006501	C-terminal protein lipidation	5.35E-07
GO:0044262	cellular carbohydrate metabolic process	5.70E-07
GO:0006661	phosphatidylinositol biosynthetic process	6.15E-07
GO:1901658	glycosyl compound catabolic process	6.31E-07
GO:1901292	nucleoside phosphate catabolic process	6.42E-07
GO:0009166	nucleotide catabolic process	6.96E-07
GO:1901136	carbohydrate derivative catabolic process	8.11E-07
GO:0072523	purine-containing compound catabolic process	8.47E-07
GO:0006195	purine nucleotide catabolic process	8.80E-07
GO:0006695	cholesterol biosynthetic process	9.29E-07
GO:0006901	vesicle coating	9.29E-07
GO:0009109	coenzyme catabolic process	9.29E-07
GO:0030071	regulation of mitotic metaphase/anaphase transition	1.08E-06
GO:0050684	regulation of mRNA processing	1.08E-06
GO:0006400	tRNA modification	1.14E-06
GO:0009081	branched-chain amino acid metabolic process	1.14E-06
GO:0006986	response to unfolded protein	1.21E-06

GO:0000289	nuclear-transcribed mRNA poly(A) tail shortening	1.26E-06
GO:0006890	retrograde vesicle-mediated transport, Golgi to ER	1.26E-06
GO:0043603	cellular amide metabolic process	1.26E-06
GO:0006152	purine nucleoside catabolic process	1.29E-06
GO:0046130	purine ribonucleoside catabolic process	1.29E-06
GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I	1.30E-06
GO:0006325	chromatin organization	1.41E-06
GO:0071174	mitotic cell cycle spindle checkpoint	1.53E-06
GO:0000096	sulfur amino acid metabolic process	1.53E-06
GO:0051262	protein tetramerization	1.66E-06
GO:0009226	nucleotide-sugar biosynthetic process	1.73E-06
GO:0044242	cellular lipid catabolic process	1.98E-06
GO:0071173	spindle assembly checkpoint	2.06E-06
GO:0007094	mitotic cell cycle spindle assembly checkpoint	2.06E-06
GO:0000724	double-strand break repair via homologous recombination	2.12E-06
GO:0006479	protein methylation	2.17E-06
GO:0008213	protein alkylation	2.17E-06
GO:0016126	sterol biosynthetic process	2.29E-06
GO:0090305	nucleic acid phosphodiester bond hydrolysis	2.45E-06
GO:0009154	purine ribonucleotide catabolic process	2.47E-06
GO:0046356	acetyl-CoA catabolic process	2.60E-06
GO:0006694	steroid biosynthetic process	2.68E-06
GO:0006297	nucleotide-excision repair, DNA gap filling	2.80E-06

GO:0022414	reproductive process	3.02E-06
GO:0006473	protein acetylation	3.07E-06
GO:0043543	protein acylation	3.31E-06
GO:0009261	ribonucleotide catabolic process	3.54E-06
GO:0006888	ER to Golgi vesicle-mediated transport	4.21E-06
GO:0009201	ribonucleoside triphosphate biosynthetic process	4.50E-06
GO:0006475	internal protein amino acid acetylation	4.61E-06
GO:0009164	nucleoside catabolic process	4.63E-06
GO:0010948	negative regulation of cell cycle process	4.78E-06
GO:0031577	spindle checkpoint	4.81E-06
GO:0000086	G2/M transition of mitotic cell cycle	5.40E-06
GO:0009142	nucleoside triphosphate biosynthetic process	5.42E-06
GO:0000725	recombinational repair	5.48E-06
GO:0042454	ribonucleoside catabolic process	5.70E-06
GO:0009064	glutamine family amino acid metabolic process	5.80E-06
GO:0006099	tricarboxylic acid cycle	5.86E-06
GO:0007131	reciprocal meiotic recombination	5.86E-06
GO:0035825	reciprocal DNA recombination	5.86E-06
GO:0045841	negative regulation of mitotic metaphase/anaphase transition	7.01E-06
GO:0034440	lipid oxidation	7.41E-06
GO:0051289	protein homotetramerization	8.10E-06
GO:0006521	regulation of cellular amino acid metabolic process	8.20E-06
GO:0007018	microtubule-based movement	8.28E-06

GO:0006805	xenobiotic metabolic process	8.76E-06
GO:1901615	organic hydroxy compound metabolic process	9.22E-06
GO:0016310	phosphorylation	9.34E-06
GO:0070979	protein K11-linked ubiquitination	9.86E-06
GO:0051187	cofactor catabolic process	1.00E-05
GO:0009066	aspartate family amino acid metabolic process	1.00E-05
GO:0090068	positive regulation of cell cycle process	1.08E-05
GO:0042274	ribosomal small subunit biogenesis	1.15E-05
GO:0046165	alcohol biosynthetic process	1.30E-05
GO:0009069	serine family amino acid metabolic process	1.31E-05
GO:0019395	fatty acid oxidation	1.36E-05
GO:0009145	purine nucleoside triphosphate biosynthetic process	1.36E-05
GO:0006505	GPI anchor metabolic process	1.60E-05
GO:0009112	nucleobase metabolic process	1.60E-05
GO:0006541	glutamine metabolic process	1.66E-05
GO:0009124	nucleoside monophosphate biosynthetic process	1.72E-05
GO:0007093	mitotic cell cycle checkpoint	1.92E-05
GO:0006739	NADP metabolic process	2.14E-05
GO:0009206	purine ribonucleoside triphosphate biosynthetic process	2.46E-05
GO:0006383	transcription from RNA polymerase III promoter	2.71E-05
GO:0016226	iron-sulfur cluster assembly	2.97E-05
GO:0031163	metallo-sulfur cluster assembly	2.97E-05
GO:0042273	ribosomal large subunit biogenesis	2.97E-05

GO:0048205	COPI coating of Golgi vesicle	2.97E-05
GO:0048200	Golgi transport vesicle coating	2.97E-05
GO:0006308	DNA catabolic process	3.02E-05
GO:0048024	regulation of mRNA splicing, via spliceosome	3.12E-05
GO:0010608	posttranscriptional regulation of gene expression	3.41E-05
GO:0006506	GPI anchor biosynthetic process	3.70E-05
GO:0008299	isoprenoid biosynthetic process	3.70E-05
GO:0016254	preassembly of GPI anchor in ER membrane	4.00E-05
GO:0051568	histone H3-K4 methylation	4.00E-05
GO:0009067	aspartate family amino acid biosynthetic process	4.00E-05
GO:0006721	terpenoid metabolic process	4.10E-05
GO:0006643	membrane lipid metabolic process	5.19E-05
GO:0006284	base-excision repair	5.42E-05
GO:0072329	monocarboxylic acid catabolic process	5.68E-05
GO:0033238	regulation of cellular amine metabolic process	6.07E-05
GO:0018393	internal peptidyl-lysine acetylation	6.20E-05
GO:0035967	cellular response to topologically incorrect protein	6.20E-05
GO:0018394	peptidyl-lysine acetylation	6.23E-05
GO:0046488	phosphatidylinositol metabolic process	6.23E-05
GO:0022900	electron transport chain	6.61E-05
GO:0009156	ribonucleoside monophosphate biosynthetic process	6.64E-05
GO:0043484	regulation of RNA splicing	8.33E-05
GO:0043574	peroxisomal transport	8.47E-05

GO:0006625	protein targeting to peroxisome	8.47E-05
GO:0072663	establishment of protein localization to peroxisome	8.47E-05
GO:0006513	protein monoubiquitination	8.59E-05
GO:0019751	polyol metabolic process	9.05E-05
GO:0006482	protein demethylation	9.62E-05
GO:0006386	termination of RNA polymerase III transcription	9.62E-05
GO:0006385	transcription elongation from RNA polymerase III promoter	9.62E-05
GO:0008214	protein dealkylation	9.62E-05
GO:0016137	glycoside metabolic process	9.62E-05
GO:0006779	porphyrin-containing compound biosynthetic process	9.78E-05
GO:0071806	protein transmembrane transport	1.01E-04
GO:0065002	intracellular protein transmembrane transport	1.01E-04
GO:0016571	histone methylation	1.02E-04
GO:0016042	lipid catabolic process	1.26E-04
GO:0006417	regulation of translation	1.27E-04
GO:0006446	regulation of translational initiation	1.31E-04
GO:0009247	glycolipid biosynthetic process	1.34E-04
GO:0000387	spliceosomal snRNP assembly	1.42E-04
GO:0043631	RNA polyadenylation	1.42E-04
GO:0008202	steroid metabolic process	1.46E-04
GO:1901617	organic hydroxy compound biosynthetic process	1.46E-04
GO:0042398	cellular modified amino acid biosynthetic process	1.49E-04
GO:0046777	protein autophosphorylation	1.56E-04

GO:0070646	protein modification by small protein removal	1.56E-04
GO:0006839	mitochondrial transport	1.58E-04
GO:0046782	regulation of viral transcription	1.65E-04
GO:0016573	histone acetylation	1.66E-04
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	1.66E-04
GO:0007059	chromosome segregation	1.67E-04
GO:0009161	ribonucleoside monophosphate metabolic process	1.69E-04
GO:0034968	histone lysine methylation	1.72E-04
GO:0006778	porphyrin-containing compound metabolic process	1.72E-04
GO:0070988	demethylation	1.72E-04
GO:0000245	spliceosomal complex assembly	1.92E-04
GO:0001522	pseudouridine synthesis	1.92E-04
GO:0006103	2-oxoglutarate metabolic process	1.92E-04
GO:0015936	coenzyme A metabolic process	1.92E-04
GO:0006515	misfolded or incompletely synthesized protein catabolic process	1.98E-04
GO:0006069	ethanol oxidation	1.98E-04
GO:0032968	positive regulation of transcription elongation from RNA polymerase II promoter	1.98E-04
GO:0036109	alpha-linolenic acid metabolic process	1.98E-04
GO:0006366	transcription from RNA polymerase II promoter	2.15E-04
GO:0006754	ATP biosynthetic process	2.18E-04
GO:0001510	RNA methylation	2.30E-04
GO:0016577	histone demethylation	2.30E-04
GO:0042375	quinone cofactor metabolic process	2.30E-04

GO:0046174	polyol catabolic process	2.30E-04
GO:0009070	serine family amino acid biosynthetic process	2.30E-04
GO:0000097	sulfur amino acid biosynthetic process	2.30E-04
GO:0007126	meiosis	2.51E-04
GO:0006984	ER-nucleus signaling pathway	2.60E-04
GO:0044743	intracellular protein transmembrane import	2.60E-04
GO:0061136	regulation of proteasomal protein catabolic process	2.64E-04
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	2.68E-04
GO:0002495	antigen processing and presentation of peptide antigen via MHC class II	2.68E-04
GO:0008643	carbohydrate transport	2.70E-04
GO:0033014	tetrapyrrole biosynthetic process	2.74E-04
GO:0016197	endosomal transport	3.05E-04
GO:0045839	negative regulation of mitosis	3.12E-04
GO:0046164	alcohol catabolic process	3.12E-04
GO:0051784	negative regulation of nuclear division	3.12E-04
GO:0009123	nucleoside monophosphate metabolic process	3.86E-04
GO:0043623	cellular protein complex assembly	4.02E-04
GO:0017038	protein import	4.03E-04
GO:0006879	cellular iron ion homeostasis	4.19E-04
GO:0048524	positive regulation of viral reproduction	4.20E-04
GO:0006518	peptide metabolic process	4.28E-04
GO:0071897	DNA biosynthetic process	4.30E-04
GO:0006361	transcription initiation from RNA polymerase I promoter	4.30E-04

GO:0032786	positive regulation of DNA-dependent transcription, elongation	4.34E-04
GO:0007031	peroxisome organization	4.34E-04
GO:0033365	protein localization to organelle	4.38E-04
GO:0046467	membrane lipid biosynthetic process	4.47E-04
GO:0072527	pyrimidine-containing compound metabolic process	4.47E-04
GO:0032268	regulation of cellular protein metabolic process	4.66E-04
GO:0046134	pyrimidine nucleoside biosynthetic process	4.67E-04
GO:0010565	regulation of cellular ketone metabolic process	4.67E-04
GO:0006575	cellular modified amino acid metabolic process	4.93E-04
GO:0031400	negative regulation of protein modification process	4.94E-04
GO:0006734	NADH metabolic process	5.11E-04
GO:0071044	histone mRNA catabolic process	5.11E-04
GO:0050686	negative regulation of mRNA processing	5.11E-04
GO:0051571	positive regulation of histone H3-K4 methylation	5.11E-04
GO:0006468	protein phosphorylation	5.38E-04
GO:1901661	quinone metabolic process	5.47E-04
GO:0070076	histone lysine demethylation	5.47E-04
GO:0048208	COPII vesicle coating	5.47E-04
GO:0006891	intra-Golgi vesicle-mediated transport	5.53E-04
GO:0009220	pyrimidine ribonucleotide biosynthetic process	5.53E-04
GO:0022607	cellular component assembly	5.77E-04
GO:0072528	pyrimidine-containing compound biosynthetic process	6.12E-04
GO:0034308	primary alcohol metabolic process	6.59E-04

GO:0006362	transcription elongation from RNA polymerase I promoter	6.59E-04
GO:0019674	NAD metabolic process	6.71E-04
GO:0009218	pyrimidine ribonucleotide metabolic process	6.71E-04
GO:0034620	cellular response to unfolded protein	6.78E-04
GO:0032269	negative regulation of cellular protein metabolic process	6.96E-04
GO:0006213	pyrimidine nucleoside metabolic process	6.98E-04
GO:0030163	protein catabolic process	7.15E-04
GO:0006950	response to stress	7.18E-04
GO:0033559	unsaturated fatty acid metabolic process	7.33E-04
GO:0006221	pyrimidine nucleotide biosynthetic process	8.74E-04
GO:0007292	female gamete generation	8.74E-04
GO:0006783	heme biosynthetic process	8.87E-04
GO:0090329	regulation of DNA-dependent DNA replication	8.87E-04
GO:0018022	peptidyl-lysine methylation	9.67E-04
GO:0006516	glycoprotein catabolic process	9.67E-04
GO:0051790	short-chain fatty acid biosynthetic process	9.67E-04

Appendix vii – Overrepresented G.O. Function terms (*H. sapiens*).

GO Term	Description	P-value
GO:0003824	catalytic activity	8.84E-277
GO:1901265	nucleoside phosphate binding	6.38E-150
GO:0000166	nucleotide binding	1.29E-149
GO:0036094	small molecule binding	1.51E-142
GO:0005524	ATP binding	9.56E-116
GO:0032559	adenyl ribonucleotide binding	1.23E-115
GO:0030554	adenyl nucleotide binding	1.61E-115
GO:0032549	ribonucleoside binding	1.15E-110
GO:0032553	ribonucleotide binding	3.66E-110
GO:0001882	nucleoside binding	1.03E-109
GO:0032550	purine ribonucleoside binding	1.68E-109
GO:0035639	purine ribonucleoside triphosphate binding	3.00E-109
GO:0001883	purine nucleoside binding	7.02E-109
GO:0017076	purine nucleotide binding	9.83E-109
GO:0032555	purine ribonucleotide binding	3.43E-108
GO:0043168	anion binding	2.19E-95
GO:0016787	hydrolase activity	3.51E-78
GO:1901363	heterocyclic compound binding	7.45E-76
GO:0097159	organic cyclic compound binding	7.00E-75
GO:0003723	RNA binding	4.68E-71
GO:00168	ATPase activity	4.16E

87		-66
GO:0016740	transferase activity	1.50E-65
GO:0016817	hydrolase activity, acting on acid anhydrides	1.67E-65
GO:0016462	pyrophosphatase activity	2.35E-65
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	2.80E-65
GO:0017111	nucleoside-triphosphatase activity	7.12E-64
GO:0042623	ATPase activity, coupled	6.03E-58
GO:0043167	ion binding	3.05E-49
GO:0016491	oxidoreductase activity	2.71E-43
GO:0003674	molecular_function	1.10E-36
GO:0004386	helicase activity	9.50E-36
GO:0016874	ligase activity	7.47E-35
GO:0048037	cofactor binding	4.46E-34
GO:0016772	transferase activity, transferring phosphorus-containing groups	3.45E-33
GO:0008026	ATP-dependent helicase activity	2.66E-29
GO:0070035	purine NTP-dependent helicase activity	2.66E-29
GO:0005488	binding	1.69E-28
GO:0050662	coenzyme binding	3.96E-26
GO:0016741	transferase activity, transferring one-carbon groups	2.85E-25
GO:0003735	structural constituent of ribosome	2.09E-24
GO:0008168	methyltransferase activity	1.24E-23
GO:0016879	ligase activity, forming carbon-nitrogen bonds	1.67E-21
GO:0016301	kinase activity	4.57E-21

GO:0008094	DNA-dependent ATPase activity	9.57E-21
GO:0016773	phosphotransferase activity, alcohol group as acceptor	6.97E-20
GO:0015399	primary active transmembrane transporter activity	1.77E-19
GO:0015405	P-P-bond-hydrolysis-driven transmembrane transporter activity	1.77E-19
GO:0008135	translation factor activity, nucleic acid binding	1.77E-19
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	7.33E-19
GO:0051540	metal cluster binding	2.15E-18
GO:0051536	iron-sulfur cluster binding	2.15E-18
GO:0016779	nucleotidyltransferase activity	3.48E-18
GO:0043492	ATPase activity, coupled to movement of substances	5.10E-18
GO:0042626	ATPase activity, coupled to transmembrane movement of substances	5.10E-18
GO:0000287	magnesium ion binding	7.98E-18
GO:0016820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	1.10E-17
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	3.54E-16
GO:0016881	acid-amino acid ligase activity	1.50E-15
GO:0004674	protein serine/threonine kinase activity	1.75E-15
GO:0022804	active transmembrane transporter activity	5.12E-15
GO:0009055	electron carrier activity	5.48E-15
GO:0003743	translation initiation factor activity	2.06E-14
GO:0004812	aminoacyl-tRNA ligase activity	2.49E-14
GO:0016875	ligase activity, forming carbon-oxygen bonds	2.49E-14
GO:0016876	ligase activity, forming aminoacyl-tRNA and related compounds	2.49E-14
GO:0016788	hydrolase activity, acting on ester bonds	5.17E-14

GO:0019787	small conjugating protein ligase activity	6.86E-14
GO:0051082	unfolded protein binding	1.96E-13
GO:0051539	4 iron, 4 sulfur cluster binding	2.10E-13
GO:0004518	nuclease activity	5.43E-13
GO:0004842	ubiquitin-protein ligase activity	9.65E-13
GO:0003678	DNA helicase activity	2.37E-12
GO:0004527	exonuclease activity	3.41E-12
GO:0004672	protein kinase activity	7.96E-12
GO:0016853	isomerase activity	8.66E-12
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	1.20E-11
GO:0005506	iron ion binding	3.03E-11
GO:0042625	ATPase activity, coupled to transmembrane movement of ions	4.47E-11
GO:0016903	oxidoreductase activity, acting on the aldehyde or oxo group of donors	1.03E-10
GO:0004003	ATP-dependent DNA helicase activity	3.29E-10
GO:0000049	tRNA binding	8.20E-10
GO:0050660	flavin adenine dinucleotide binding	9.69E-10
GO:0051287	NAD binding	1.20E-09
GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	1.42E-09
GO:0043021	ribonucleoprotein complex binding	4.21E-09
GO:0003724	RNA helicase activity	5.08E-09
GO:0034062	RNA polymerase activity	6.07E-09
GO:0003899	DNA-directed RNA polymerase activity	6.07E-09
GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	7.08E-09

GO:0019829	cation-transporting ATPase activity	1.26E-08
GO:0031072	heat shock protein binding	2.14E-08
GO:0008276	protein methyltransferase activity	3.52E-08
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	3.74E-08
GO:0016859	cis-trans isomerase activity	4.09E-08
GO:0003755	peptidyl-prolyl cis-trans isomerase activity	9.46E-08
GO:0003729	mRNA binding	1.12E-07
GO:0050661	NADP binding	1.69E-07
GO:0005515	protein binding	2.51E-07
GO:0016829	lyase activity	2.89E-07
GO:0008408	3'-5' exonuclease activity	4.17E-07
GO:0008186	RNA-dependent ATPase activity	4.67E-07
GO:0016620	oxidoreductase activity, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor	5.01E-07
GO:0043022	ribosome binding	5.01E-07
GO:0004722	protein serine/threonine phosphatase activity	5.05E-07
GO:0051213	dioxygenase activity	5.64E-07
GO:0008170	N-methyltransferase activity	6.98E-07
GO:0004497	monooxygenase activity	8.09E-07
GO:0008565	protein transporter activity	1.01E-06
GO:0020037	heme binding	1.07E-06
GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	1.11E-06
GO:0003777	microtubule motor activity	1.11E-06
GO:0004004	ATP-dependent RNA helicase activity	1.14E-06

GO:0008173	RNA methyltransferase activity	1.26E-06
GO:0003684	damaged DNA binding	1.38E-06
GO:0003676	nucleic acid binding	1.53E-06
GO:0016702	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen	2.17E-06
GO:0022892	substrate-specific transporter activity	2.59E-06
GO:0030170	pyridoxal phosphate binding	2.61E-06
GO:0016791	phosphatase activity	5.45E-06
GO:0034061	DNA polymerase activity	5.86E-06
GO:0016746	transferase activity, transferring acyl groups	6.17E-06
GO:0016278	lysine N-methyltransferase activity	6.60E-06
GO:0016279	protein-lysine N-methyltransferase activity	6.60E-06
GO:0004532	exoribonuclease activity	6.82E-06
GO:0016769	transferase activity, transferring nitrogenous groups	6.91E-06
GO:0016651	oxidoreductase activity, acting on NADH or NADPH	7.61E-06
GO:0005215	transporter activity	8.30E-06
GO:0046906	tetrapyrrole binding	8.83E-06
GO:0008233	peptidase activity	9.00E-06
GO:0016709	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NADH or NADPH as one donor, and incorporation of one atom of oxygen	9.86E-06
GO:0042054	histone methyltransferase activity	1.08E-05
GO:0016796	exonuclease activity, active with either ribo- or deoxyribonucleic acids and producing 5'-phosphomonoesters	1.31E-05
GO:0004519	endonuclease activity	1.41E-05
GO:0070011	peptidase activity, acting on L-amino acid peptides	1.42E-05
GO:00036	single-stranded DNA binding	1.48E

97		-05
GO:0008483	transaminase activity	1.60E-05
GO:0005048	signal sequence binding	1.60E-05
GO:0016747	transferase activity, transferring acyl groups other than amino-acyl groups	1.65E-05
GO:0016896	exoribonuclease activity, producing 5'-phosphomonoesters	1.66E-05
GO:0042578	phosphoric ester hydrolase activity	1.78E-05
GO:0022857	transmembrane transporter activity	2.03E-05
GO:0003887	DNA-directed DNA polymerase activity	2.93E-05
GO:0016667	oxidoreductase activity, acting on a sulfur group of donors	3.32E-05
GO:0016757	transferase activity, transferring glycosyl groups	3.67E-05
GO:0042393	histone binding	3.78E-05
GO:0032451	demethylase activity	4.00E-05
GO:0016758	transferase activity, transferring hexosyl groups	4.64E-05
GO:0004298	threonine-type endopeptidase activity	6.47E-05
GO:0070003	threonine-type peptidase activity	6.47E-05
GO:0048038	quinone binding	7.68E-05
GO:0004659	prenyltransferase activity	7.68E-05
GO:0000030	mannosyltransferase activity	8.47E-05
GO:0004033	aldo-keto reductase (NADP) activity	9.62E-05
GO:0000175	3'-5'-exoribonuclease activity	9.62E-05
GO:0016712	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	9.78E-05
GO:0018024	histone-lysine N-methyltransferase activity	1.07E-04
GO:0043169	cation binding	1.16E-04

GO:0005525	GTP binding	1.27E-04
GO:0016835	carbon-oxygen lyase activity	1.31E-04
GO:0019843	rRNA binding	1.34E-04
GO:0008238	exopeptidase activity	1.60E-04
GO:0003774	motor activity	1.65E-04
GO:0015297	antiporter activity	1.87E-04
GO:0003746	translation elongation factor activity	1.92E-04
GO:0051537	2 iron, 2 sulfur cluster binding	1.92E-04
GO:0000339	RNA cap binding	1.98E-04
GO:0032452	histone demethylase activity	1.98E-04
GO:0004576	oligosaccharyl transferase activity	1.98E-04
GO:0005388	calcium-transporting ATPase activity	1.98E-04
GO:0022891	substrate-specific transmembrane transporter activity	2.01E-04
GO:0032561	guanyl ribonucleotide binding	2.04E-04
GO:0019001	guanyl nucleotide binding	2.04E-04
GO:0046961	proton-transporting ATPase activity, rotational mechanism	2.30E-04
GO:0046872	metal ion binding	2.36E-04
GO:0030145	manganese ion binding	2.47E-04
GO:0019200	carbohydrate kinase activity	3.07E-04
GO:0004520	endodeoxyribonuclease activity	3.42E-04
GO:0015932	nucleobase-containing compound transmembrane transporter activity	4.30E-04
GO:0004709	MAP kinase kinase kinase activity	4.34E-04
GO:0008353	RNA polymerase II carboxy-terminal domain kinase activity	4.34E-04

GO:0004540	ribonuclease activity	4.48E-04
GO:0004579	dolichyl-diphosphooligosaccharide-protein glycotransferase activity	5.11E-04
GO:0004012	phospholipid-translocating ATPase activity	5.11E-04
GO:0008641	small protein activating enzyme activity	5.11E-04
GO:0016634	oxidoreductase activity, acting on the CH-CH group of donors, oxygen as acceptor	5.11E-04
GO:0030515	snoRNA binding	5.11E-04
GO:0046933	hydrogen ion transporting ATP synthase activity, rotational mechanism	5.11E-04
GO:0030983	mismatched DNA binding	5.47E-04
GO:0070577	histone acetyl-lysine binding	5.47E-04
GO:0017069	snRNA binding	5.47E-04
GO:0016866	intramolecular transferase activity	5.53E-04
GO:0016229	steroid dehydrogenase activity	5.53E-04
GO:0003924	GTPase activity	5.61E-04
GO:0016765	transferase activity, transferring alkyl or aryl (other than methyl) groups	6.47E-04
GO:0016628	oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	6.59E-04
GO:0043130	ubiquitin binding	7.15E-04
GO:0016706	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	7.78E-04
GO:0003727	single-stranded RNA binding	7.78E-04
GO:0016836	hydro-lyase activity	8.74E-04
GO:0070330	aromatase activity	8.87E-04
GO:0004190	aspartic-type endopeptidase activity	9.67E-04
GO:0070001	aspartic-type peptidase activity	9.67E-04

Appendix viii – Overrepresented G.O. Cellular Component terms (*H. sapiens*)

GO Term	Description	P-value
GO:0044446	intracellular organelle part	1.14E-120
GO:0044422	organelle part	8.58E-119
GO:0044444	cytoplasmic part	2.32E-108
GO:0044424	intracellular part	2.68E-107
GO:0044428	nuclear part	1.09E-67
GO:0030529	ribonucleoprotein complex	2.71E-62
GO:0005829	cytosol	4.65E-62
GO:0032991	macromolecular complex	1.43E-60
GO:0043229	intracellular organelle	4.54E-49
GO:0043226	organelle	1.36E-48
GO:0005739	mitochondrion	3.73E-45
GO:0005654	nucleoplasm	5.24E-38
GO:0043227	membrane-bounded organelle	1.97E-37
GO:0043231	intracellular membrane-bounded organelle	2.46E-37
GO:0044464	cell part	1.64E-34
GO:0031090	organelle membrane	1.11E-31
GO:0005681	spliceosomal complex	3.34E-30
GO:0043228	non-membrane-bounded organelle	2.42E-28
GO:0043232	intracellular non-membrane-bounded organelle	2.42E-28
GO:0005759	mitochondrial matrix	1.84E-27
GO:0043234	protein complex	3.49E-27
GO:0044429	mitochondrial part	7.78E-27
GO:0005730	nucleolus	2.57E-24
GO:0044391	ribosomal subunit	2.82E-23
GO:0044445	cytosolic part	2.28E-22
GO:0005789	endoplasmic reticulum membrane	7.60E-21
GO:0071013	catalytic step 2 spliceosome	1.40E-20
GO:0022625	cytosolic large ribosomal subunit	7.78E-17
GO:0044432	endoplasmic reticulum part	2.37E-16
GO:0015934	large ribosomal subunit	1.25E-15
GO:0005737	cytoplasm	1.97E-15
GO:0005743	mitochondrial inner membrane	7.01E-15
GO:0022627	cytosolic small ribosomal subunit	8.35E-14
GO:0070013	intracellular organelle lumen	4.11E-13
GO:0019866	organelle inner membrane	8.85E-13
GO:0031966	mitochondrial membrane	2.68E-12
GO:0031974	membrane-enclosed lumen	2.92E-11
GO:0005643	nuclear pore	5.81E-11

GO:0000151	ubiquitin ligase complex	6.59E-11
GO:0000502	proteasome complex	7.01E-11
GO:0042579	microbody	7.94E-11
GO:0030532	small nuclear ribonucleoprotein complex	8.42E-11
GO:0005634	nucleus	3.68E-10
GO:0043233	organelle lumen	4.93E-10
GO:0044427	chromosomal part	5.72E-10
GO:0046930	pore complex	9.05E-10
GO:0030880	RNA polymerase complex	1.37E-09
GO:0016604	nuclear body	2.64E-09
GO:0000428	DNA-directed RNA polymerase complex	3.22E-09
GO:0055029	nuclear DNA-directed RNA polymerase complex	3.22E-09
GO:0044451	nucleoplasm part	5.13E-09
GO:0005777	peroxisome	6.98E-09
GO:0015935	small ribosomal subunit	1.14E-08
GO:0005840	ribosome	1.80E-08
GO:0005689	U12-type spliceosomal complex	3.12E-08
GO:0044454	nuclear chromosome part	3.61E-08
GO:0005694	chromosome	1.65E-07
GO:0044438	microbody part	2.33E-07
GO:0044439	peroxisomal part	2.33E-07
GO:0022624	proteasome accessory complex	2.59E-07
GO:0005783	endoplasmic reticulum	2.69E-07
GO:0016607	nuclear speck	2.96E-07
GO:0000781	chromosome, telomeric region	3.23E-07
GO:0005815	microtubule organizing center	3.35E-07
GO:0034708	methyltransferase complex	3.57E-07
GO:0005778	peroxisomal membrane	3.57E-07
GO:0031903	microbody membrane	3.57E-07
GO:0031461	cullin-RING ubiquitin ligase complex	6.15E-07
GO:0005852	eukaryotic translation initiation factor 3 complex	6.69E-07
GO:0035097	histone methyltransferase complex	6.98E-07
GO:0016585	chromatin remodeling complex	1.87E-06
GO:0005874	microtubule	3.32E-06
GO:0071339	MLL1 complex	5.86E-06
GO:0044665	MLL1/2 complex	5.86E-06
GO:0000152	nuclear ubiquitin ligase complex	5.86E-06
GO:0030117	membrane coat	1.72E-05
GO:0005813	centrosome	3.17E-05
GO:0015630	microtubule cytoskeleton	3.66E-05

GO:0042470	melanosome	6.20E-05
GO:0048770	pigment granule	6.20E-05
GO:0005680	anaphase-promoting complex	6.47E-05
GO:0000784	nuclear chromosome, telomeric region	7.68E-05
GO:0005675	holo TFIIH complex	7.68E-05
GO:0030684	preribosome	9.62E-05
GO:0005732	small nucleolar ribonucleoprotein complex	9.62E-05
GO:0005665	DNA-directed RNA polymerase II, core complex	9.62E-05
GO:0008023	transcription elongation factor complex	1.42E-04
GO:0008250	oligosaccharyltransferase complex	1.98E-04
GO:0042555	MCM complex	1.98E-04
GO:0005929	cilium	2.01E-04
GO:0005782	peroxisomal matrix	2.10E-04
GO:0009295	nucleoid	2.10E-04
GO:0031907	microbody lumen	2.10E-04
GO:0033202	DNA helicase complex	2.30E-04
GO:0031011	Ino80 complex	2.30E-04
GO:0097346	INO80-type complex	2.30E-04
GO:0008287	protein serine/threonine phosphatase complex	2.47E-04
GO:0000793	condensed chromosome	3.40E-04
GO:0030120	vesicle coat	4.07E-04
GO:0042645	mitochondrial nucleoid	4.07E-04
GO:0000123	histone acetyltransferase complex	4.40E-04
GO:0005838	proteasome regulatory particle	5.11E-04
GO:0019773	proteasome core complex, alpha-subunit complex	5.11E-04
GO:0005669	transcription factor TFIID complex	6.59E-04
GO:0005875	microtubule associated complex	7.81E-04
GO:0005839	proteasome core complex	8.87E-04
GO:0000178	exosome (RNase complex)	8.87E-04
GO:0000139	Golgi membrane	9.41E-04
GO:0035267	NuA4 histone acetyltransferase complex	9.67E-04

Appendix ix – Underrepresented G.O. Process terms (*H. sapiens*)

GO Term	Description	P-value
GO:0022610	biological adhesion	1.47E-28
GO:0007155	cell adhesion	1.47E-28
GO:0051239	regulation of multicellular organismal process	6.65E-26
GO:0050793	regulation of developmental process	9.10E-26
GO:0032502	developmental process	1.93E-24
GO:0007166	cell surface receptor signaling pathway	5.01E-23
GO:2000026	regulation of multicellular organismal development	6.99E-22
GO:0050789	regulation of biological process	1.09E-21
GO:0048583	regulation of response to stimulus	2.95E-21
GO:0050794	regulation of cellular process	1.93E-20
GO:0048856	anatomical structure development	4.60E-20
GO:0007186	G-protein coupled receptor signaling pathway	4.77E-20
GO:0030154	cell differentiation	5.97E-20
GO:0048869	cellular developmental process	6.04E-20
GO:0044699	single-organism process	1.42E-19
GO:0045595	regulation of cell differentiation	5.46E-19
GO:0065007	biological regulation	6.92E-19
GO:0022603	regulation of anatomical structure morphogenesis	7.83E-18
GO:0048584	positive regulation of response to stimulus	1.05E-17
GO:0032501	multicellular organismal process	1.27E-17

GO:0009653	anatomical structure morphogenesis	1.58E-17
GO:0051094	positive regulation of developmental process	5.42E-17
GO:0044707	single-multicellular organism process	5.90E-17
GO:0007165	signal transduction	7.54E-17
GO:0002682	regulation of immune system process	1.24E-16
GO:0007267	cell-cell signaling	1.32E-16
GO:0040011	locomotion	2.48E-16
GO:0009888	tissue development	3.79E-16
GO:0044700	single organism signaling	5.02E-16
GO:0023052	signaling	5.02E-16
GO:0048646	anatomical structure formation involved in morphogenesis	6.82E-16
GO:0007154	cell communication	2.66E-15
GO:0010646	regulation of cell communication	3.77E-15
GO:0016337	cell-cell adhesion	5.13E-15
GO:0023051	regulation of signaling	8.94E-15
GO:0051093	negative regulation of developmental process	9.35E-15
GO:0016477	cell migration	2.09E-14
GO:0006952	defense response	2.81E-14
GO:0042127	regulation of cell proliferation	4.11E-14
GO:0006955	immune response	9.35E-14
GO:0006935	chemotaxis	1.65E-13
GO:0042330	taxis	1.65E-13
GO:0009966	regulation of signal transduction	1.88E-13

GO:0051240	positive regulation of multicellular organismal process	2.16E-13
GO:0023056	positive regulation of signaling	5.10E-13
GO:0045597	positive regulation of cell differentiation	6.22E-13
GO:0010647	positive regulation of cell communication	6.96E-13
GO:0045596	negative regulation of cell differentiation	8.44E-13
GO:0009611	response to wounding	1.82E-12
GO:0032101	regulation of response to external stimulus	2.02E-12
GO:0050776	regulation of immune response	3.60E-12
GO:0009967	positive regulation of signal transduction	3.86E-12
GO:0006355	regulation of transcription, DNA-dependent	4.95E-12
GO:0048731	system development	7.60E-12
GO:0006954	inflammatory response	2.17E-11
GO:0060284	regulation of cell development	1.02E-10
GO:0050900	leukocyte migration	1.05E-10
GO:0048518	positive regulation of biological process	1.15E-10
GO:0048870	cell motility	1.59E-10
GO:0050896	response to stimulus	1.69E-10
GO:2001141	regulation of RNA biosynthetic process	2.06E-10
GO:0048513	organ development	2.39E-10
GO:0022604	regulation of cell morphogenesis	2.63E-10
GO:0007156	homophilic cell adhesion	2.90E-10
GO:0002684	positive regulation of immune system process	3.00E-10
GO:0009605	response to external stimulus	5.53E-10

GO:0050865	regulation of cell activation	8.54E-10
GO:0032879	regulation of localization	1.14E-09
GO:0002694	regulation of leukocyte activation	2.52E-09
GO:0001775	cell activation	2.59E-09
GO:0008284	positive regulation of cell proliferation	2.92E-09
GO:0002696	positive regulation of leukocyte activation	3.13E-09
GO:0050867	positive regulation of cell activation	4.63E-09
GO:0002376	immune system process	6.08E-09
GO:0051249	regulation of lymphocyte activation	6.69E-09
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	9.73E-09
GO:0048522	positive regulation of cellular process	1.03E-08
GO:0051047	positive regulation of secretion	1.05E-08
GO:0048585	negative regulation of response to stimulus	1.38E-08
GO:0051046	regulation of secretion	1.49E-08
GO:0051251	positive regulation of lymphocyte activation	1.60E-08
GO:0009889	regulation of biosynthetic process	1.73E-08
GO:0008285	negative regulation of cell proliferation	1.73E-08
GO:0031326	regulation of cellular biosynthetic process	1.97E-08
GO:0030334	regulation of cell migration	2.62E-08
GO:0040012	regulation of locomotion	2.65E-08
GO:0003008	system process	2.67E-08
GO:0051252	regulation of RNA metabolic process	2.79E-08
GO:0048519	negative regulation of biological process	2.99E-08

GO:0007389	pattern specification process	3.04E-08
GO:0010740	positive regulation of intracellular protein kinase cascade	3.29E-08
GO:0030278	regulation of ossification	3.33E-08
GO:2000027	regulation of organ morphogenesis	3.34E-08
GO:0050795	regulation of behavior	3.34E-08
GO:2000112	regulation of cellular macromolecule biosynthetic process	3.73E-08
GO:0051716	cellular response to stimulus	3.94E-08
GO:0009887	organ morphogenesis	4.78E-08
GO:0061061	muscle structure development	5.69E-08
GO:2000145	regulation of cell motility	6.53E-08
GO:0050727	regulation of inflammatory response	7.86E-08
GO:0051270	regulation of cellular component movement	8.42E-08
GO:0010556	regulation of macromolecule biosynthetic process	9.16E-08
GO:0007517	muscle organ development	1.07E-07
GO:0048729	tissue morphogenesis	1.40E-07
GO:0051050	positive regulation of transport	1.47E-07
GO:0010468	regulation of gene expression	1.69E-07
GO:0019219	regulation of nucleobase-containing compound metabolic process	2.04E-07
GO:0001817	regulation of cytokine production	2.05E-07
GO:0051960	regulation of nervous system development	2.06E-07
GO:0050767	regulation of neurogenesis	2.34E-07
GO:0045321	leukocyte activation	2.35E-07
GO:0048523	negative regulation of cellular process	2.73E-07

GO:0048598	embryonic morphogenesis	2.76E-07
GO:0030155	regulation of cell adhesion	3.36E-07
GO:0051049	regulation of transport	3.73E-07
GO:0030855	epithelial cell differentiation	3.86E-07
GO:0010627	regulation of intracellular protein kinase cascade	4.56E-07
GO:0045664	regulation of neuron differentiation	4.78E-07
GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	5.20E-07
GO:0070663	regulation of leukocyte proliferation	5.30E-07
GO:0050863	regulation of T cell activation	7.74E-07
GO:0032944	regulation of mononuclear cell proliferation	9.03E-07
GO:0045765	regulation of angiogenesis	9.12E-07
GO:1901342	regulation of vasculature development	9.29E-07
GO:0045893	positive regulation of transcription, DNA-dependent	1.02E-06
GO:0006959	humoral immune response	1.07E-06
GO:0032103	positive regulation of response to external stimulus	1.10E-06
GO:0007218	neuropeptide signaling pathway	1.11E-06
GO:0050670	regulation of lymphocyte proliferation	1.18E-06
GO:0061138	morphogenesis of a branching epithelium	1.21E-06
GO:0002521	leukocyte differentiation	1.25E-06
GO:0044057	regulation of system process	1.29E-06
GO:0010628	positive regulation of gene expression	1.40E-06
GO:0050778	positive regulation of immune response	1.44E-06
GO:0010769	regulation of cell morphogenesis involved in differentiation	1.59E-06

GO:0045165	cell fate commitment	1.60E-06
GO:0050870	positive regulation of T cell activation	1.85E-06
GO:0050730	regulation of peptidyl-tyrosine phosphorylation	2.72E-06
GO:0051128	regulation of cellular component organization	2.75E-06
GO:0006928	cellular component movement	2.89E-06
GO:0001763	morphogenesis of a branching structure	3.07E-06
GO:0043410	positive regulation of MAPK cascade	3.47E-06
GO:0002683	negative regulation of immune system process	3.49E-06
GO:0050678	regulation of epithelial cell proliferation	3.76E-06
GO:0031294	lymphocyte costimulation	3.89E-06
GO:0060341	regulation of cellular localization	4.08E-06
GO:0043062	extracellular structure organization	4.16E-06
GO:0048705	skeletal system morphogenesis	4.16E-06
GO:0008544	epidermis development	4.37E-06
GO:0031424	keratinization	4.40E-06
GO:0002009	morphogenesis of an epithelium	4.64E-06
GO:0051171	regulation of nitrogen compound metabolic process	4.77E-06
GO:0030198	extracellular matrix organization	5.27E-06
GO:0031295	T cell costimulation	5.37E-06
GO:0050707	regulation of cytokine secretion	5.37E-06
GO:0048562	embryonic organ morphogenesis	6.33E-06
GO:0006357	regulation of transcription from RNA polymerase II promoter	7.77E-06
GO:0080134	regulation of response to stress	7.97E-06

GO:0007610	behavior	8.65E-06
GO:0031328	positive regulation of cellular biosynthetic process	8.68E-06
GO:0009891	positive regulation of biosynthetic process	8.96E-06
GO:0002697	regulation of immune effector process	8.98E-06
GO:0008037	cell recognition	1.02E-05
GO:0042742	defense response to bacterium	1.08E-05
GO:0051480	cytosolic calcium ion homeostasis	1.08E-05
GO:0008360	regulation of cell shape	1.09E-05
GO:0031589	cell-substrate adhesion	1.09E-05
GO:0050708	regulation of protein secretion	1.09E-05
GO:0070887	cellular response to chemical stimulus	1.37E-05
GO:0070167	regulation of biomineral tissue development	1.41E-05
GO:0060326	cell chemotaxis	1.43E-05
GO:0007160	cell-matrix adhesion	1.50E-05
GO:0040013	negative regulation of locomotion	1.52E-05
GO:0030336	negative regulation of cell migration	1.58E-05
GO:0042472	inner ear morphogenesis	1.74E-05
GO:0007187	G-protein coupled receptor signaling pathway, coupled to cyclic nucleotide second messenger	1.75E-05
GO:0040008	regulation of growth	1.82E-05
GO:0050920	regulation of chemotaxis	1.90E-05
GO:0051271	negative regulation of cellular component movement	2.28E-05
GO:0046649	lymphocyte activation	2.33E-05
GO:0048520	positive regulation of behavior	2.39E-05

GO:0070665	positive regulation of leukocyte proliferation	2.45E-05
GO:0001525	angiogenesis	2.50E-05
GO:0042221	response to chemical stimulus	2.59E-05
GO:0002768	immune response-regulating cell surface receptor signaling pathway	2.63E-05
GO:0002687	positive regulation of leukocyte migration	2.67E-05
GO:0006968	cellular defense response	2.75E-05
GO:0051051	negative regulation of transport	3.01E-05
GO:0031347	regulation of defense response	3.11E-05
GO:0032946	positive regulation of mononuclear cell proliferation	3.20E-05
GO:0050817	coagulation	3.21E-05
GO:0007596	blood coagulation	3.21E-05
GO:0009968	negative regulation of signal transduction	3.21E-05
GO:0007599	hemostasis	3.31E-05
GO:0001819	positive regulation of cytokine production	3.42E-05
GO:0042391	regulation of membrane potential	3.43E-05
GO:0045892	negative regulation of transcription, DNA-dependent	3.48E-05
GO:0002706	regulation of lymphocyte mediated immunity	3.67E-05
GO:0002688	regulation of leukocyte chemotaxis	3.67E-05
GO:0045666	positive regulation of neuron differentiation	3.67E-05
GO:0050714	positive regulation of protein secretion	3.72E-05
GO:0050877	neurological system process	3.75E-05
GO:0010629	negative regulation of gene expression	4.10E-05
GO:0032102	negative regulation of response to external stimulus	4.18E-05

GO:0050671	positive regulation of lymphocyte proliferation	4.18E-05
GO:0045637	regulation of myeloid cell differentiation	4.25E-05
GO:0002703	regulation of leukocyte mediated immunity	4.26E-05
GO:0002685	regulation of leukocyte migration	4.26E-05
GO:0045667	regulation of osteoblast differentiation	4.37E-05
GO:0040017	positive regulation of locomotion	4.43E-05
GO:0007411	axon guidance	4.61E-05
GO:0045785	positive regulation of cell adhesion	4.75E-05
GO:2000146	negative regulation of cell motility	4.75E-05
GO:0007204	elevation of cytosolic calcium ion concentration	4.75E-05
GO:0032989	cellular component morphogenesis	4.78E-05
GO:0080090	regulation of primary metabolic process	4.91E-05
GO:0048704	embryonic skeletal system morphogenesis	5.02E-05
GO:0031348	negative regulation of defense response	5.02E-05
GO:0060548	negative regulation of cell death	5.46E-05
GO:0048754	branching morphogenesis of a tube	5.57E-05
GO:0043269	regulation of ion transport	5.66E-05
GO:0008277	regulation of G-protein coupled receptor protein signaling pathway	5.75E-05
GO:0030335	positive regulation of cell migration	5.75E-05
GO:0030029	actin filament-based process	6.33E-05
GO:0002429	immune response-activating cell surface receptor signaling pathway	6.38E-05
GO:0042113	B cell activation	6.38E-05
GO:0044708	single-organism behavior	6.64E-05

GO:0009617	response to bacterium	6.66E-05
GO:0043408	regulation of MAPK cascade	6.91E-05
GO:0030500	regulation of bone mineralization	6.93E-05
GO:0006351	transcription, DNA-dependent	7.09E-05
GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	7.36E-05
GO:2001233	regulation of apoptotic signaling pathway	7.54E-05
GO:0030799	regulation of cyclic nucleotide metabolic process	7.55E-05
GO:0008217	regulation of blood pressure	7.69E-05
GO:0048608	reproductive structure development	7.69E-05
GO:2000147	positive regulation of cell motility	7.83E-05
GO:0023057	negative regulation of signaling	8.10E-05
GO:0042692	muscle cell differentiation	8.23E-05
GO:0002253	activation of immune response	8.73E-05
GO:0019222	regulation of metabolic process	9.37E-05
GO:0007264	small GTPase mediated signal transduction	9.62E-05
GO:0045685	regulation of glial cell differentiation	9.66E-05
GO:0007268	synaptic transmission	9.72E-05
GO:0030595	leukocyte chemotaxis	9.91E-05
GO:0055082	cellular chemical homeostasis	9.96E-05
GO:0051272	positive regulation of cellular component movement	1.05E-04
GO:0007399	nervous system development	1.11E-04
GO:0043069	negative regulation of programmed cell death	1.14E-04
GO:0010648	negative regulation of cell communication	1.16E-04

GO:0003013	circulatory system process	1.16E-04
GO:0010557	positive regulation of macromolecule biosynthetic process	1.19E-04
GO:0003018	vascular process in circulatory system	1.23E-04
GO:0034330	cell junction organization	1.28E-04
GO:1900371	regulation of purine nucleotide biosynthetic process	1.30E-04
GO:0030808	regulation of nucleotide biosynthetic process	1.30E-04
GO:0030802	regulation of cyclic nucleotide biosynthetic process	1.30E-04
GO:0002690	positive regulation of leukocyte chemotaxis	1.31E-04
GO:0050728	negative regulation of inflammatory response	1.31E-04
GO:0002819	regulation of adaptive immune response	1.33E-04
GO:0052548	regulation of endopeptidase activity	1.35E-04
GO:0001658	branching involved in ureteric bud morphogenesis	1.36E-04
GO:0002699	positive regulation of immune effector process	1.36E-04
GO:0050851	antigen receptor-mediated signaling pathway	1.36E-04
GO:0055074	calcium ion homeostasis	1.37E-04
GO:0007275	multicellular organismal development	1.40E-04
GO:0030182	neuron differentiation	1.41E-04
GO:0006874	cellular calcium ion homeostasis	1.49E-04
GO:0070374	positive regulation of ERK1 and ERK2 cascade	1.55E-04
GO:0050866	negative regulation of cell activation	1.55E-04
GO:0042129	regulation of T cell proliferation	1.55E-04
GO:0032940	secretion by cell	1.59E-04
GO:0043066	negative regulation of apoptotic process	1.64E-04

GO:0046903	secretion	1.67E-04
GO:0050921	positive regulation of chemotaxis	1.76E-04
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	1.76E-04
GO:0001708	cell fate specification	1.79E-04
GO:0050679	positive regulation of epithelial cell proliferation	1.98E-04
GO:0002761	regulation of myeloid leukocyte differentiation	2.01E-04
GO:0007188	adenylate cyclase-modulating G-protein coupled receptor signaling pathway	2.01E-04
GO:0001501	skeletal system development	2.03E-04
GO:0065008	regulation of biological quality	2.05E-04
GO:0003012	muscle system process	2.09E-04
GO:0052547	regulation of peptidase activity	2.16E-04
GO:0072503	cellular divalent inorganic cation homeostasis	2.17E-04
GO:0010959	regulation of metal ion transport	2.22E-04
GO:0007631	feeding behavior	2.22E-04
GO:0051173	positive regulation of nitrogen compound metabolic process	2.25E-04
GO:0010720	positive regulation of cell development	2.33E-04
GO:0045669	positive regulation of osteoblast differentiation	2.45E-04
GO:0060560	developmental growth involved in morphogenesis	2.45E-04
GO:0006140	regulation of nucleotide metabolic process	2.54E-04
GO:0008406	gonad development	2.60E-04
GO:0050878	regulation of body fluid levels	2.62E-04
GO:0002705	positive regulation of leukocyte mediated immunity	2.70E-04
GO:0002708	positive regulation of lymphocyte mediated immunity	2.70E-04

GO:0032652	regulation of interleukin-1 production	2.70E-04
GO:0007189	adenylate cyclase-activating G-protein coupled receptor signaling pathway	2.70E-04
GO:0035019	somatic stem cell maintenance	2.70E-04
GO:0006873	cellular ion homeostasis	2.77E-04
GO:0070848	response to growth factor stimulus	2.87E-04
GO:0003006	developmental process involved in reproduction	2.88E-04
GO:1900542	regulation of purine nucleotide metabolic process	2.97E-04
GO:0060627	regulation of vesicle-mediated transport	3.04E-04
GO:0043900	regulation of multi-organism process	3.05E-04
GO:0010810	regulation of cell-substrate adhesion	3.06E-04
GO:0007600	sensory perception	3.06E-04
GO:0051924	regulation of calcium ion transport	3.16E-04
GO:0000902	cell morphogenesis	3.16E-04
GO:0051253	negative regulation of RNA metabolic process	3.18E-04
GO:0051048	negative regulation of secretion	3.19E-04
GO:0001649	osteoblast differentiation	3.35E-04
GO:0050715	positive regulation of cytokine secretion	3.35E-04
GO:0060021	palate development	3.35E-04
GO:0071363	cellular response to growth factor stimulus	3.63E-04
GO:0061448	connective tissue development	3.67E-04
GO:0009607	response to biotic stimulus	3.76E-04
GO:0002673	regulation of acute inflammatory response	3.81E-04
GO:0001657	ureteric bud development	3.81E-04

GO:0001764	neuron migration	3.87E-04
GO:0034329	cell junction assembly	3.93E-04
GO:0030098	lymphocyte differentiation	3.99E-04
GO:0051897	positive regulation of protein kinase B signaling cascade	3.99E-04
GO:0030183	B cell differentiation	3.99E-04
GO:0046887	positive regulation of hormone secretion	3.99E-04
GO:0060333	interferon-gamma-mediated signaling pathway	3.99E-04
GO:0051250	negative regulation of lymphocyte activation	4.05E-04
GO:0051336	regulation of hydrolase activity	4.52E-04
GO:0045744	negative regulation of G-protein coupled receptor protein signaling pathway	4.58E-04
GO:0050880	regulation of blood vessel size	4.58E-04
GO:0090183	regulation of kidney development	4.58E-04
GO:0035150	regulation of tube size	4.58E-04
GO:0007229	integrin-mediated signaling pathway	4.58E-04
GO:0035023	regulation of Rho protein signal transduction	4.58E-04
GO:0002252	immune effector process	4.60E-04
GO:0031323	regulation of cellular metabolic process	4.68E-04
GO:0051241	negative regulation of multicellular organismal process	4.72E-04
GO:0009890	negative regulation of biosynthetic process	4.89E-04
GO:0071346	cellular response to interferon-gamma	4.92E-04
GO:0071356	cellular response to tumor necrosis factor	5.34E-04
GO:0050729	positive regulation of inflammatory response	5.34E-04
GO:0050864	regulation of B cell activation	5.34E-04

GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	5.34E-04
GO:0002526	acute inflammatory response	5.34E-04
GO:0034340	response to type I interferon	5.34E-04
GO:0071357	cellular response to type I interferon	5.34E-04
GO:0050868	negative regulation of T cell activation	5.34E-04
GO:0060337	type I interferon-mediated signaling pathway	5.34E-04
GO:0002762	negative regulation of myeloid leukocyte differentiation	5.36E-04
GO:0023021	termination of signal transduction	5.36E-04
GO:0007167	enzyme linked receptor protein signaling pathway	5.43E-04
GO:0042476	odontogenesis	5.56E-04
GO:0031344	regulation of cell projection organization	6.02E-04
GO:0042110	T cell activation	6.03E-04
GO:0071345	cellular response to cytokine stimulus	6.13E-04
GO:0003002	regionalization	6.20E-04
GO:0019220	regulation of phosphate metabolic process	6.25E-04
GO:0006936	muscle contraction	6.29E-04
GO:0072507	divalent inorganic cation homeostasis	6.35E-04
GO:0030817	regulation of cAMP biosynthetic process	6.39E-04
GO:0030814	regulation of cAMP metabolic process	6.39E-04
GO:0048660	regulation of smooth muscle cell proliferation	6.39E-04
GO:0032844	regulation of homeostatic process	6.92E-04
GO:0002695	negative regulation of leukocyte activation	7.13E-04
GO:190037	positive regulation of purine nucleotide biosynthetic process	7.14E

3		-04
GO:0045981	positive regulation of nucleotide metabolic process	7.14E-04
GO:1900544	positive regulation of purine nucleotide metabolic process	7.14E-04
GO:0002700	regulation of production of molecular mediator of immune response	7.14E-04
GO:0030801	positive regulation of cyclic nucleotide metabolic process	7.14E-04
GO:0030804	positive regulation of cyclic nucleotide biosynthetic process	7.14E-04
GO:0030810	positive regulation of nucleotide biosynthetic process	7.14E-04
GO:0045598	regulation of fat cell differentiation	7.14E-04
GO:0042475	odontogenesis of dentin-containing tooth	7.14E-04
GO:0030279	negative regulation of ossification	7.55E-04
GO:0045168	cell-cell signaling involved in cell fate commitment	7.55E-04
GO:0031128	developmental induction	7.55E-04
GO:0060537	muscle tissue development	7.59E-04
GO:0090066	regulation of anatomical structure size	7.65E-04
GO:0031327	negative regulation of cellular biosynthetic process	7.69E-04
GO:0070372	regulation of ERK1 and ERK2 cascade	7.73E-04
GO:0008016	regulation of heart contraction	7.85E-04
GO:0044087	regulation of cellular component biogenesis	7.88E-04
GO:0051216	cartilage development	8.29E-04
GO:0042108	positive regulation of cytokine biosynthetic process	8.54E-04
GO:0007507	heart development	8.89E-04
GO:0010941	regulation of cell death	8.93E-04
GO:0030036	actin cytoskeleton organization	9.24E-04

GO:0010001	glial cell differentiation	9.53E-04
GO:0061097	regulation of protein tyrosine kinase activity	9.53E-04
GO:0043270	positive regulation of ion transport	9.60E-04
GO:0030534	adult behavior	9.64E-04
GO:0032880	regulation of protein localization	9.86E-04

Appendix x - Underrepresented G.O. Function terms (*H. sapiens*)

GO Term	Description	P-value
GO:0004872	receptor activity	3.20E-51
GO:0038023	signaling receptor activity	3.64E-51
GO:0004888	transmembrane signaling receptor activity	8.72E-50
GO:0004871	signal transducer activity	1.39E-41
GO:0060089	molecular transducer activity	1.39E-41
GO:0004930	G-protein coupled receptor activity	1.37E-31
GO:0001071	nucleic acid binding transcription factor activity	4.74E-27
GO:0003700	sequence-specific DNA binding transcription factor activity	6.02E-27
GO:0005102	receptor binding	8.85E-24
GO:0043565	sequence-specific DNA binding	9.78E-21
GO:0004984	olfactory receptor activity	3.58E-20
GO:0005125	cytokine activity	6.99E-16
GO:0000981	sequence-specific DNA binding RNA polymerase II transcription factor activity	7.03E-15
GO:0005179	hormone activity	5.53E-11
GO:0005126	cytokine receptor binding	1.16E-10
GO:0003677	DNA binding	1.60E-10
GO:0001664	G-protein coupled receptor binding	1.02E-09
GO:0008083	growth factor activity	3.06E-09
GO:0005539	glycosaminoglycan binding	6.06E-09
GO:0004857	enzyme inhibitor activity	1.42E-08

GO:0000976	transcription regulatory region sequence-specific DNA binding	3.99E-08
GO:0001653	peptide receptor activity	7.13E-08
GO:0008528	G-protein coupled peptide receptor activity	7.13E-08
GO:0004896	cytokine receptor activity	2.82E-07
GO:0030414	peptidase inhibitor activity	3.10E-07
GO:0005509	calcium ion binding	3.95E-07
GO:0003705	RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity	5.09E-07
GO:0030234	enzyme regulator activity	6.35E-07
GO:0004866	endopeptidase inhibitor activity	7.07E-07
GO:0097367	carbohydrate derivative binding	1.06E-06
GO:0061135	endopeptidase regulator activity	1.53E-06
GO:0008201	heparin binding	2.10E-06
GO:0061134	peptidase regulator activity	2.17E-06
GO:0008009	chemokine activity	2.22E-06
GO:0000975	regulatory region DNA binding	2.73E-06
GO:0001067	regulatory region nucleic acid binding	2.73E-06
GO:0044212	transcription regulatory region DNA binding	4.69E-06
GO:0042379	chemokine receptor binding	5.37E-06
GO:0004879	ligand-activated sequence-specific DNA binding RNA polymerase II transcription factor activity	1.74E-05
GO:0019904	protein domain specific binding	1.82E-05
GO:0008270	zinc ion binding	1.83E-05
GO:0035591	signaling adaptor activity	3.67E-05
GO:0004867	serine-type endopeptidase inhibitor activity	3.72E-05

GO:0000982	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity	5.67E-05
GO:0008047	enzyme activator activity	6.89E-05
GO:0019199	transmembrane receptor protein kinase activity	6.93E-05
GO:0001228	RNA polymerase II transcription regulatory region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription	9.13E-05
GO:0003707	steroid hormone receptor activity	9.51E-05
GO:0008289	lipid binding	1.03E-04
GO:0005201	extracellular matrix structural constituent	1.31E-04
GO:0005543	phospholipid binding	1.48E-04
GO:0005216	ion channel activity	1.57E-04
GO:0001012	RNA polymerase II regulatory region DNA binding	2.33E-04
GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	3.07E-04
GO:0003779	actin binding	3.23E-04
GO:0070851	growth factor receptor binding	3.67E-04
GO:0030594	neurotransmitter receptor activity	3.81E-04
GO:0008301	DNA binding, bending	3.99E-04
GO:0001077	RNA polymerase II core promoter proximal region sequence-specific DNA binding transcription factor activity involved in positive regulation of transcription	3.99E-04
GO:0005070	SH3/SH2 adaptor activity	4.58E-04
GO:0019955	cytokine binding	4.58E-04
GO:0008134	transcription factor binding	5.24E-04
GO:0019838	growth factor binding	5.97E-04
GO:0022803	passive transmembrane transporter activity	9.06E-04
GO:0015267	channel activity	9.06E-04

Appendix xi - Underrepresented G.O. Cellular Component terms (*H. sapiens*)

GO Term	Description	P-value
GO:0005576	extracellular region	2.29E-49
GO:0044421	extracellular region part	9.20E-39
GO:0005886	plasma membrane	1.16E-37
GO:0005615	extracellular space	4.14E-27
GO:0044459	plasma membrane part	2.29E-24
GO:0031224	intrinsic to membrane	1.09E-22
GO:0031226	intrinsic to plasma membrane	1.09E-20
GO:0016021	integral to membrane	1.67E-19
GO:0005887	integral to plasma membrane	2.42E-19
GO:0005882	intermediate filament	6.72E-15
GO:0044425	membrane part	2.38E-14
GO:0045095	keratin filament	3.67E-11
GO:0031225	anchored to membrane	5.53E-11
GO:0030054	cell junction	7.28E-11
GO:0009897	external side of plasma membrane	9.47E-11
GO:0031012	extracellular matrix	1.39E-10
GO:0005578	proteinaceous extracellular matrix	2.42E-10
GO:0044420	extracellular matrix part	4.14E-09
GO:0043235	receptor complex	1.96E-07
GO:0005581	collagen	1.07E-06
GO:0045202	synapse	1.28E-06
GO:0009986	cell surface	5.99E-06
GO:0005796	Golgi lumen	1.23E-05
GO:0031983	vesicle lumen	2.45E-05
GO:0060205	cytoplasmic membrane-bounded vesicle lumen	2.45E-05
GO:0097060	synaptic membrane	6.41E-05
GO:0070161	anchoring junction	6.54E-05
GO:0034774	secretory granule lumen	6.85E-05
GO:0005912	adherens junction	1.03E-04
GO:0043005	neuron projection	1.30E-04
GO:0031093	platelet alpha granule lumen	1.36E-04
GO:0016020	membrane	1.51E-04
GO:0030055	cell-substrate junction	1.59E-04
GO:0034702	ion channel complex	1.96E-04
GO:0044456	synapse part	2.22E-04
GO:0045211	postsynaptic membrane	2.42E-04

GO:0030175	filopodium	2.70E-04
GO:0005924	cell-substrate adherens junction	2.99E-04
GO:0030141	secretory granule	4.95E-04
GO:0030424	axon	6.81E-04
GO:0005925	focal adhesion	6.83E-04
GO:0005911	cell-cell junction	9.70E-04

Appendix xii – Overrepresented 1-2 h G.O. Process terms

GO Term	Description	P-value
GO:0016817	hydrolase activity, acting on acid anhydrides	1.86E-13
GO:0016462	pyrophosphatase activity	2.62E-13
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	3.62E-13
GO:0017111	nucleoside-triphosphatase activity	4.52E-13
GO:0000166	nucleotide binding	8.49E-12
GO:1901265	nucleoside phosphate binding	9.33E-12
GO:0036094	small molecule binding	1.02E-11
GO:0005515	protein binding	3.51E-10
GO:1901363	heterocyclic compound binding	1.16E-09
GO:0097159	organic cyclic compound binding	1.21E-09
GO:0035639	purine ribonucleoside triphosphate binding	1.93E-09
GO:0032555	purine ribonucleotide binding	6.74E-09
GO:0032553	ribonucleotide binding	6.74E-09
GO:0017076	purine nucleotide binding	1.23E-08
GO:0005524	ATP binding	1.57E-07
GO:0005488	binding	2.55E-07
GO:0016887	ATPase activity	4.06E-07
GO:0032559	adenyl ribonucleotide binding	4.97E-07
GO:0030554	adenyl nucleotide binding	6.58E-07
GO:0042623	ATPase activity, coupled	1.45E-05
GO:0004386	helicase activity	2.78E-05
GO:000392	GTPase activity	5.26E-

4		05
GO:0051082	unfolded protein binding	8.57E-05
GO:0004738	pyruvate dehydrogenase activity	8.77E-05
GO:0004739	pyruvate dehydrogenase (acetyl-transferring) activity	8.77E-05
GO:0003723	RNA binding	1.19E-04
GO:0003676	nucleic acid binding	1.45E-04
GO:0030507	spectrin binding	3.58E-04
GO:0008026	ATP-dependent helicase activity	4.79E-04
GO:0070035	purine NTP-dependent helicase activity	4.79E-04
GO:0003688	DNA replication origin binding	5.39E-04
GO:0016787	hydrolase activity	5.56E-04
GO:0003918	DNA topoisomerase (ATP-hydrolyzing) activity	5.99E-04
GO:0051117	ATPase binding	9.86E-04

Appendix xiii - Overrepresented 1-2 h G.O. Function terms

GO Term	Description	P-value
GO:0016071	mRNA metabolic process	3.97E-27
GO:0006397	mRNA processing	4.22E-24
GO:0008380	RNA splicing	3.02E-23
GO:0000375	RNA splicing, via transesterification reactions	1.08E-22
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	5.34E-22
GO:0000398	mRNA splicing, via spliceosome	5.34E-22
GO:0022402	cell cycle process	1.08E-18
GO:0007049	cell cycle	1.27E-17
GO:0000278	mitotic cell cycle	1.38E-15
GO:0022403	cell cycle phase	2.51E-15
GO:0006396	RNA processing	6.68E-14
GO:0000082	G1/S transition of mitotic cell cycle	1.41E-13
GO:0071842	cellular component organization at cellular level	1.84E-13
GO:0006996	organelle organization	1.96E-13
GO:0006139	nucleobase-containing compound metabolic process	2.70E-13
GO:0000216	M/G1 transition of mitotic cell cycle	4.22E-13
GO:0016043	cellular component organization	6.55E-13
GO:0071841	cellular component organization or biogenesis at cellular level	8.51E-13
GO:0071840	cellular component organization or biogenesis	2.81E-12
GO:0046483	heterocycle metabolic process	6.31E-12
GO:0000075	cell cycle checkpoint	7.10E-12

GO:0034622	cellular macromolecular complex assembly	9.58E-12
GO:0044260	cellular macromolecule metabolic process	1.43E-11
GO:0051301	cell division	1.59E-11
GO:0006259	DNA metabolic process	1.97E-11
GO:0006725	cellular aromatic compound metabolic process	1.98E-11
GO:0034641	cellular nitrogen compound metabolic process	3.18E-11
GO:0090304	nucleic acid metabolic process	3.76E-11
GO:0051704	multi-organism process	3.89E-11
GO:0043170	macromolecule metabolic process	5.15E-11
GO:0016032	viral reproduction	5.79E-11
GO:0006807	nitrogen compound metabolic process	6.45E-11
GO:1901360	organic cyclic compound metabolic process	9.20E-11
GO:0071156	regulation of cell cycle arrest	1.89E-10
GO:0065003	macromolecular complex assembly	3.51E-10
GO:0034621	cellular macromolecular complex subunit organization	3.56E-10
GO:0010564	regulation of cell cycle process	7.95E-10
GO:0009987	cellular process	1.25E-09
GO:0071844	cellular component assembly at cellular level	1.54E-09
GO:0051320	S phase	2.42E-09
GO:0000084	S phase of mitotic cell cycle	2.42E-09
GO:0044238	primary metabolic process	2.81E-09
GO:0046907	intracellular transport	2.94E-09
GO:0071704	organic substance metabolic process	3.73E-09

GO:0043933	macromolecular complex subunit organization	4.09E-09
GO:0006260	DNA replication	4.45E-09
GO:0044265	cellular macromolecule catabolic process	4.75E-09
GO:0051084	'de novo' posttranslational protein folding	6.51E-09
GO:0006458	'de novo' protein folding	6.71E-09
GO:0010467	gene expression	7.18E-09
GO:0006270	DNA replication initiation	7.99E-09
GO:0022607	cellular component assembly	8.24E-09
GO:0009057	macromolecule catabolic process	1.22E-08
GO:0044237	cellular metabolic process	1.87E-08
GO:0051701	interaction with host	2.11E-08
GO:0048285	organelle fission	2.27E-08
GO:0000280	nuclear division	5.53E-08
GO:0007067	mitosis	5.53E-08
GO:0022415	viral reproductive process	5.68E-08
GO:0044419	interspecies interaction between organisms	7.13E-08
GO:0006334	nucleosome assembly	1.13E-07
GO:0051028	mRNA transport	1.43E-07
GO:0044703	multi-organism reproductive process	1.45E-07
GO:0051649	establishment of localization in cell	1.53E-07
GO:0031145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	1.61E-07
GO:0051726	regulation of cell cycle	1.65E-07
GO:0016044	cellular membrane organization	1.79E-07

GO:0002478	antigen processing and presentation of exogenous peptide antigen	1.79E-07
GO:0051258	protein polymerization	1.79E-07
GO:0050658	RNA transport	2.22E-07
GO:0050657	nucleic acid transport	2.22E-07
GO:0051236	establishment of RNA localization	2.22E-07
GO:0008152	metabolic process	2.30E-07
GO:0061024	membrane organization	2.49E-07
GO:0048002	antigen processing and presentation of peptide antigen	2.58E-07
GO:0019884	antigen processing and presentation of exogenous antigen	3.05E-07
GO:0015682	ferric iron transport	6.77E-07
GO:0033572	transferrin transport	6.77E-07
GO:0065004	protein-DNA complex assembly	7.87E-07
GO:0006405	RNA export from nucleus	8.15E-07
GO:0009141	nucleoside triphosphate metabolic process	9.05E-07
GO:0019882	antigen processing and presentation	9.54E-07
GO:0044710	single-organism metabolic process	1.14E-06
GO:0006368	transcription elongation from RNA polymerase II promoter	1.35E-06
GO:0071103	DNA conformation change	1.35E-06
GO:0006325	chromatin organization	1.54E-06
GO:0034728	nucleosome organization	1.54E-06
GO:0090382	phagosome maturation	1.56E-06
GO:0000086	G2/M transition of mitotic cell cycle	1.62E-06
GO:0009056	catabolic process	1.81E-06

GO:0075733	intracellular transport of viral material	1.95E-06
GO:0046788	egress of virus within host cell	1.95E-06
GO:0043900	regulation of multi-organism process	2.31E-06
GO:0051276	chromosome organization	2.47E-06
GO:0051603	proteolysis involved in cellular protein catabolic process	2.59E-06
GO:0044248	cellular catabolic process	2.98E-06
GO:0051443	positive regulation of ubiquitin-protein ligase activity	3.14E-06
GO:0009199	ribonucleoside triphosphate metabolic process	3.46E-06
GO:0071824	protein-DNA complex subunit organization	3.62E-06
GO:0006511	ubiquitin-dependent protein catabolic process	3.70E-06
GO:0031124	mRNA 3'-end processing	3.79E-06
GO:0051437	positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	3.85E-06
GO:0019048	virus-host interaction	4.07E-06
GO:0051351	positive regulation of ligase activity	4.08E-06
GO:0006406	mRNA export from nucleus	4.71E-06
GO:0000209	protein polyubiquitination	5.06E-06
GO:0009144	purine nucleoside triphosphate metabolic process	5.17E-06
GO:0006268	DNA unwinding involved in replication	5.42E-06
GO:0019941	modification-dependent protein catabolic process	5.44E-06
GO:0009203	ribonucleoside triphosphate catabolic process	5.84E-06
GO:0009207	purine ribonucleoside triphosphate catabolic process	5.84E-06
GO:0051439	regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	6.10E-06
GO:0006370	7-methylguanosine mRNA capping	6.38E-06

GO:0051436	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	6.63E-06
GO:0051168	nuclear export	6.73E-06
GO:0050434	positive regulation of viral transcription	6.86E-06
GO:0015931	nucleobase-containing compound transport	6.99E-06
GO:0009143	nucleoside triphosphate catabolic process	7.44E-06
GO:0048524	positive regulation of viral reproduction	7.47E-06
GO:0051438	regulation of ubiquitin-protein ligase activity	7.51E-06
GO:0009261	ribonucleotide catabolic process	7.72E-06
GO:0009146	purine nucleoside triphosphate catabolic process	7.72E-06
GO:0009154	purine ribonucleotide catabolic process	7.72E-06
GO:0043632	modification-dependent macromolecule catabolic process	7.93E-06
GO:0050792	regulation of viral reproduction	8.20E-06
GO:0006457	protein folding	8.71E-06
GO:0009259	ribonucleotide metabolic process	9.30E-06
GO:0051352	negative regulation of ligase activity	9.40E-06
GO:0051444	negative regulation of ubiquitin-protein ligase activity	9.40E-06
GO:0051340	regulation of ligase activity	9.45E-06
GO:0009205	purine ribonucleoside triphosphate metabolic process	1.02E-05
GO:0036260	RNA capping	1.15E-05
GO:0009452	7-methylguanosine RNA capping	1.15E-05
GO:0070979	protein K11-linked ubiquitination	1.54E-05
GO:0043902	positive regulation of multi-organism process	1.63E-05
GO:0046782	regulation of viral transcription	1.70E-05

GO:0010498	proteasomal protein catabolic process	1.71E-05
GO:0043161	proteasomal ubiquitin-dependent protein catabolic process	1.71E-05
GO:0061418	regulation of transcription from RNA polymerase II promoter in response to hypoxia	1.84E-05
GO:0006281	DNA repair	1.96E-05
GO:0006826	iron ion transport	2.07E-05
GO:0070423	nucleotide-binding oligomerization domain containing signaling pathway	2.67E-05
GO:0045184	establishment of protein localization	3.02E-05
GO:0016070	RNA metabolic process	3.29E-05
GO:0031397	negative regulation of protein ubiquitination	3.31E-05
GO:0007017	microtubule-based process	3.35E-05
GO:0032508	DNA duplex unwinding	3.41E-05
GO:0016567	protein ubiquitination	3.45E-05
GO:0006271	DNA strand elongation involved in DNA replication	3.61E-05
GO:0002753	cytoplasmic pattern recognition receptor signaling pathway	3.61E-05
GO:0051169	nuclear transport	3.67E-05
GO:0051318	G1 phase	3.74E-05
GO:0000080	G1 phase of mitotic cell cycle	3.74E-05
GO:2000243	positive regulation of reproductive process	4.52E-05
GO:0006913	nucleocytoplasmic transport	4.66E-05
GO:0043618	regulation of transcription from RNA polymerase II promoter in response to stress	4.85E-05
GO:0006401	RNA catabolic process	5.09E-05
GO:0000226	microtubule cytoskeleton organization	5.32E-05
GO:0032392	DNA geometric change	5.44E-05

GO:0015031	protein transport	5.99E-05
GO:0022616	DNA strand elongation	6.06E-05
GO:0000083	regulation of transcription involved in G1/S phase of mitotic cell cycle	6.17E-05
GO:0006974	response to DNA damage stimulus	6.47E-05
GO:0032480	negative regulation of type I interferon production	6.57E-05
GO:0007249	I-kappaB kinase/NF-kappaB cascade	6.57E-05
GO:0009150	purine ribonucleotide metabolic process	6.69E-05
GO:0006195	purine nucleotide catabolic process	7.44E-05
GO:0006354	DNA-dependent transcription, elongation	7.63E-05
GO:0007051	spindle organization	7.63E-05
GO:0006818	hydrogen transport	9.16E-05
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	9.27E-05
GO:0015988	energy coupled proton transmembrane transport, against electrochemical gradient	9.39E-05
GO:0015991	ATP hydrolysis coupled proton transport	9.39E-05
GO:0022414	reproductive process	9.96E-05
GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	1.09E-04
GO:0000956	nuclear-transcribed mRNA catabolic process	1.16E-04
GO:2000241	regulation of reproductive process	1.18E-04
GO:0006163	purine nucleotide metabolic process	1.19E-04
GO:0007220	Notch receptor processing	1.20E-04
GO:0072523	purine-containing compound catabolic process	1.26E-04
GO:0031323	regulation of cellular metabolic process	1.28E-04
GO:0043623	cellular protein complex assembly	1.32E-04

GO:0043620	regulation of DNA-dependent transcription in response to stress	1.42E-04
GO:0071174	mitotic cell cycle spindle checkpoint	1.45E-04
GO:0031577	spindle checkpoint	1.45E-04
GO:0007093	mitotic cell cycle checkpoint	1.54E-04
GO:0006265	DNA topological change	1.56E-04
GO:0031123	RNA 3'-end processing	1.63E-04
GO:0006283	transcription-coupled nucleotide-excision repair	1.63E-04
GO:0008150	biological_process	1.68E-04
GO:0016568	chromatin modification	1.71E-04
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	1.73E-04
GO:0002495	antigen processing and presentation of peptide antigen via MHC class II	1.73E-04
GO:0010994	free ubiquitin chain polymerization	1.76E-04
GO:0050794	regulation of cellular process	1.92E-04
GO:0032446	protein modification by small protein conjugation	1.94E-04
GO:0070683	mitotic cell cycle G2/M transition decatenation checkpoint	2.08E-04
GO:0031398	positive regulation of protein ubiquitination	2.25E-04
GO:0015992	proton transport	2.31E-04
GO:0000041	transition metal ion transport	2.32E-04
GO:0044267	cellular protein metabolic process	2.34E-04
GO:0035872	nucleotide-binding domain, leucine rich repeat containing receptor signaling pathway	2.35E-04
GO:0006402	mRNA catabolic process	2.45E-04
GO:0000395	mRNA 5'-splice site recognition	2.53E-04
GO:0009166	nucleotide catabolic process	2.60E-04

GO:0050789	regulation of biological process	2.61E-04
GO:0009117	nucleotide metabolic process	2.63E-04
GO:0016570	histone modification	2.71E-04
GO:0010608	posttranscriptional regulation of gene expression	2.80E-04
GO:0007254	JNK cascade	2.99E-04
GO:0006461	protein complex assembly	3.06E-04
GO:0035083	cilium axoneme assembly	3.18E-04
GO:0051488	activation of anaphase-promoting complex activity	3.20E-04
GO:1901292	nucleoside phosphate catabolic process	3.22E-04
GO:0065007	biological regulation	3.25E-04
GO:0016569	covalent chromatin modification	3.26E-04
GO:0006879	cellular iron ion homeostasis	3.29E-04
GO:0071900	regulation of protein serine/threonine kinase activity	3.31E-04
GO:0071456	cellular response to hypoxia	3.41E-04
GO:0036294	cellular response to decreased oxygen levels	3.41E-04
GO:0007010	cytoskeleton organization	3.68E-04
GO:0045087	innate immune response	3.87E-04
GO:0050790	regulation of catalytic activity	4.15E-04
GO:0000245	spliceosomal complex assembly	4.25E-04
GO:0022618	ribonucleoprotein complex assembly	4.30E-04
GO:0071453	cellular response to oxygen levels	4.48E-04
GO:0006952	defense response	4.48E-04
GO:0060255	regulation of macromolecule metabolic process	4.61E-04

GO:0006366	transcription from RNA polymerase II promoter	4.73E-04
GO:0019222	regulation of metabolic process	4.75E-04
GO:0034645	cellular macromolecule biosynthetic process	4.76E-04
GO:0002756	MyD88-independent toll-like receptor signaling pathway	4.96E-04
GO:0034138	toll-like receptor 3 signaling pathway	4.96E-04
GO:0035666	TRIF-dependent toll-like receptor signaling pathway	4.96E-04
GO:0050684	regulation of mRNA processing	4.99E-04
GO:0010458	exit from mitosis	5.09E-04
GO:0071822	protein complex subunit organization	5.17E-04
GO:0050852	T cell receptor signaling pathway	5.52E-04
GO:0009059	macromolecule biosynthetic process	5.69E-04
GO:0009893	positive regulation of metabolic process	5.76E-04
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	5.84E-04
GO:0050690	regulation of defense response to virus by virus	6.04E-04
GO:0006184	GTP catabolic process	6.08E-04
GO:1901185	negative regulation of ERBB signaling pathway	6.11E-04
GO:0042059	negative regulation of epidermal growth factor receptor signaling pathway	6.11E-04
GO:0055072	iron ion homeostasis	6.13E-04
GO:0055003	cardiac myofibril assembly	6.36E-04
GO:0044270	cellular nitrogen compound catabolic process	6.77E-04
GO:0006376	mRNA splice site selection	6.94E-04
GO:0046434	organophosphate catabolic process	7.03E-04
GO:0072659	protein localization to plasma membrane	7.34E-04

GO:0071826	ribonucleoprotein complex subunit organization	7.47E-04
GO:0034130	toll-like receptor 1 signaling pathway	7.47E-04
GO:0034134	toll-like receptor 2 signaling pathway	7.47E-04
GO:0006753	nucleoside phosphate metabolic process	7.48E-04
GO:0000187	activation of MAPK activity	7.53E-04
GO:0010604	positive regulation of macromolecule metabolic process	7.86E-04
GO:1901069	guanosine-containing compound catabolic process	7.92E-04
GO:0048619	embryonic hindgut morphogenesis	7.99E-04
GO:0006413	translational initiation	8.53E-04
GO:0032479	regulation of type I interferon production	8.80E-04
GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I	9.21E-04
GO:0010389	regulation of G2/M transition of mitotic cell cycle	9.39E-04
GO:0031396	regulation of protein ubiquitination	9.73E-04
GO:0043484	regulation of RNA splicing	9.98E-04

Appendix xiv - Overrepresented 1-2 h G.O. Cellular Component terms

GO Term	Description	P-value
GO:0032991	macromolecular complex	4.26E-22
GO:0005681	spliceosomal complex	5.87E-21
GO:0071013	catalytic step 2 spliceosome	5.24E-16
GO:0005654	nucleoplasm	1.63E-15
GO:0044428	nuclear part	7.73E-14
GO:0044424	intracellular part	1.45E-13
GO:0030529	ribonucleoprotein complex	1.79E-13
GO:0044422	organelle part	3.50E-13
GO:0005829	cytosol	3.53E-13
GO:0043234	protein complex	6.13E-13
GO:0044446	intracellular organelle part	1.51E-12
GO:0005634	nucleus	1.87E-10
GO:0044427	chromosomal part	1.26E-08
GO:0016607	nuclear speck	5.10E-08
GO:0043226	organelle	8.46E-08
GO:0043227	membrane-bounded organelle	9.39E-08
GO:0043229	intracellular organelle	1.09E-07
GO:0043231	intracellular membrane-bounded organelle	1.12E-07
GO:0042555	MCM complex	9.03E-07
GO:0005874	microtubule	1.01E-06
GO:0000808	origin recognition complex	1.94E-06
GO:0005664	nuclear origin of replication recognition complex	1.94E-06
GO:0044454	nuclear chromosome part	2.29E-06
GO:0016604	nuclear body	3.58E-06
GO:0000786	nucleosome	1.14E-05
GO:0005819	spindle	2.08E-05
GO:0044430	cytoskeletal part	2.29E-05
GO:0005680	anaphase-promoting complex	3.29E-05
GO:0005665	DNA-directed RNA polymerase II, core complex	3.48E-05
GO:0015630	microtubule cytoskeleton	3.50E-05
GO:0042470	melanosome	4.82E-05
GO:0048770	pigment granule	4.82E-05
GO:0005832	chaperonin-containing T-complex	5.58E-05
GO:0044444	cytoplasmic part	7.42E-05
GO:0032993	protein-DNA complex	9.30E-05
GO:0030666	endocytic vesicle membrane	9.99E-05
GO:0016460	myosin II complex	1.08E-04

GO:0000151	ubiquitin ligase complex	1.12E-04
GO:0031461	cullin-RING ubiquitin ligase complex	1.51E-04
GO:0044451	nucleoplasm part	1.69E-04
GO:0043228	non-membrane-bounded organelle	1.77E-04
GO:0043232	intracellular non-membrane-bounded organelle	1.77E-04
GO:0000152	nuclear ubiquitin ligase complex	1.78E-04
GO:0005672	transcription factor TFIIA complex	2.18E-04
GO:0022624	proteasome accessory complex	2.67E-04
GO:0005694	chromosome	2.85E-04
GO:0000796	condensin complex	4.08E-04
GO:0045254	pyruvate dehydrogenase complex	4.57E-04
GO:0005859	muscle myosin complex	5.20E-04
GO:0044464	cell part	6.60E-04
GO:0044445	cytosolic part	7.77E-04
GO:0005689	U12-type spliceosomal complex	8.88E-04

Appendix xv – Overrepresented 3-5 h G.O. Process terms

GO Term	Description	P-value
GO:0016071	mRNA metabolic process	1.65E-25
GO:0008380	RNA splicing	4.35E-24
GO:0000375	RNA splicing, via transesterification reactions	7.69E-24
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	9.68E-24
GO:0000398	mRNA splicing, via spliceosome	9.68E-24
GO:0006397	mRNA processing	5.68E-22
GO:0034622	cellular macromolecular complex assembly	5.36E-14
GO:0000278	mitotic cell cycle	7.36E-14
GO:0006396	RNA processing	1.02E-13
GO:0034621	cellular macromolecular complex subunit organization	2.44E-13
GO:0007049	cell cycle	1.86E-12
GO:0016032	viral reproduction	4.35E-12
GO:0022402	cell cycle process	6.10E-12
GO:0006139	nucleobase-containing compound metabolic process	1.78E-11
GO:0006458	'de novo' protein folding	2.72E-11
GO:0071842	cellular component organization at cellular level	3.31E-11
GO:0051084	'de novo' posttranslational protein folding	4.50E-11
GO:0000216	M/G1 transition of mitotic cell cycle	7.26E-11
GO:0022403	cell cycle phase	1.49E-10
GO:0071841	cellular component organization or biogenesis at cellular level	1.59E-10
GO:0016043	cellular component organization	2.30E-10

GO:0044238	primary metabolic process	2.38E-10
GO:0034641	cellular nitrogen compound metabolic process	2.76E-10
GO:0031145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	2.84E-10
GO:0046483	heterocycle metabolic process	3.04E-10
GO:0043170	macromolecule metabolic process	4.00E-10
GO:0006725	cellular aromatic compound metabolic process	4.93E-10
GO:0010467	gene expression	5.20E-10
GO:0006807	nitrogen compound metabolic process	5.98E-10
GO:0006259	DNA metabolic process	6.04E-10
GO:0000075	cell cycle checkpoint	6.52E-10
GO:0071840	cellular component organization or biogenesis	9.15E-10
GO:0048002	antigen processing and presentation of peptide antigen	9.76E-10
GO:0071844	cellular component assembly at cellular level	9.78E-10
GO:0000082	G1/S transition of mitotic cell cycle	1.15E-09
GO:0006996	organelle organization	1.36E-09
GO:0051437	positive regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	1.46E-09
GO:0002478	antigen processing and presentation of exogenous peptide antigen	1.50E-09
GO:0006405	RNA export from nucleus	1.55E-09
GO:0044260	cellular macromolecule metabolic process	1.58E-09
GO:0006406	mRNA export from nucleus	1.59E-09
GO:0051443	positive regulation of ubiquitin-protein ligase activity	2.21E-09
GO:0019884	antigen processing and presentation of exogenous antigen	2.67E-09
GO:0051439	regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	2.74E-09

GO:0051168	nuclear export	2.94E-09
GO:0051028	mRNA transport	2.95E-09
GO:0051351	positive regulation of ligase activity	4.52E-09
GO:0090304	nucleic acid metabolic process	4.73E-09
GO:0019882	antigen processing and presentation	6.05E-09
GO:0051436	negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	6.83E-09
GO:1901360	organic cyclic compound metabolic process	7.80E-09
GO:0051438	regulation of ubiquitin-protein ligase activity	8.05E-09
GO:0051320	S phase	8.38E-09
GO:0000084	S phase of mitotic cell cycle	8.38E-09
GO:0046907	intracellular transport	8.39E-09
GO:0051352	negative regulation of ligase activity	1.03E-08
GO:0051444	negative regulation of ubiquitin-protein ligase activity	1.03E-08
GO:0051704	multi-organism process	1.26E-08
GO:0051340	regulation of ligase activity	1.59E-08
GO:0031398	positive regulation of protein ubiquitination	1.75E-08
GO:0071156	regulation of cell cycle arrest	2.23E-08
GO:0071704	organic substance metabolic process	2.84E-08
GO:0050658	RNA transport	2.94E-08
GO:0050657	nucleic acid transport	2.94E-08
GO:0051236	establishment of RNA localization	2.94E-08
GO:0008152	metabolic process	3.06E-08
GO:0000209	protein polyubiquitination	3.60E-08

GO:0070979	protein K11-linked ubiquitination	3.71E-08
GO:0044237	cellular metabolic process	3.79E-08
GO:0065003	macromolecular complex assembly	5.79E-08
GO:0044265	cellular macromolecule catabolic process	5.96E-08
GO:0010564	regulation of cell cycle process	6.69E-08
GO:0009987	cellular process	7.41E-08
GO:0043933	macromolecular complex subunit organization	7.77E-08
GO:0051649	establishment of localization in cell	9.58E-08
GO:0051258	protein polymerization	1.08E-07
GO:0009057	macromolecule catabolic process	1.93E-07
GO:0044710	single-organism metabolic process	2.65E-07
GO:0006260	DNA replication	2.80E-07
GO:0010498	proteasomal protein catabolic process	3.33E-07
GO:0043161	proteasomal ubiquitin-dependent protein catabolic process	3.33E-07
GO:0075733	intracellular transport of viral material	4.43E-07
GO:0046788	egress of virus within host cell	4.43E-07
GO:0006376	mRNA splice site selection	4.96E-07
GO:0022607	cellular component assembly	6.73E-07
GO:0009056	catabolic process	8.66E-07
GO:0015931	nucleobase-containing compound transport	9.00E-07
GO:0031397	negative regulation of protein ubiquitination	1.20E-06
GO:0000389	mRNA 3'-splice site recognition	1.25E-06
GO:0006913	nucleocytoplasmic transport	1.69E-06

GO:0006457	protein folding	1.75E-06
GO:0022415	viral reproductive process	1.88E-06
GO:0051169	nuclear transport	2.39E-06
GO:0044248	cellular catabolic process	2.45E-06
GO:0009144	purine nucleoside triphosphate metabolic process	2.50E-06
GO:0051701	interaction with host	2.56E-06
GO:0009141	nucleoside triphosphate metabolic process	2.59E-06
GO:0016070	RNA metabolic process	2.96E-06
GO:0009205	purine ribonucleoside triphosphate metabolic process	4.36E-06
GO:0061418	regulation of transcription from RNA polymerase II promoter in response to hypoxia	5.08E-06
GO:0044419	interspecies interaction between organisms	5.48E-06
GO:0031124	mRNA 3'-end processing	5.52E-06
GO:0048024	regulation of mRNA splicing, via spliceosome	5.94E-06
GO:0007220	Notch receptor processing	5.96E-06
GO:0031396	regulation of protein ubiquitination	6.97E-06
GO:0070423	nucleotide-binding oligomerization domain containing signaling pathway	7.63E-06
GO:0009199	ribonucleoside triphosphate metabolic process	7.86E-06
GO:0009259	ribonucleotide metabolic process	8.47E-06
GO:0006368	transcription elongation from RNA polymerase II promoter	9.09E-06
GO:0071824	protein-DNA complex subunit organization	9.47E-06
GO:0007017	microtubule-based process	9.71E-06
GO:0022618	ribonucleoprotein complex assembly	9.96E-06
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II	1.09E-05

GO:0043618	regulation of transcription from RNA polymerase II promoter in response to stress	1.10E-05
GO:0007249	I-kappaB kinase/NF-kappaB cascade	1.10E-05
GO:0016044	cellular membrane organization	1.12E-05
GO:0061024	membrane organization	1.18E-05
GO:0002479	antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	1.29E-05
GO:0042590	antigen processing and presentation of exogenous peptide antigen via MHC class I	1.29E-05
GO:0032480	negative regulation of type I interferon production	1.30E-05
GO:0002753	cytoplasmic pattern recognition receptor signaling pathway	1.30E-05
GO:0000080	G1 phase of mitotic cell cycle	1.31E-05
GO:0051603	proteolysis involved in cellular protein catabolic process	1.42E-05
GO:0000280	nuclear division	1.46E-05
GO:0007067	mitosis	1.46E-05
GO:0009150	purine ribonucleotide metabolic process	1.49E-05
GO:0050684	regulation of mRNA processing	1.50E-05
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	1.52E-05
GO:0043623	cellular protein complex assembly	1.57E-05
GO:0051318	G1 phase	1.58E-05
GO:0006511	ubiquitin-dependent protein catabolic process	1.71E-05
GO:0071826	ribonucleoprotein complex subunit organization	1.78E-05
GO:0043900	regulation of multi-organism process	1.81E-05
GO:0031123	RNA 3'-end processing	1.81E-05
GO:0009117	nucleotide metabolic process	1.91E-05
GO:0006325	chromatin organization	1.92E-05

GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II	2.14E-05
GO:0002495	antigen processing and presentation of peptide antigen via MHC class II	2.14E-05
GO:0043620	regulation of DNA-dependent transcription in response to stress	2.21E-05
GO:0019941	modification-dependent protein catabolic process	2.24E-05
GO:0051276	chromosome organization	2.27E-05
GO:0065004	protein-DNA complex assembly	2.33E-05
GO:0044703	multi-organism reproductive process	2.42E-05
GO:0010604	positive regulation of macromolecule metabolic process	2.66E-05
GO:0017015	regulation of transforming growth factor beta receptor signaling pathway	2.66E-05
GO:0043632	modification-dependent macromolecule catabolic process	2.76E-05
GO:0034728	nucleosome organization	3.12E-05
GO:0006334	nucleosome assembly	3.13E-05
GO:0006281	DNA repair	3.16E-05
GO:0035872	nucleotide-binding domain, leucine rich repeat containing receptor signaling pathway	3.24E-05
GO:0048285	organelle fission	3.68E-05
GO:1901185	negative regulation of ERBB signaling pathway	3.74E-05
GO:0042059	negative regulation of epidermal growth factor receptor signaling pathway	3.74E-05
GO:0055086	nucleobase-containing small molecule metabolic process	4.11E-05
GO:0071456	cellular response to hypoxia	4.13E-05
GO:0036294	cellular response to decreased oxygen levels	4.13E-05
GO:0009893	positive regulation of metabolic process	4.36E-05
GO:0015988	energy coupled proton transmembrane transport, against electrochemical gradient	4.38E-05
GO:0015991	ATP hydrolysis coupled proton transport	4.38E-05

GO:0043484	regulation of RNA splicing	4.43E-05
GO:0015682	ferric iron transport	4.57E-05
GO:0033572	transferrin transport	4.57E-05
GO:0071453	cellular response to oxygen levels	4.71E-05
GO:0007254	JNK cascade	4.71E-05
GO:0030512	negative regulation of transforming growth factor beta receptor signaling pathway	4.79E-05
GO:0050690	regulation of defense response to virus by virus	5.29E-05
GO:0006977	DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest	5.29E-05
GO:0072431	signal transduction involved in mitotic cell cycle G1/S transition DNA damage checkpoint	5.29E-05
GO:0072413	signal transduction involved in mitotic cell cycle checkpoint	5.29E-05
GO:0072474	signal transduction involved in mitotic cell cycle G1/S checkpoint	5.29E-05
GO:0072404	signal transduction involved in G1/S transition checkpoint	5.29E-05
GO:0006096	glycolysis	5.81E-05
GO:0032268	regulation of cellular protein metabolic process	5.84E-05
GO:0006270	DNA replication initiation	5.86E-05
GO:0008150	biological_process	6.04E-05
GO:0016567	protein ubiquitination	6.10E-05
GO:0006369	termination of RNA polymerase II transcription	6.18E-05
GO:0051726	regulation of cell cycle	6.40E-05
GO:0070507	regulation of microtubule cytoskeleton organization	6.86E-05
GO:0072395	signal transduction involved in cell cycle checkpoint	7.11E-05
GO:0072422	signal transduction involved in DNA damage checkpoint	7.11E-05
GO:0072401	signal transduction involved in DNA integrity checkpoint	7.11E-05

GO:0050434	positive regulation of viral transcription	7.49E-05
GO:0034138	toll-like receptor 3 signaling pathway	7.77E-05
GO:1901184	regulation of ERBB signaling pathway	7.77E-05
GO:0050852	T cell receptor signaling pathway	7.77E-05
GO:0042058	regulation of epidermal growth factor receptor signaling pathway	7.77E-05
GO:0035666	TRIF-dependent toll-like receptor signaling pathway	7.77E-05
GO:0035083	cilium axoneme assembly	8.24E-05
GO:0002756	MyD88-independent toll-like receptor signaling pathway	8.46E-05
GO:0034130	toll-like receptor 1 signaling pathway	9.30E-05
GO:0034134	toll-like receptor 2 signaling pathway	9.30E-05
GO:0006163	purine nucleotide metabolic process	9.39E-05
GO:0090068	positive regulation of cell cycle process	9.63E-05
GO:0090382	phagosome maturation	9.97E-05
GO:0019048	virus-host interaction	9.97E-05
GO:0050792	regulation of viral reproduction	1.01E-04
GO:0007018	microtubule-based movement	1.05E-04
GO:0006753	nucleoside phosphate metabolic process	1.07E-04
GO:0006417	regulation of translation	1.07E-04
GO:0032479	regulation of type I interferon production	1.13E-04
GO:0032886	regulation of microtubule-based process	1.18E-04
GO:0000395	mRNA 5'-splice site recognition	1.21E-04
GO:0050851	antigen receptor-mediated signaling pathway	1.31E-04
GO:0002755	MyD88-dependent toll-like receptor signaling pathway	1.46E-04

GO:0090092	regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	1.49E-04
GO:0009203	ribonucleoside triphosphate catabolic process	1.50E-04
GO:0009207	purine ribonucleoside triphosphate catabolic process	1.50E-04
GO:0010994	free ubiquitin chain polymerization	1.53E-04
GO:0002429	immune response-activating cell surface receptor signaling pathway	1.58E-04
GO:0032446	protein modification by small protein conjugation	1.58E-04
GO:0001578	microtubule bundle formation	1.65E-04
GO:0008063	Toll signaling pathway	1.73E-04
GO:0051301	cell division	1.74E-04
GO:0009143	nucleoside triphosphate catabolic process	1.75E-04
GO:0006366	transcription from RNA polymerase II promoter	1.80E-04
GO:0034142	toll-like receptor 4 signaling pathway	1.86E-04
GO:0006508	proteolysis	1.89E-04
GO:0019538	protein metabolic process	1.90E-04
GO:0051403	stress-activated MAPK cascade	1.92E-04
GO:0032270	positive regulation of cellular protein metabolic process	1.95E-04
GO:0090101	negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway	2.02E-04
GO:0009146	purine nucleoside triphosphate catabolic process	2.03E-04
GO:0009154	purine ribonucleotide catabolic process	2.03E-04
GO:0031098	stress-activated protein kinase signaling cascade	2.05E-04
GO:0048524	positive regulation of viral reproduction	2.10E-04
GO:0031325	positive regulation of cellular metabolic process	2.19E-04
GO:0002768	immune response-regulating cell surface receptor signaling pathway	2.21E-04

GO:0051092	positive regulation of NF-kappaB transcription factor activity	2.21E-04
GO:0009261	ribonucleotide catabolic process	2.42E-04
GO:0008624	induction of apoptosis by extracellular signals	2.55E-04
GO:0008543	fibroblast growth factor receptor signaling pathway	2.55E-04
GO:0051246	regulation of protein metabolic process	2.58E-04
GO:0006370	7-methylguanosine mRNA capping	2.62E-04
GO:0006271	DNA strand elongation involved in DNA replication	2.67E-04
GO:0006974	response to DNA damage stimulus	2.79E-04
GO:0072521	purine-containing compound metabolic process	2.82E-04
GO:0006401	RNA catabolic process	2.85E-04
GO:0071900	regulation of protein serine/threonine kinase activity	2.85E-04
GO:0045087	innate immune response	3.01E-04
GO:0044267	cellular protein metabolic process	3.11E-04
GO:0002224	toll-like receptor signaling pathway	3.12E-04
GO:0071158	positive regulation of cell cycle arrest	3.12E-04
GO:0043086	negative regulation of catalytic activity	3.53E-04
GO:0060255	regulation of macromolecule metabolic process	3.55E-04
GO:0055003	cardiac myofibril assembly	3.57E-04
GO:0006911	phagocytosis, engulfment	3.68E-04
GO:0046782	regulation of viral transcription	3.77E-04
GO:0036260	RNA capping	3.80E-04
GO:0009452	7-methylguanosine RNA capping	3.80E-04
GO:0048522	positive regulation of cellular process	3.91E-04

GO:0043902	positive regulation of multi-organism process	3.98E-04
GO:0033238	regulation of cellular amine metabolic process	4.00E-04
GO:0035082	axoneme assembly	4.04E-04
GO:0071774	response to fibroblast growth factor stimulus	4.07E-04
GO:0002221	pattern recognition receptor signaling pathway	4.07E-04
GO:0007179	transforming growth factor beta receptor signaling pathway	4.07E-04
GO:0044344	cellular response to fibroblast growth factor stimulus	4.07E-04
GO:0031323	regulation of cellular metabolic process	4.13E-04
GO:0006354	DNA-dependent transcription, elongation	4.18E-04
GO:0002758	innate immune response-activating signal transduction	4.40E-04
GO:0010458	exit from mitosis	4.42E-04
GO:0051247	positive regulation of protein metabolic process	4.53E-04
GO:0002218	activation of innate immune response	4.63E-04
GO:0001818	negative regulation of cytokine production	4.63E-04
GO:0019222	regulation of metabolic process	4.77E-04
GO:0030330	DNA damage response, signal transduction by p53 class mediator	4.77E-04
GO:0007173	epidermal growth factor receptor signaling pathway	4.77E-04
GO:0010608	posttranscriptional regulation of gene expression	4.81E-04
GO:0038127	ERBB signaling pathway	4.95E-04
GO:0043652	engulfment of apoptotic cell	5.37E-04
GO:0000187	activation of MAPK activity	5.39E-04
GO:0016052	carbohydrate catabolic process	5.45E-04
GO:0048025	negative regulation of mRNA splicing, via spliceosome	5.47E-04

GO:0046039	GTP metabolic process	5.48E-04
GO:0030239	myofibril assembly	5.79E-04
GO:0006195	purine nucleotide catabolic process	5.88E-04
GO:0022616	DNA strand elongation	5.89E-04
GO:0007219	Notch signaling pathway	6.13E-04
GO:0010324	membrane invagination	6.34E-04
GO:0034655	nucleobase-containing compound catabolic process	6.37E-04
GO:0072523	purine-containing compound catabolic process	6.39E-04
GO:0033554	cellular response to stress	6.75E-04
GO:0000381	regulation of alternative mRNA splicing, via spliceosome	6.80E-04
GO:0042770	signal transduction in response to DNA damage	6.86E-04
GO:0006826	iron ion transport	6.91E-04
GO:0045089	positive regulation of innate immune response	7.06E-04
GO:0031400	negative regulation of protein modification process	7.32E-04
GO:0046034	ATP metabolic process	7.42E-04
GO:0072331	signal transduction by p53 class mediator	7.91E-04
GO:0006892	post-Golgi vesicle-mediated transport	8.11E-04
GO:0043123	positive regulation of I-kappaB kinase/NF-kappaB cascade	8.26E-04
GO:0048518	positive regulation of biological process	8.36E-04
GO:0000165	MAPK cascade	8.57E-04
GO:0007178	transmembrane receptor protein serine/threonine kinase signaling pathway	8.71E-04
GO:0000226	microtubule cytoskeleton organization	8.81E-04
GO:0000041	transition metal ion transport	8.88E-04

GO:0060972	left/right pattern formation	9.66E-04
GO:0031113	regulation of microtubule polymerization	9.74E-04
GO:0006310	DNA recombination	9.97E-04

Appendix xvi - Overrepresented 3-5 h G.O. Function terms

GO Term	Description	P-value
GO:0016817	hydrolase activity, acting on acid anhydrides	8.04E-14
GO:0016462	pyrophosphatase activity	1.19E-13
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	1.55E-13
GO:0017111	nucleoside-triphosphatase activity	2.11E-13
GO:0036094	small molecule binding	1.41E-09
GO:0000166	nucleotide binding	1.50E-09
GO:1901265	nucleoside phosphate binding	1.57E-09
GO:0005515	protein binding	1.04E-07
GO:0016887	ATPase activity	1.74E-07
GO:1901363	heterocyclic compound binding	7.59E-07
GO:0097159	organic cyclic compound binding	7.59E-07
GO:0003723	RNA binding	1.53E-06
GO:0035639	purine ribonucleoside triphosphate binding	2.92E-05
GO:0036002	pre-mRNA binding	2.96E-05
GO:0017076	purine nucleotide binding	3.71E-05
GO:0032555	purine ribonucleotide binding	4.36E-05
GO:0032553	ribonucleotide binding	4.36E-05
GO:0042623	ATPase activity, coupled	4.97E-05
GO:0004738	pyruvate dehydrogenase activity	5.39E-05
GO:0004739	pyruvate dehydrogenase (acetyl-transferring) activity	5.39E-05
GO:0016787	hydrolase activity	5.43E-05

GO:0003924	GTPase activity	7.13E-05
GO:0005488	binding	9.37E-05
GO:0019829	cation-transporting ATPase activity	1.33E-04
GO:0046961	proton-transporting ATPase activity, rotational mechanism	1.93E-04
GO:0051082	unfolded protein binding	2.36E-04
GO:0003676	nucleic acid binding	3.79E-04
GO:0005524	ATP binding	6.20E-04
GO:0030554	adenyl nucleotide binding	6.28E-04
GO:0016624	oxidoreductase activity, acting on the aldehyde or oxo group of donors, disulfide as acceptor	6.83E-04
GO:0008327	methyl-CpG binding	7.80E-04
GO:0032559	adenyl ribonucleotide binding	8.64E-04
GO:0004386	helicase activity	9.05E-04

Appendix xvii - Overrepresented 3-5 h G.O. Cellular Component terms

GO Term	Description	P-value
GO:0005681	spliceosomal complex	1.03E-22
GO:0032991	macromolecular complex	2.01E-21
GO:0044422	organelle part	4.31E-20
GO:0044446	intracellular organelle part	8.27E-20
GO:0071013	catalytic step 2 spliceosome	1.74E-17
GO:0044424	intracellular part	9.13E-17
GO:0005654	nucleoplasm	9.82E-17
GO:0030529	ribonucleoprotein complex	5.48E-14
GO:0005829	cytosol	7.29E-14
GO:0044428	nuclear part	2.06E-13
GO:0016607	nuclear speck	6.47E-12
GO:0005874	microtubule	1.35E-11
GO:0043234	protein complex	2.48E-11
GO:0043229	intracellular organelle	8.08E-11
GO:0043226	organelle	1.17E-10
GO:0016604	nuclear body	9.62E-09
GO:0044430	cytoskeletal part	2.51E-08
GO:0005634	nucleus	2.44E-07
GO:0043228	non-membrane-bounded organelle	2.60E-07
GO:0043232	intracellular non-membrane-bounded organelle	2.60E-07
GO:0044444	cytoplasmic part	8.29E-07
GO:0022624	proteasome accessory complex	8.91E-07
GO:0015630	microtubule cytoskeleton	2.34E-06
GO:0044464	cell part	3.24E-06
GO:0043231	intracellular membrane-bounded organelle	3.30E-06
GO:0000786	nucleosome	3.52E-06
GO:0043227	membrane-bounded organelle	3.82E-06
GO:0005680	anaphase-promoting complex	1.13E-05
GO:0044451	nucleoplasm part	1.80E-05
GO:0032993	protein-DNA complex	1.92E-05
GO:0005813	centrosome	2.26E-05
GO:0044427	chromosomal part	2.55E-05
GO:0030666	endocytic vesicle membrane	3.69E-05
GO:0005856	cytoskeleton	6.27E-05
GO:0000152	nuclear ubiquitin ligase complex	6.59E-05
GO:0005832	chaperonin-containing T-complex	7.03E-05
GO:0042470	melanosome	8.36E-05
GO:0048770	pigment granule	8.36E-05

Appendix xviii – Overrepresented 6-8 h G.O. Process terms

GO Term	Description	P-value
GO:0007017	microtubule-based process	1.05E-12
GO:0007018	microtubule-based movement	5.41E-11
GO:0006996	organelle organization	5.21E-08
GO:0022402	cell cycle process	7.01E-08
GO:0000278	mitotic cell cycle	4.63E-07
GO:0071842	cellular component organization at cellular level	1.06E-06
GO:0001539	ciliary or flagellar motility	1.18E-06
GO:0000086	G2/M transition of mitotic cell cycle	1.57E-06
GO:0034622	cellular macromolecular complex assembly	1.79E-06
GO:0071844	cellular component assembly at cellular level	2.93E-06
GO:0071841	cellular component organization or biogenesis at cellular level	3.06E-06
GO:0006091	generation of precursor metabolites and energy	3.52E-06
GO:0016043	cellular component organization	5.41E-06
GO:0051258	protein polymerization	6.37E-06
GO:0010766	negative regulation of sodium ion transport	1.24E-05
GO:0071840	cellular component organization or biogenesis	1.41E-05
GO:0034621	cellular macromolecular complex subunit organization	1.47E-05
GO:0034199	activation of protein kinase A activity	2.16E-05
GO:0009141	nucleoside triphosphate metabolic process	2.20E-05
GO:0006006	glucose metabolic process	2.26E-05
GO:000692	cellular component movement	2.46E-

8		05
GO:0000375	RNA splicing, via transesterification reactions	3.16E-05
GO:0043623	cellular protein complex assembly	3.36E-05
GO:0009206	purine ribonucleoside triphosphate biosynthetic process	4.07E-05
GO:0070286	axonemal dynein complex assembly	4.96E-05
GO:0046907	intracellular transport	5.38E-05
GO:0009145	purine nucleoside triphosphate biosynthetic process	5.60E-05
GO:0051649	establishment of localization in cell	6.04E-05
GO:0055086	nucleobase-containing small molecule metabolic process	6.05E-05
GO:0009144	purine nucleoside triphosphate metabolic process	6.27E-05
GO:0030388	fructose 1,6-bisphosphate metabolic process	6.54E-05
GO:0009260	ribonucleotide biosynthetic process	6.71E-05
GO:0065003	macromolecular complex assembly	7.03E-05
GO:0016266	O-glycan processing	7.25E-05
GO:0070979	protein K11-linked ubiquitination	7.30E-05
GO:0007010	cytoskeleton organization	8.34E-05
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	9.21E-05
GO:0000398	mRNA splicing, via spliceosome	9.21E-05
GO:0006309	apoptotic DNA fragmentation	1.05E-04
GO:0010564	regulation of cell cycle process	1.14E-04
GO:0000209	protein polyubiquitination	1.23E-04
GO:0006096	glycolysis	1.26E-04
GO:0009220	pyrimidine ribonucleotide biosynthetic process	1.33E-04

GO:0005996	monosaccharide metabolic process	1.34E-04
GO:0071377	cellular response to glucagon stimulus	1.37E-04
GO:0043266	regulation of potassium ion transport	1.45E-04
GO:0003341	cilium movement	1.49E-04
GO:1901016	regulation of potassium ion transmembrane transporter activity	1.52E-04
GO:1901379	regulation of potassium ion transmembrane transport	1.52E-04
GO:0000075	cell cycle checkpoint	1.69E-04
GO:0000737	DNA catabolic process, endonucleolytic	1.72E-04
GO:0022403	cell cycle phase	1.75E-04
GO:0019318	hexose metabolic process	1.81E-04
GO:0000226	microtubule cytoskeleton organization	1.83E-04
GO:0009201	ribonucleoside triphosphate biosynthetic process	1.89E-04
GO:0043933	macromolecular complex subunit organization	2.09E-04
GO:0060286	flagellar cell motility	2.10E-04
GO:0060285	ciliary cell motility	2.10E-04
GO:0006493	protein O-linked glycosylation	2.24E-04
GO:0009628	response to abiotic stimulus	2.41E-04
GO:0009259	ribonucleotide metabolic process	2.44E-04
GO:0046034	ATP metabolic process	2.56E-04
GO:0033146	regulation of intracellular estrogen receptor signaling pathway	2.63E-04
GO:0006112	energy reserve metabolic process	2.65E-04
GO:0009205	purine ribonucleoside triphosphate metabolic process	2.67E-04
GO:0016052	carbohydrate catabolic process	2.67E-04

GO:0009218	pyrimidine ribonucleotide metabolic process	2.70E-04
GO:0051234	establishment of localization	2.85E-04
GO:0060088	auditory receptor cell stereocilium organization	3.25E-04
GO:0071156	regulation of cell cycle arrest	3.41E-04
GO:0022607	cellular component assembly	3.51E-04
GO:0009152	purine ribonucleotide biosynthetic process	3.55E-04
GO:0009150	purine ribonucleotide metabolic process	3.66E-04
GO:0009142	nucleoside triphosphate biosynthetic process	3.79E-04
GO:1901135	carbohydrate derivative metabolic process	3.81E-04
GO:0009719	response to endogenous stimulus	3.98E-04
GO:0008543	fibroblast growth factor receptor signaling pathway	3.99E-04
GO:0044702	single organism reproductive process	4.08E-04
GO:0033762	response to glucagon stimulus	4.20E-04
GO:0009056	catabolic process	4.23E-04
GO:0006810	transport	4.31E-04
GO:0072522	purine-containing compound biosynthetic process	4.55E-04
GO:0006753	nucleoside phosphate metabolic process	4.57E-04
GO:0007049	cell cycle	4.97E-04
GO:0006221	pyrimidine nucleotide biosynthetic process	5.07E-04
GO:0075733	intracellular transport of viral material	5.30E-04
GO:0046788	egress of virus within host cell	5.30E-04
GO:0009199	ribonucleoside triphosphate metabolic process	5.34E-04
GO:0046121	deoxyribonucleoside catabolic process	5.96E-04

GO:0009120	deoxyribonucleoside metabolic process	5.96E-04
GO:0043967	histone H4 acetylation	5.98E-04
GO:0090288	negative regulation of cellular response to growth factor stimulus	5.99E-04
GO:0015980	energy derivation by oxidation of organic compounds	6.17E-04
GO:0006754	ATP biosynthetic process	6.27E-04
GO:0042752	regulation of circadian rhythm	6.34E-04
GO:0006833	water transport	6.37E-04
GO:0006007	glucose catabolic process	6.49E-04
GO:0000281	cytokinesis after mitosis	6.73E-04
GO:0006094	gluconeogenesis	7.04E-04
GO:0008380	RNA splicing	7.38E-04
GO:0034654	nucleobase-containing compound biosynthetic process	7.71E-04
GO:0033148	positive regulation of intracellular estrogen receptor signaling pathway	7.79E-04
GO:0006334	nucleosome assembly	7.92E-04
GO:0072521	purine-containing compound metabolic process	8.20E-04
GO:0048521	negative regulation of behavior	8.27E-04
GO:0071774	response to fibroblast growth factor stimulus	8.35E-04
GO:0044344	cellular response to fibroblast growth factor stimulus	8.35E-04
GO:0031116	positive regulation of microtubule polymerization	8.46E-04
GO:0071824	protein-DNA complex subunit organization	8.68E-04
GO:0019319	hexose biosynthetic process	8.83E-04
GO:0008039	synaptic target recognition	8.95E-04
GO:0033205	cell cycle cytokinesis	8.98E-04

GO:0043484	regulation of RNA splicing	9.10E-04
GO:0000280	nuclear division	9.17E-04
GO:0007067	mitosis	9.17E-04
GO:0046365	monosaccharide catabolic process	9.54E-04
GO:0042044	fluid transport	9.55E-04
GO:0006164	purine nucleotide biosynthetic process	9.58E-04

Appendix xix - Overrepresented 6-8 h G.O. Function terms

GO Term	Description	P-value
GO:0017111	nucleoside-triphosphatase activity	5.34E-13
GO:0016817	hydrolase activity, acting on acid anhydrides	1.72E-12
GO:0016462	pyrophosphatase activity	2.85E-12
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	3.76E-12
GO:0003774	motor activity	1.32E-08
GO:0019829	cation-transporting ATPase activity	7.07E-07
GO:0036094	small molecule binding	9.53E-07
GO:0000166	nucleotide binding	1.95E-06
GO:1901265	nucleoside phosphate binding	2.07E-06
GO:0016887	ATPase activity	4.48E-06
GO:0030197	extracellular matrix constituent, lubricant activity	9.78E-06
GO:0005388	calcium-transporting ATPase activity	1.19E-05
GO:0003777	microtubule motor activity	2.36E-05
GO:0008603	cAMP-dependent protein kinase regulator activity	2.57E-05
GO:0017076	purine nucleotide binding	2.74E-05
GO:0032555	purine ribonucleotide binding	4.66E-05
GO:0032553	ribonucleotide binding	4.66E-05
GO:0030554	adenyl nucleotide binding	4.93E-05
GO:0019871	sodium channel inhibitor activity	7.30E-05
GO:0042625	ATPase activity, coupled to transmembrane movement of ions	9.14E-05
GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	9.16E-05

GO:0035639	purine ribonucleoside triphosphate binding	9.38E-05
GO:0032559	adenyl ribonucleotide binding	9.41E-05
GO:0004332	fructose-bisphosphate aldolase activity	1.17E-04
GO:0042623	ATPase activity, coupled	1.94E-04
GO:0005524	ATP binding	1.94E-04
GO:0003924	GTPase activity	2.48E-04
GO:0005198	structural molecule activity	3.02E-04
GO:0070061	fructose binding	3.75E-04
GO:0016820	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	5.34E-04
GO:0008184	glycogen phosphorylase activity	5.84E-04
GO:0005525	GTP binding	6.10E-04
GO:0032561	guanyl ribonucleotide binding	7.08E-04
GO:0019001	guanyl nucleotide binding	7.08E-04
GO:0019904	protein domain specific binding	8.37E-04
GO:0043394	proteoglycan binding	8.91E-04
GO:0015399	primary active transmembrane transporter activity	9.06E-04
GO:0015405	P-P-bond-hydrolysis-driven transmembrane transporter activity	9.06E-04
GO:0016830	carbon-carbon lyase activity	9.35E-04
GO:0019899	enzyme binding	9.36E-04
GO:0051082	unfolded protein binding	9.69E-04

Appendix xx - Overrepresented 6-8 h G.O. Cellular Component terms

GO Term	Description	P-value
GO:0044422	organelle part	5.67E-08

GO:0044430	cytoskeletal part	7.67E-08
GO:0044446	intracellular organelle part	9.96E-08
GO:0005930	axoneme	4.20E-07
GO:0005796	Golgi lumen	5.93E-07
GO:0005952	cAMP-dependent protein kinase complex	8.40E-07
GO:0005815	microtubule organizing center	1.98E-06
GO:0044463	cell projection part	2.21E-06
GO:0005874	microtubule	3.48E-06
GO:0043228	non-membrane-bounded organelle	3.60E-06
GO:0043232	intracellular non-membrane-bounded organelle	3.60E-06
GO:0015630	microtubule cytoskeleton	5.52E-06
GO:0005813	centrosome	9.42E-06
GO:0030286	dynein complex	1.57E-05
GO:0005929	cilium	2.20E-05
GO:0044424	intracellular part	2.62E-05
GO:0005654	nucleoplasm	5.24E-05
GO:0000786	nucleosome	5.92E-05
GO:0044447	axoneme part	8.21E-05

Appendix xxi – Overrepresented 1-2 h G.O. Process terms (*A. thaliana*)

GO Term	Description	P-value
GO:0034728	nucleosome organization	2.44E-14
GO:0071824	protein-DNA complex subunit organization	2.44E-14
GO:0006334	nucleosome assembly	2.44E-14
GO:0065004	protein-DNA complex assembly	2.44E-14
GO:0051276	chromosome organization	4.41E-11
GO:0006996	organelle organization	1.97E-10
GO:0006325	chromatin organization	1.64E-09
GO:0007010	cytoskeleton organization	1.34E-08
GO:0051567	histone H3-K9 methylation	6.51E-08
GO:0016569	covalent chromatin modification	4.40E-07
GO:0016568	chromatin modification	4.40E-07
GO:0016570	histone modification	9.87E-07
GO:0034968	histone lysine methylation	1.07E-06
GO:0016043	cellular component organization	1.10E-06
GO:0022402	cell cycle process	1.27E-06
GO:0000910	cytokinesis	1.61E-06
GO:0022403	cell cycle phase	1.72E-06
GO:0040029	regulation of gene expression, epigenetic	1.99E-06
GO:0051726	regulation of cell cycle	2.28E-06
GO:0006275	regulation of DNA replication	2.55E-06
GO:2000602	regulation of interphase of mitotic cell cycle	2.62E-06
GO:001038	regulation of G2/M transition of mitotic cell cycle	2.62E-

9		06
GO:0016458	gene silencing	3.23E-06
GO:0034622	cellular macromolecular complex assembly	3.72E-06
GO:0071840	cellular component organization or biogenesis	3.95E-06
GO:0065003	macromolecular complex assembly	4.07E-06
GO:0010564	regulation of cell cycle process	4.29E-06
GO:0000911	cytokinesis by cell plate formation	4.37E-06
GO:0033205	cell cycle cytokinesis	4.37E-06
GO:0006346	methylation-dependent chromatin silencing	4.57E-06
GO:0006342	chromatin silencing	5.44E-06
GO:0045814	negative regulation of gene expression, epigenetic	5.44E-06
GO:0043933	macromolecular complex subunit organization	1.14E-05
GO:0003006	developmental process involved in reproduction	1.48E-05
GO:0022607	cellular component assembly	1.65E-05
GO:0016571	histone methylation	1.87E-05
GO:0031047	gene silencing by RNA	2.82E-05
GO:0048449	floral organ formation	3.07E-05
GO:0022414	reproductive process	3.31E-05
GO:0031048	chromatin silencing by small RNA	3.62E-05
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	3.70E-05
GO:0051172	negative regulation of nitrogen compound metabolic process	3.70E-05
GO:0006479	protein methylation	4.35E-05
GO:0008213	protein alkylation	4.35E-05

GO:0010558	negative regulation of macromolecule biosynthetic process	4.62E-05
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	4.62E-05
GO:0009890	negative regulation of biosynthetic process	4.62E-05
GO:0031327	negative regulation of cellular biosynthetic process	4.62E-05
GO:0007346	regulation of mitotic cell cycle	4.95E-05
GO:0010629	negative regulation of gene expression	6.11E-05
GO:0006270	DNA replication initiation	6.44E-05
GO:0044763	single-organism cellular process	6.59E-05
GO:0007017	microtubule-based process	6.60E-05
GO:0048645	organ formation	6.61E-05
GO:0000226	microtubule cytoskeleton organization	6.70E-05
GO:0008283	cell proliferation	9.25E-05
GO:0045892	negative regulation of transcription, DNA-dependent	9.87E-05
GO:0051253	negative regulation of RNA metabolic process	9.87E-05
GO:0010605	negative regulation of macromolecule metabolic process	1.16E-04
GO:0009892	negative regulation of metabolic process	1.16E-04
GO:0031324	negative regulation of cellular metabolic process	1.48E-04
GO:0006366	transcription from RNA polymerase II promoter	1.61E-04
GO:0051052	regulation of DNA metabolic process	1.73E-04
GO:0048523	negative regulation of cellular process	1.87E-04
GO:0043414	macromolecule methylation	2.40E-04
GO:0006352	DNA-dependent transcription, initiation	2.92E-04
GO:0007067	mitosis	2.95E-04

GO:0000280	nuclear division	2.98E-04
GO:0016572	histone phosphorylation	3.01E-04
GO:0032259	methylation	3.18E-04
GO:0016192	vesicle-mediated transport	3.18E-04
GO:0008380	RNA splicing	3.32E-04
GO:0006397	mRNA processing	3.37E-04
GO:0006367	transcription initiation from RNA polymerase II promoter	3.70E-04
GO:0010393	galacturonan metabolic process	4.51E-04
GO:0045488	pectin metabolic process	4.51E-04
GO:0006413	translational initiation	4.63E-04
GO:0006306	DNA methylation	5.06E-04
GO:0006305	DNA alkylation	5.06E-04
GO:0044728	DNA methylation or demethylation	5.06E-04
GO:0048285	organelle fission	5.42E-04
GO:0051322	anaphase	5.49E-04
GO:0052546	cell wall pectin metabolic process	5.64E-04
GO:0032502	developmental process	5.66E-04
GO:0010498	proteasomal protein catabolic process	5.84E-04
GO:0006094	gluconeogenesis	6.61E-04
GO:0019319	hexose biosynthetic process	6.61E-04
GO:0052541	plant-type cell wall cellulose metabolic process	6.95E-04
GO:0006304	DNA modification	6.96E-04
GO:0015988	energy coupled proton transmembrane transport, against electrochemical gradient	7.50E-04

GO:0015991	ATP hydrolysis coupled proton transport	7.50E-04
GO:0044699	single-organism process	7.96E-04
GO:0046364	monosaccharide biosynthetic process	8.19E-04

Appendix xxii - Overrepresented 1-2 h G.O. Function terms (*A. thaliana*)

GO Term	Description	P-value
GO:0003676	nucleic acid binding	4.05E-05
GO:0003677	DNA binding	5.35E-05
GO:0003743	translation initiation factor activity	1.10E-04
GO:1901363	heterocyclic compound binding	1.98E-04
GO:0097159	organic cyclic compound binding	1.98E-04
GO:0003723	RNA binding	2.04E-04
GO:0008135	translation factor activity, nucleic acid binding	4.62E-04
GO:0005488	binding	8.08E-04

Appendix xxiii - Overrepresented 1-2 h G.O. Cellular Component terms (*A. thaliana*)

GO Term	Description	P-value
GO:0000786	nucleosome	2.44E-14
GO:0032993	protein-DNA complex	2.44E-14
GO:0005634	nucleus	4.60E-14
GO:0044427	chromosomal part	2.75E-12
GO:0005829	cytosol	4.49E-11
GO:0044428	nuclear part	3.39E-05
GO:0044445	cytosolic part	5.17E-05
GO:0005667	transcription factor complex	3.70E-04

Appendix xxiv – Overrepresented 3-5 h G.O. Process terms (*A. thaliana*)

GO Term	Description	P-value
GO:0034728	nucleosome organization	1.58E-12
GO:0071824	protein-DNA complex subunit organization	1.58E-12
GO:0006334	nucleosome assembly	1.58E-12
GO:0065004	protein-DNA complex assembly	1.58E-12
GO:0006325	chromatin organization	4.11E-07
GO:0007010	cytoskeleton organization	4.14E-07
GO:0034622	cellular macromolecular complex assembly	1.95E-06
GO:0065003	macromolecular complex assembly	2.04E-06
GO:0051567	histone H3-K9 methylation	2.58E-06
GO:0043933	macromolecular complex subunit organization	5.39E-06
GO:0006996	organelle organization	6.20E-06
GO:0022607	cellular component assembly	8.38E-06
GO:0051276	chromosome organization	9.34E-06
GO:0034968	histone lysine methylation	1.82E-05
GO:0051726	regulation of cell cycle	2.16E-05
GO:0016569	covalent chromatin modification	3.32E-05
GO:0000911	cytokinesis by cell plate formation	3.83E-05
GO:0033205	cell cycle cytokinesis	3.83E-05
GO:0016568	chromatin modification	3.93E-05
GO:0010498	proteasomal protein catabolic process	5.01E-05
GO:0006304	DNA modification	5.13E-05
GO:0006306	DNA methylation	5.13E-05
GO:0006305	DNA alkylation	5.13E-05
GO:0044728	DNA methylation or demethylation	5.13E-05
GO:0016571	histone methylation	6.38E-05
GO:0006275	regulation of DNA replication	7.29E-05
GO:0008380	RNA splicing	9.51E-05
GO:0006366	transcription from RNA polymerase II promoter	9.88E-05
GO:0006479	protein methylation	1.17E-04
GO:0008213	protein alkylation	1.17E-04
GO:0040029	regulation of gene expression, epigenetic	1.44E-04
GO:0016570	histone modification	1.49E-04
GO:0000910	cytokinesis	1.57E-04
GO:0010564	regulation of cell cycle process	1.82E-04
GO:0006270	DNA replication initiation	2.17E-04
GO:0010393	galacturonan metabolic process	2.65E-04
GO:0045488	pectin metabolic process	2.65E-04
GO:0006397	mRNA processing	2.69E-04
GO:0044763	single-organism cellular process	2.74E-04

GO:0008283	cell proliferation	2.75E-04
GO:0006352	DNA-dependent transcription, initiation	4.05E-04
GO:0048453	sepal formation	4.29E-04
GO:0048451	petal formation	4.29E-04
GO:0016043	cellular component organization	5.03E-04
GO:0006342	chromatin silencing	5.37E-04
GO:0045814	negative regulation of gene expression, epigenetic	5.37E-04
GO:0006346	methylation-dependent chromatin silencing	5.47E-04
GO:0006413	translational initiation	5.63E-04
GO:0048449	floral organ formation	6.25E-04
GO:0016192	vesicle-mediated transport	6.53E-04
GO:0022403	cell cycle phase	6.63E-04
GO:0051017	actin filament bundle assembly	7.72E-04
GO:0071840	cellular component organization or biogenesis	9.24E-04
GO:0048645	organ formation	9.76E-04

Appendix xxv - Overrepresented 3-5 h G.O. Function terms (*A. thaliana*)

GO Term	Description	P-value
GO:0003723	RNA binding	1.03E-07
GO:0003677	DNA binding	7.31E-05
GO:0003743	translation initiation factor activity	8.76E-05
GO:0003979	UDP-glucose 6-dehydrogenase activity	9.46E-05
GO:0003676	nucleic acid binding	1.02E-04
GO:0008135	translation factor activity, nucleic acid binding	2.55E-04
GO:1901363	heterocyclic compound binding	4.34E-04
GO:0097159	organic cyclic compound binding	4.34E-04

Appendix xxvi - Overrepresented 3-5 h G.O. Cellular Component terms (*A. thaliana*)

GO Term	Description	P-value
GO:0000786	nucleosome	1.58E-12
GO:0032993	protein-DNA complex	1.58E-12
GO:0044427	chromosomal part	3.09E-11
GO:0005829	cytosol	1.50E-09
GO:0005634	nucleus	5.23E-09
GO:0044428	nuclear part	6.50E-05
GO:0005802	trans-Golgi network	1.23E-04
GO:0005768	endosome	2.34E-04
GO:0030054	cell junction	5.70E-04
GO:0005911	cell-cell junction	5.70E-04
GO:0009506	plasmodesma	5.70E-04
GO:0005794	Golgi apparatus	6.91E-04
GO:0005737	cytoplasm	8.59E-04

Appendix xxvii – Overrepresented 6-8 h G.O. Process terms (*A. thaliana*)

GO Term	Description	P-value
GO:0034728	nucleosome organization	5.63E-09
GO:0071824	protein-DNA complex subunit organization	5.63E-09
GO:0006334	nucleosome assembly	5.63E-09
GO:0065004	protein-DNA complex assembly	5.63E-09
GO:0006096	glycolysis	2.48E-05
GO:0034622	cellular macromolecular complex assembly	8.16E-05
GO:0065003	macromolecular complex assembly	8.71E-05
GO:0006367	transcription initiation from RNA polymerase II promoter	2.43E-04
GO:0043933	macromolecular complex subunit organization	2.48E-04
GO:0022607	cellular component assembly	2.97E-04
GO:0051276	chromosome organization	4.23E-04
GO:0006091	generation of precursor metabolites and energy	4.54E-04
GO:0006006	glucose metabolic process	5.67E-04
GO:0010035	response to inorganic substance	8.42E-04
GO:0010393	galacturonan metabolic process	8.82E-04
GO:0045488	pectin metabolic process	8.82E-04

Appendix xxviii - Overrepresented 6-8 h G.O. Function terms (*A. thaliana*)

GO Term	Description	P-value
GO:0004332	fructose-bisphosphate aldolase activity	1.27E-07
GO:0016832	aldehyde-lyase activity	3.80E-07
GO:0016830	carbon-carbon lyase activity	5.73E-07
GO:0005199	structural constituent of cell wall	6.96E-06
GO:0016829	lyase activity	7.83E-05

Appendix xxix - Overrepresented 6-8 h G.O. Cellular Component terms (*A. thaliana*)

GO Term	Description	P-value
GO:0000786	nucleosome	5.63E-09
GO:0032993	protein-DNA complex	5.63E-09
GO:0044427	chromosomal part	7.58E-08
GO:0030054	cell junction	4.36E-06
GO:0005911	cell-cell junction	4.36E-06
GO:0009506	plasmodesma	4.36E-06
GO:0005887	integral to plasma membrane	9.42E-06
GO:0005829	cytosol	2.35E-05
GO:0005886	plasma membrane	1.89E-04
GO:0005667	transcription factor complex	2.43E-04
GO:0044459	plasma membrane part	4.91E-04

References

Alliegro, M. A., Henry, J. J., & Alliegro, M. C. (2010). Rediscovery of the nucleolinus, a dynamic RNA-rich organelle associated with the nucleolus, spindle, and centrosomes. *Proc. Natl. Acad. Sci.*, *107*(31), 13718-13723.

Altieri, D. C. (1994). Splicing of effector cell protease receptor-1 mRNA is modulated by an unusual retained intron. *Biochem.*, *33*(46), 13848-13855.

Amaldi, P. P., Felicetti, L., & Campioni, N. (1977). Flow of informational RNA from cytoplasmic poly (A)-containing particles to polyribosomes in *Artemia salina* cysts at early stages of development. *Dev. Biol.*, *59*(1), 49-61.

Andrade, L. E., Chan, E. K., Raska, I., Peebles, C. L., Roos, G., & Tan, E. M. (1991). Human autoantibody to a novel protein of the nuclear coiled body: immunological characterization and cDNA cloning of p80-coilin. *J. Exp. Med.*, *173*(6), 1407-1419.

Averbeck, N., Sunder, S., Sample, N., Wise, J. A., & Leatherwood, J. (2005). Negative control contributes to an extensive program of meiotic splicing in fission yeast. *Mol. Cell*, *18*(4), 491-498.

Bachvarova, R., & De Leon, V. (1977). Stored and polysomal ribosomes of mouse ova. *Dev. Biol.*, *58*(2), 248-254.

Barbosa-Morais, N. L., Irimia, M., Pan, Q., Xiong, H. Y., Gueroussov, S., Lee, L. J., *et al*. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science*, *338*(6114), 1587-1593.

Baskin, T. I., & Wilson, J. E. (1997). Inhibitors of protein kinases and phosphatases alter root morphology and disorganize cortical microtubules. *Plant Physiol.*, *113*(2), 493-502.

Baurén, G., & Wieslander, L. (1994). Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell*, *76*(1), 183-192.

Beck, J. S. (1961). Variations in the morphological patterns of autoimmune nuclear fluorescence. *Lancet*, *1*, 1203–1205.

Bensaude, O. (2011). Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity? *Transcription*, *2*(3), 103-108.

Bevilacqua, A., Ceriani, M. C., Capaccioli, S., & Nicolin, A. (2003). Post‐transcriptional regulation of gene expression by degradation of messenger RNAs. *J. Cell. Physiol.*, *195*(3), 356-372.

Bewley, J. D. (1973). Desiccation and protein synthesis in the moss Tortilla ruralis. *Can. J. Bot.*, *51*(1), 203-206.

Beyer, A. L., Bouton, A. H., & Miller, O. L. (1981). Correlation of hnRNP structure and nascent transcript cleavage. *Cell*, *26*(2), 155-165.

Beyer, A. L., & Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.*, *2*(6), 754-765.

Boguski, M. S., Tolstoshev, C. M., & Bassett Jr, D. E. (1994). Gene discovery in dbEST. *Science*, *265*(5181), 1993-1994.

Boothby, T., & Wolniak, S. (2011). Masked mRNA is stored with aggregated nuclear speckles and its asymmetric redistribution requires a homolog of mago nashi. *BMC Cell Biol.*, *12*(1), 45.

Boothby, T. C., Zipper, R. S., van der Weele, C. M., & Wolniak, S. M. (2013). Removal of Retained Introns Regulates Translation in the Rapidly Developing Gametophyte of *Marsilea vestita*. *Dev. Cell – In press*

Bor, Y., Swartz, J., Morrison, A., Rekosh, D., Ladomery, M., & Hammarskjöld, M. L. (2006). The Wilms' tumor 1 (WT1) gene (+ KTS isoform) functions with a CTE to enhance translation from an unspliced RNA with a retained intron. *Genes Dev.*, *20*(12), 1597-1608.

Braddock, M., Muckenthaler, M., White, M. R., Thorbum, A. M., Sommerville, J., Kingsman, A. J., & Kingsman, S. M. (1994). Intron-less RNA injected into the nucleus of Xenopus oocytes accesses a regulated translation control pathway. *Nucleic Acids Res.*, *22*(24), 5255-5264.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., *et al*. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.*, *18*(6), 630-634.

Brown, J. M., Green, J., das Neves, R. P., Wallace, H. A., Smith, A. J., Hughes, J., *et al*. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.*, *182*(6), 1083-1097.

Buckley, P. T., Lee, M. T., Sul, J. Y., Miyashiro, K. Y., Bell, T. J., Fisher, S. A., *et al*. (2011). Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons. *Neuron*, *69*(5), 877-884.

Cajal, S. R. (1910). El nucleo de las celulas piramidales del cerebro humano y de algunos mamiferos. *Trab. Lab. Invest. Biol.*, *8*: 27–62.

Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M., & Buell, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. *BMC Genomics*, *7*(1), 327.

Capco, D. G., & Jeffery, W. R. (1979). Origin and spatial distribution of maternal messenger RNA during oogenesis of an insect, Oncopeltus fasciatus. *J. Cell Sci.*, *39*(1), 63-76.

Carter, K. C., Taneja, K. L., & Lawrence, J. B. (1991). Discrete nuclear domains of poly (A) RNA and their relationship to the functional organization of the nucleus. *J. Cell Biol.*, *115*(5), 1191-1202.

Caspersson, T., & Schultz, J. (1939). Pentose nucleotides in the cytoplasm of growing tissues. *Nature*, *143*(3623), 602-3.

Chang, Y. F., Imam, J. S., & Wilkinson, M. F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.*, *76*, 51-74.

Clark, F., & Thanaraj, T. A. (2002). Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum. Mol. Genet.*, *11*(4), 451-464.

Clement, J. Q., Qian, L., Kaplinsky, N., & Wilkinson, M. F. (1999). The stability and fate of a spliced intron from vertebrate cells. *RNA*, *5*(2), 206-220.

Cmarko, D., Verschure, P. J., Martin, T. E., Dahmus, M. E., Krause, S., Fu, X. D., *et al*. (1999). Ultrastructural analysis of transcription and splicing in the cell nucleus after bromo-UTP microinjection. *Mol. Biol. Cell*, *10*(1), 211-223.

Cohen, S. S. (1998). *A Guide to the Polyamines.* pp. 481-511 New York: Oxford University Press.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinform.*, *21*(18), 3674-3676.

Curry, A. M., Williams, B. D., & Rosenbaum, J. L. (1992). Sequence analysis reveals homology between two proteins of the flagellar radial spoke. *Mol. Cell Biol.*, *12*(9), 3967-3977.

Curtis, D., Apfeld, J., & Lehmann, R. (1995). Nanos is an evolutionarily conserved organizer of anterior-posterior polarity. *Development*, *121*(6), 1899-1910.

Czarnecka, E., Edelman, L., Schöffl, F., & Key, J. L. (1984). Comparative analysis of physical stress responses in soybean seedlings using cloned heat shock cDNAs. *Plant Mol. Biol.*, *3*(1), 45-58.

Dace, H., Sherwin, H. W., Illing, N., & Farrant, J. M. (1998). Use of metabolic inhibitors to elucidate mechanisms of recovery from desiccation stress in the resurrection plant Xerophyta humilis. *Plant Growth Regul.*, *24*(3), 171-177.

Davidson, E. H. (1986). Gene activity in early development. 3rd Ed. pp. 670 New York: Academic press.

Deeb, F. (2009). The role of spermidine in the regulation of development and differentiation in spermatids of Marsilea vestita. Ph.D. University of Maryland, College Park, MD.

Deeb, F., van der Weele, C. M., & Wolniak, S. M. (2010). Spermidine is a morphogenetic determinant for cell fate specification in the male gametophyte of the water fern Marsilea vestita. *Plant Cell*, *22*(11), 3678-3691.

Denis, M. M., Tolley, N. D., Bunting, M., Schwertz, H., Jiang, H., Lindemann, S., *et al*. (2005). Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. *Cell*, *122*(3), 379-391.

Der, J. P., Barker, M. S., Wickett, N. J., W dePamphilis, C., & Wolf, P. G. (2011). De novo characterization of the gametophyte transcriptome in bracken fern, Pteridium aquilinum. *BMC genomics*, *12*(1), 99.

Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., & Pagano, A. (2007). The expanding RNA polymerase III transcriptome. *Trends Genet.*, *23*(12), 614-622.

Dhindsa, R. S., & Bewley, J. D. (1977). Water stress and protein synthesis V. Protein synthesis, protein stability, and membrane permeability in a drought-sensitive and a drought-tolerant moss. *Plant Physiol.*, *59*(2), 295-300.

Dirksen, W. P., Sun, Q., & Rottman, F. M. (1995). Multiple splicing signals control alternative intron retention of bovine growth hormone pre-mRNA. *J. Biol. Chem.*, *270*(10), 5346-5352.

Dostie, J., Lejbkowicz, F., & Sonenberg, N. (2000). Nuclear eukaryotic initiation factor 4E (eIF4E) colocalizes with splicing factors in speckles. *J. Cell Biol.*, *148*(2), 239-246.

Dure, L., & Waters, L. (1965). Long-lived messenger RNA: evidence from cotton seed germination. *Science. 147*, 410.

Dutcher, S. K., & Trabuco, E. C. (1998). The UNI3 gene is required for assembly of basal bodies of Chlamydomonas and encodes δ-tubulin, a new member of the tubulin superfamily. *Mol. Biol. Cell*, *9*(6), 1293-1308.

Ebihara, K., Masuhiro, Y., Kitamoto, T., Suzawa, M., Uematsu, Y., Yoshizawa, T., *et al*. (1996). Intron retention generates a novel isoform of the murine vitamin D receptor that acts in a dominant negative way on the vitamin D signaling pathway. *Mol. Cell. Biol.*, *16*(7), 3393-3400.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC bioinform.*, *10*(1), 48.

Fakan, S., & Bernhard, W. (1971). Localisation of rapidly and slowly labelled nuclear RNA as visualized by high resolution autoradiography. *Exp. Cell Res.*, *67*(1), 129-141.

Fakan, S., & Nobis, P. (1978). Ultrastructural localization of transcription sites and of RNA distribution during the cell cycle of synchronized CHO cells. *Exp. Cell Res.*, *113*(2), 327-337.

Ferreira, J. A., Carmo-Fonseca, M., & Lamond, A. I. (1994). Differential interaction of splicing snRNPs with coiled bodies and interchromatin granules during mitosis and assembly of daughter cell nuclei. *J. Cell Biol.*, *126*(1), 11-23.

Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., *et al*. (2010). Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Res.*, *20*(1), 45-58.

Filipowicz, W., Bhattacharyya, S. N., & Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat. Rev. Genet.*, *9*(2), 102-114.

Flemr, M., Ma, J., Schultz, R. M., & Svoboda, P. (2010). P-body loss is concomitant with formation of a messenger RNA storage domain in mouse oocytes. *Biol. Reprod.*, *82*(5), 1008-1017.

Fox, C. A., Sheets, M. D., & Wickens, M. P. (1989). Poly (A) addition during maturation of frog oocytes: distinct nuclear and cytoplasmic activities and regulation by the sequence UUUUUAU. *Genes Dev.*, *3*(12b), 2151-2162.

Fox, A. H., Lam, Y. W., Leung, A. K., Lyon, C. E., Andersen, J., Mann, M., & Lamond, A. I. (2002). Paraspeckles: a novel nuclear domain. *Curr. Biol., CB*, *12*(1), 13.

Fu, X. D. (1995). The superfamily of arginine/serine-rich splicing factors. *RNA*, *1*(7), 663.

Galante, P. A. F., Sakabe, N. J., Kirschbaum-Slager, N., & De Souza, S. J. (2004). Detection and evaluation of intron retention events in the human transcriptome. *RNA*, *10*(5), 757-765.

Gall, J. G. (2000). Cajal bodies: the first 100 years. *Annu. Rev. Cell Dev. Biol.*, *16*(1), 273-300.

Gama-Carvalho, M., Krauss, R. D., Chiang, L., Valcárcel, J., Green, M. R., & Carmo-Fonseca, M. (1997). Targeting of U2AF65 to sites of active splicing in the nucleus. *J. Cell Biol.*, *137*(5), 975-987.

Gebauer, F., Merendino, L., Hentze, M. W., & Valcárcel, J. (1998). The Drosophila splicing regulator sex-lethal directly inhibits translation of male-specific-lethal 2 mRNA. *RNA*, *4*(2), 142-150.

Geimer, S., Clees, J., Melkonian, M., & Lechtreck, K. F. (1998). A novel w95-kD protein is located in a linker between cytoplasmic microtubules and basal bodies in a green flagellate and forms striated filaments in vitro. *J. Cell Biol.*, *140*(5), 1149-1158.

Gerhard, D. S., Wagner, L., Feingold, E. A., Shenmen, C. M., Grouse, L. H., Schuler, G., *et al*. (2004). The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.*, *14*(10b), 2121.

Girard, C., Will, C. L., Peng, J., Makarov, E. M., Kastner, B., Lemm, I., *et al*. (2012). Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat Commun.*, *3*, 994.

Gorgoni, B., & Gray, N. K. (2004). The roles of cytoplasmic poly (A)-binding proteins in regulating gene expression: a developmental perspective. *Brief. Funct. Genomic. Proteomic.*, *3*(2), 125-141.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., *et al*. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.*, *29*(7), 644-652.

Grosfeld, H., & Littauer, U. Z. (1975). Cryptic form of mRNA in dormant Artemia salina cysts. *Biochem. Biophys. Res. Commu.*, *67*(1), 176-181.

Gross, P. R., & Cousineau, G. H. (1963). Effects of actinomycin D on macromolecule synthesis and early development in sea urchin eggs. *Biochem. Biophys. Res. Commu.*, *10*, 321.

Gross, P. R., & Cousineau, G. H. (1964). Macromolecule synthesis and the influence of actinomycin on early development. *Exp. Cell Res.*, *33*, 368.

Hall, L. L., Smith, K. P., Byron, M., & Lawrence, J. B. (2006). Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.*, *288*(7), 664-675.

Hampson, R. K., & Rottman, F. M. (1987). Alternative processing of bovine growth hormone mRNA: nonsplicing of the final intron predicts a high molecular weight variant of bovine growth hormone. *Proc. Natl. Acad. Sci.*, *84*(9), 2673-2677.

Harris, B., & Dure III, L. (1978). Developmental regulation in cotton seed germination: polyadenylation of stored messenger RNA. *Biochem.*, *17*(16), 3250-3256.

Hart, P. E., & Wolniak, S. M. (1998). Spermiogenesis in Marsilea vestita: a temporal correlation between centrin expression and blepharoplast differentiation. *Cell motility and the cytoskeleton*, *41*(1), 39-48.

Hart, P. E., & Wolniak, S. M. (1999). Molecular cloning of a centrin homolog from Marsilea vestita and evidence for its translational control during spermiogenesis. *Biochem. Cell Biol.*, *77*(2), 101-108.

Harvey, E. B., 1936. Parthenogenetic merogony or cleavage without nuclei in *Arbacia punctulata. Biol. Bull.*, **71**, 101–121.

Herr, A. J., Jensen, M. B., Dalmay, T., & Baulcombe, D. C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science*, *308*(5718), 118-120.

Hoskins, A. A., Friedman, L. J., Gallagher, S. S., Crawford, D. J., Anderson, E. G., Wombacher, R., *et al*. (2011). Ordered and dynamic assembly of single spliceosomes. *Science*, *331*(6022), 1289-1295.

Huang, S., & Spector, D. L. (1991). Nascent pre-mRNA transcripts are associated with nuclear regions enriched in splicing factors. *Genes Dev.*, *5*(12a), 2288-2302.

Huang, S., Deerinck, T. J., Ellisman, M. H., & Spector, D. L. (1994). In vivo analysis of the stability and transport of nuclear poly (A)+ RNA. *J. Cell Biol.*, *126*(4), 877-900.

Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Res.*, *9*(9), 868-877.

Iida, K., Seki, M., Sakurai, T., Satou, M., Akiyama, K., Toyoda, T., *et al*. (2004). Genome-wide analysis of alternative pre-mRNA splicing in Arabidopsis thaliana based on full-length cDNA sequences. *Nucleic acids Res*, *32*(17), 5096-5103.

Jaillon, O., Bouhouche, K., Gout, J. F., Aury, J. M., Noel, B., Saudemont, B., *et al*. (2008). Translational control of intron splicing in eukaryotes. *Nature*, *451*(7176), 359-362.

Johnson, C., Primorac, D., McKinstry, M., McNeil, J., Rowe, D., & Lawrence, J. B. (2000). Tracking COL1A1 RNA in osteogenesis imperfecta splice-defective transcripts initiate transport from the gene but are retained within the SC35 domain. *J. Cell Biol.*, *150*(3), 417-432.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., *et al*. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.*, *3*(9), 576-583.

Kakkar, R. K., & Sawhney, V. K. (2002). Polyamine research in plants–a changing perspective. *Physiol. Plant.*, *116*(3), 281-292.

Kalyna, M., Simpson, C. G., Syed, N. H., Lewandowska, D., Marquez, Y., Kusenda, B., *et al*. (2012). Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis. *Nucleic Acids Res.*, *40*(6), 2454-2469.

Kan, Zhengyan, and Warren Gish. (2002). Selecting for functional alternative splices in ESTs. *Genome Res., 12*, 1837-1845.

Kedinger, C., Gniazdowski, M., Mandel Jr, J. L., Gissinger, F., & Chambon, P. (1970). Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. *Biochem. Biophys. Res. Commu.*, *38*(1), 165.

Kim, H. Y., & Gladyshev, V. N. (2006). Alternative first exon splicing regulates subcellular distribution of methionine sulfoxide reductases. *BMC Mol. Biol.*, *7*(1), 11.

Kimura, M., Kotani, S., Hattori, T., Sumi, N., Yoshioka, T., Todokoro, K., & Okano, Y. (1997). Cell cycle-dependent expression and spindle pole localization of a novel human protein kinase, Aik, related to Aurora of Drosophila and yeast Ipl1. *J. Biol. Chem.*, *272*(21), 13766-13771.

Khodor, Y. L., Rodriguez, J., Abruzzi, K. C., Tang, C. H. A., Marr, M. T., & Rosbash, M. (2011). Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes Dev.*, *25*(23), 2502-2512.

Klink, V. P. (2001). Patterns of translation during microgametogenesis in *Marsilea vestita*. Ph.D. University of Maryland, College Park, MD.

Klink, V. P., & Wolniak, S. M. (2001). Centrin is necessary for the formation of the motile apparatus in spermatids of Marsilea. *Mol. Biol. Cell*, *12*(3), 761-776.

Klink, V. P., & Wolniak, S. M. (2003). Changes in the abundance and distribution of conserved centrosomal, cytoskeletal and ciliary proteins during spermiogenesis in Marsilea vestita. *Cell Motil. Cytoskeleton*, *56*(1), 57-73.

Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., *et al*. (2006). CAGE: cap analysis of gene expression. *Nat. Methods*, *3*(3), 211-222.

Koornneef, M., Bentsink, L., & Hilhorst, H. (2002). Seed dormancy and germination. *Curr. Opin. Plant Biol.*, *5*(1), 33-36.

Kornberg, R. D. (1999). Eukaryotic transcriptional control. *Trends Biochem. Sci.*, *24*(12), 46-49.

Kornblihtt, A. R., Schor, I. E., Alló, M., Dujardin, G., Petrillo, E., & Muñoz, M. J. (2013). Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol*., *14*, 153-165.

Koroleva, O. A., Calder, G., Pendle, A. F., Kim, S. H., Lewandowska, D., Simpson, C. G., *et al*. (2009a). Dynamic behavior of Arabidopsis eIF4A-III, putative core protein of exon junction complex: fast relocation to nucleolus and splicing speckles under hypoxia. *Plant Cell.*, *21*(5), 1592-1606.

Koroleva, O., Brown, J., & Shaw, P. E. (2009b). Localization of eIF4A-III in the nucleolus and splicing speckles is an indicator of plant stress. *Plant Signal. Behav.*, *4*(12), 1148-1151.

Krause, S., Fakan, S., Weis, K., & Wahle, E. (1994). Immunodetection of poly (A) binding protein II in the cell nucleus. *Exp. Cell Res.*, (1), 75-82.

Kurnick, N. B. (1955). Pyronin Y in the methyl-green-pyronin histological stain. *Biotech Histochem.*, *30*(5), 213-230.

Lamond, A. I., & Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.*, *4*(8), 605-612.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, *10*(3), R25.

Larsson, S. H., Charlieu, J. P., Miyagawa, K., Engelkamp, D., Rassoulzadegan, M., Ross, A., *et al*. (1995). Subnuclear localization of WT1 in splicing or transcription factor domains is regulated by alternative splicing. *Cell*, *81*(3), 391-401.

Le Hir, H., Gatfield, D., Braun, I. C., Forler, D., & Izaurralde, E. (2001). The protein Mago provides a link between splicing and mRNA localization. *EMBO*, *2*(12), 1119-1124.

LeDizet, M., & Piperno, G. (1995). The light chain p28 associates with a subset of inner dynein arm heavy chains in Chlamydomonas axonemes. *Mol. Biol. Cell*, *6*(6), 697.

Leser, G. P., Fakan, S., Martin, T. E. (1989). Ultrastructural distribution of ribonucleoprotein complexes duirng mitosis. snRNP antigens are contained in mitotic granule clusters. *Eur. J. Cell Biol., 50,* 376–389.

Li, Q., Imataka, H., Morino, S., Rogers, G. W. Jr., Richter-Cook, N. J., Merrick, W. C., Sonenberg, N. (1999). Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. *Mol. Cell Biol., 19,* 7336–7346.

Lindell, T. J., Weinberg, F., Morris, P. W., Roeder, R. G., & Rutter, W. J. (1970). Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science*, *170*(956), 447.

Luo, M. J., & Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proc. Natl. Acad. Sci.*, *96*(26), 14937-14942.

Malatesta, M., Perdoni, F., Battistelli, S., Muller, S., & Zancanaro, C. (2009). The cell nuclei of skeletal muscle cells are transcriptionally active in hibernating edible dormice. *BMC Cell Biol.*, *10*(1), 19.

Mansilla, A., López-Sánchez, C., De La Rosa, E. J., García-Martínez, V., Martínez-Salas, E., De Pablo, F., & Hernández-Sánchez, C. (2005). Developmental regulation of a proinsulin messenger RNA generated by intron retention. *EMBO*, *6*(12), 1182-1187.

Martin, O. C., Gunawardane, R. N., Iwamatsu, A., & Zheng, Y. (1998). Xgrip109: a γ tubulin–associated protein with an essential role in γ tubulin ring complex (γTuRC) assembly and centrosome function. *J. Cell Biol.*, *141*(3), 675-687.

Maquat, L. E., & Carmichael, G. G. (2001). Quality control of mRNA function. *Cell*, *104*(2), 173.

Marquez, Y., Brown, J. W., Simpson, C., Barta, A., & Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. *Genome Res.*, *22*(6), 1184-1195.

Marrs, K. A., & Walbot, V. (1997). Expression and RNA splicing of the maize glutathione S-transferase Bronze2 gene is regulated by cadmium and other stresses. *Plant Physiol.*, *113*(1), 93-102.

Mastrangelo, A. M., Belloni, S., Barilli, S., Ruperti, B., Di Fonzo, N., Stanca, A. M., & Cattivelli, L. (2005). Low temperature promotes intron retention in two e-cor genes of durum wheat. *Planta*, *221*(5), 705-715.

McCullough, A. J., & Berget, S. M. (1997). G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol. Cell. Biol.*, *17*(8), 4562-4571.

McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., & Richter, J. D. (1989). Poly (A) elongation during Xenopus oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.*, *3*(6), 803-815.

McGuire, A. M., Pearson, M. D., Neafsey, D. E., & Galagan, J. E. (2008). Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol*, *9*(3), R50.

Melčák, I., Cermanová, Š., Jirsová, K., Koberna, K., Malínský, J., & Raška, I. (2000). Nuclear pre-mRNA compartmentalization: trafficking of released transcripts to splicing factor reservoirs. *Mol. Biol. Cell*, *11*(2), 497-510.

Merkin, J., Russell, C., Chen, P., & Burge, C. B. (2012). Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science*,*338*(6114), 1593-1599.

Moen, P. T., Johnson, C. V., Byron, M., Shopland, L. S., Ivana, L., Imbalzano, A. N., & Lawrence, J. B. (2004). Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis. *Mol. Biol. Cell*, *15*(1), 197-206.

Moran, R. C. (2004). *A natural history of ferns*. Portland, Timber Pr.

Morozova, O., Hirst, M., & Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annu. Rev. Genomics Hum. Genet.*, *10*, 135-151.

Mount, S. M. (1982). A catalogue of splice junction sequences. *Nucleic Acids Res.*, *10*(2), 459-472.

Mount, S. M., Burks, C., Herts, G., Stormo, G. D., White, O., & Fields, C. (1992). Splicing signals in Drosophila: intron size, information content, and consensus sequences. *Nucleic Acids Res.*, *20*(16), 4255-4262.

Mourier, T., & Jeffares, D. C. (2003). Eukaryotic intron loss. *Science*, *300*(5624), 1393-1393.

Muthukrishnan, S., Filipowicz, W., Sierra, J. M., Both, G. W., Shatkin, A., & Ochoa, S. (1975). mRNA methylation and protein synthesis in extracts from embryos of brine shrimp, Artemia salina. *J. Biol. Chem.*, *250*(24), 9336-9341.

Nakabayashi, K., Okamoto, M., Koshiba, T., Kamiya, Y., & Nambara, E. (2005). Genome‐wide profiling of stored mRNA in Arabidopsis thaliana seed germination: epigenetic and genetic regulation of transcription in seed. *Plant J.*, *41*(5), 697-709.

Nakamura, M., Masuda, H., Horii, J., Kuma, K. I., Yokoyama, N., Ohba, T., *et al*. (1998). When overexpressed, a novel centrosomal protein, RanBPM, causes ectopic microtubule nucleation similar to γ-tubulin. *J. Cell Biol.*, *143*(4), 1041-1052.

Nakayasu, H., & Ueda, K. (1984). Small nuclear RNA-protein complex anchors on the actin filaments in bovine lymphocyte nuclear matrix. *Cell Struct. Funct.*, *9*(4), 317.

Ner-Gaon, H., Leviatan, N., Rubin, E., & Fluhr, R. (2007). Comparative cross-species alternative splicing in plants. *Plant Physiol.*, *144*(3), 1632-1641.

Neugebauer, K. M., & Roth, M. B. (1997). Transcription units as RNA processing units. *Genes Dev.*, *11*(24), 3279-3285.

Neugebauer, K. M. (2002). On the importance of being co-transcriptional. *J. Cell Sci.*, *115*(20), 3865-3871.

Nott, A., Meislin, S. H., & Moore, M. J. (2003). A quantitative analysis of intron effects on mammalian gene expression. *RNA*, *9*(5), 607-617.

Oakley, B. R., Oakley, C. E., Yoon, Y., & Jung, M. K. (1990). γ-Tubulin is a component of the spindle pole body that is essential for microtubule function in Aspergillus nidulans. *Cell*, *61*(7), 1289-1301.

O'Brien, T. P., & McCully, M. E. (1981). *The study of plant structure: principles and selected methods*. *Melbourne, Termarcarphi Pty. LTD*.

Oliver, M. J., & Bewley, J. D. (1997). Desiccation-tolerance of plant tissues: a mechanistic overview. *Hortic. Rev.*, *18*, 171-213.

Ohler, U., Shomron, N., & Burge, C. B. (2005). Recognition of unknown conserved alternatively spliced exons. *PLoS Comput. Biol.*, *1*(2), e15.

Okoniewski, M. J., & Miller, C. J. (2006). Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC bioinform.*, *7*(1), 276.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986). Splicing of messenger RNA precursors. *Annu. Rev. Biochem.*, *55*(1), 1119-1150.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, *40*(12), 1413-1415.

Pandya-Jones, A., & Black, D. L. (2009). Co-transcriptional splicing of constitutive and alternative exons. *RNA*, *15*(10), 1896-1908.

Paris, J., Osborne, H. B., Couturier, A., Le Guellec, R., & Philippe, M. (1988). Changesin the polyadenylation of specific stable RNA during the early development of *Xenopus laevis*. *Gene*, *72*(1), 169-176.

Park, N. I., Yeung, E. C., & Muench, D. G. (2009). Mago Nashi is involved in meristem organization, pollen formation, and seed development in Arabidopsis. *Plant Sci.*, *176*(4), 461-469.

Pennisi, E. (2003). A low number wins the GeneSweep pool. *Science*, *300*(5625), 1484-1484.

Perraud, M., Gioud, M., & Monier, J. C. (1979). Intranuclear structures of monkey kidney cells recognised by immunofluorescence and immuno-electron microscopy using anti-ribonucleoprotein antibodies. *Annales d'immunologie*, *130*(5), 635.

Prasanth, K. V., Sacco-Bubulya, P. A., Prasanth, S. G., & Spector, D. L. (2003). Sequential entry of components of gene expression machinery into daughter nuclei. *Mol. Biol. Cell*, *14*(3), 1043-1057.

Prasanth, K. V., Prasanth, S. G., Xuan, Z., Hearn, S., Freier, S. M., Bennett, C. F., *et al*. (2005). Regulating gene expression through RNA nuclear retention. *Cell*, *123*(2), 249-263.

Proweller, A., & Butler, S. (1994). Efficient translation of poly (A)-deficient mRNAs in Saccharomyces cerevisiae. *Genes Dev.*, *8*(21), 2629-2640.

Rafiq, M., Suen, C. K., Choudhury, N., Joannou, C. L., White, K. N., & Evans, R. W. (1997). Expression of recombinant human ceruloplasmin–an absolute requirement for splicing signals in the expression cassette. *FEBS Lett.*, *407*(2), 132-136.

Raff, R. A., Colot, H. V., Selvig, S. E., & Gross, P. R. (1972). Oogenetic origin of messenger RNA for embryonic synthesis of microtubule proteins. *Nature, 235,* 211-214.

Rajjou, L., Gallardo, K., Debeaujon, I., Vandekerckhove, J., Job, C., & Job, D. (2004). The effect of α-amanitin on the Arabidopsis seed proteome highlights the distinct roles of stored and neosynthesized mRNAs during germination. *Plant Physiol.*, *134*(4), 1598-1613.

Reich, C.I., VanHoy, R.W., Porter, G.L., Wise, J.A. (1992). Mutations at the 3' splice site can be suppressed by compensatory base changes in the U1 snRNA in fission yeast. Cell, *69,* 1159-1169.

Reuter, R., Appel, B., Rinke, J., & Lührmann, R. (1985). Localization and structure of snRNPs during mitosis: immunofluorescent and biochemical studies. *Exp. Cell Res.*, *159*(1), 63-79.

Rino, J., & Carmo-Fonseca, M. (2009). The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends Cell Biol.*, *19*(8), 375-384.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nat. biotechnol.*, *29*(1), 24-26.

Rochester, D. E., Winer, J. A., & Shah, D. M. (1986). The structure and expression of maize genes encoding the major heat shock protein, hsp70. *EMBO*, *5*(3), 451.

Romano, M., Marcucci, R., & Baralle, F. E. (2001). Splicing of constitutive upstream introns is essential for the recognition of intra-exonic suboptimal splice sites in the thrombopoietin gene. *Nucleic Acids Res.*, *29*(4), 886-894.

Romfo, C. M., Alvarez, C. J., Van Heeckeren, W. J., Webb, C. J., & Wise, J. A. (2000). Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. *Mol. Cell. Biol.*, *20*(21), 7955-7970.

Rosenthal, E. T., Tansey, T. R., & Ruderman, J. V. (1983). Sequence-specific adenylations and deadenylations accompany changes in the translation of maternal messenger RNA after fertilization of Spisula oocytes. *J. Mol. Biol.*, *166*(3), 309.

Rosenthal, E. T., Bahr, M., & Wilt, F. H. (1993). Regulation of Maternal Messenger RNA Translation during Oogenesis and Embryogenesis in *Urechis caupo*. *Dev. Biol.*, *155*(2), 297-306.

Roy, S. W., & Gilbert, W. (2006). The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.*, *7*(3), 211-221.

Roy, S. W. (2006). Intron-rich ancestors. *Trends Genet.*, *22*(9), 468-471.

Royce, T. E., Rozowsky, J. S., & Gerstein, M. B. (2007). Assessing the need for sequence-based normalization in tiling microarray experiments. *Bioinform.*, *23*(8), 988-997.

Russell, J., & Zomerdijk, J. C. B. M. (2006, December). The RNA polymerase I transcription machinery. *Biochem. Soc. Symp., 73*, 203-216.

Ryu, W. S., & Mertz, J. E. (1989). Simian virus 40 late transcripts lacking excisable intervening sequences are defective in both stability in the nucleus and transport to the cytoplasm. *J. Virol.*, *63*(10), 4386-4394.

Sallés, F. J., Richards, W. G., & Strickland, S. (1999). Assaying the polyadenylation state of mRNAs. *Methods*, *17*(1), 38-45.

Sakabe, N. J., & de Souza, S. J. (2007). Sequence features responsible for intron retention in human. *BMC genomics*, *8*(1), 59.

Scott, H. B., & Oliver, M. J. (1994). Accumulation and polysomal recruitment of transcripts in response to desiccation and rehydration of the moss Tortula ruralis. *J. Exp. Bot.*, *45*(5), 577-583.

Sharp, L. W. (1914). Spermatogenesis in Marsilia. *Bot. Gaz.*, 419-431.

Silver, D. L., Watkins-Chow, D. E., Schreck, K. C., Pierfelice, T. J., Larson, D. M., Burnetti, A. J., *et al*. (2010). The exon junction complex component Magoh controls brain size by regulating neural stem cell division. *Nat. Neurosci.*, *13*(5), 551-558.

Simon, R., Tassan, J. P., & Richter, J. D. (1992). Translational control by poly (A) elongation during Xenopus development: differential repression and enhancement by a novel cytoplasmic polyadenylation element. *Genes Dev.*, *6*(12b), 2580-2591.

Sims, R. J. III, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev., 18*, 2437-2468.

Smith, K. P., Moen, P. T., Wydner, K. L., Coleman, J. R., & Lawrence, J. B. (1999). Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *J. Cell Biol.*, *144*(4), 617-629.

Sommerville, J., & Ladomery, M. (1996). Masking of mRNA by Y-box proteins. *FASEB*, *10*(4), 435-443.

Spector, D. L., Schrier, W. H., Busch, H. (1983). Immunoelectron microscopic localization of snRNPs. *Biol. Cell, 49*, 1–10.

Spector, D. L., Smith, H. C. (1986). Redistribution of U-snRNPs during mitosis. *Exp. Cell Res.*, *163*, 87–94.

Spector, D. L., Fu, X. D., & Maniatis, T. (1991). Associations between distinct pre-mRNA splicing components and the cell nucleus. *EMBO*, *10*(11), 3467.

Spector, D. L., & Lamond, A. I. (2011). Nuclear speckles. *Cold Spring Harbor Perspect. Biol., 3*(2).

Spiegel, S., & Marcus, A. (1975). Polyribosome formation in early wheat embryo germination independent of either transcription or polyadenylation. *Nature, 256,* 228-230.

Spirin, A. S., & Nemer, M. (1965). Messenger RNA in early sea-urchin embryos: cytoplasmic particles. *Science*, *150*(693), 214-217.

Stamm, S., Zhu, J., Nakai, K., Stoilov, P., Stoss, O., & Zhang, M. Q. (2000). An alternative-exon database and its statistical analysis. *DNA Cell Biol.*, *19*(12), 739-756.

Sterner, D. A., & Berget, S. M. (1993). In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. *Mol. Cell. Biol.*, *13*(5), 2677-2687.

Strzelecka, M., Trowitzsch, S., Weber, G., Lührmann, R., Oates, A. C., & Neugebauer, K. M. (2010). Coilin-dependent snRNP assembly is essential for zebrafish embryogenesis. *Nat. Struct. Biol.*, *17*(4), 403-409.

Sugnet, C. W., Kent, W. J., Ares Jr, M., & Haussler, D. (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput., 9*, 66-77.

Swift, H. (1959). Studies on nuclear fine structure. *Brookhaven Symp. Biol., 12,* 134.

Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., & Brown, J. W. (2012). Alternative splicing in plants–coming of age. *Trends Plant Sci., 17*(10), 616-623.

Tabor, C. W., & Tabor, H. (1984). Polyamines. *Annu. Rev. Biochem.*, *53*(1), 749-790.

Talerico, M., & Berget, S. M. (1994). Intron definition in splicing of small Drosophila introns. *Mol. Cell. Biol.*, *14*(5), 3434-3445.

Tange, T. Ø., Nott, A., & Moore, M. J. (2004). The ever-increasing complexities of the exon junction complex. *Curr. Opin. Cell Biol.*, *16*(3), 279-284.

Tate, W. P., & Marshall, C. J. (1991). Post-dormancy transcription and translation in the brine shrimp. *Artemia Biology*. Ed. Browne, R. A., Sorgeloos, P., & Trotmn, C. N. A. pp. 21-36. Boca Raton: CRC Press.

Thiry, M. (1995). The interchromatin granules. *Histol. Histopathol.*, *10*(4), 1035.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2012). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform*.

Toba, K., Winton, E. F., Koike, T., & Shibata, A. (1995). Simultaneous three-color analysis of the surface phenotype and DNA-RNA quantitation using 7-amino-actinomycin D and pyronin Y. *J. Immunol. Methods*, *182*(2), 193-207.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., *et al*. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.*, *28*(5), 511-515.

Tsai, C. W., & Wolniak, S. M. (2001). Cell cycle arrest allows centrin translation but not basal body formation during spermiogenesis in Marsilea. *J. Cell Sci.*, *114*(23), 4265-4272.

Tsai, C. W., Van der Weele, C. M., & Wolniak, S. M. (2004). Differential segregation and modification of mRNA during spermiogenesis in *Marsilea vestita*. *Dev. Biol., 269*(2), 319-330.

Tucker, K. E., & Matera, A. G. (2005). The Cajal body: a nuclear gathering place. *Visions of the Cell Nucleus. P. Hemmerich and S. Diekmann, editors. American Scientific Publishers, Stevenson Ranch, CA*, *159*, 171.

Valdés-Santiago, L., Cervantes-Chavez, J. A., Winkler, R., León-Ramírez, C. G., & Ruiz-Herrera, J. (2012). Phenotypic comparison of samdc and spe mutants reveals complex relationships of polyamine metabolism in Ustilago maydis. *Microbiol.*, *158*(Pt 3), 674-684.

Van Der Weele, C. M., Tsai, C. W., & Wolniak, S. M. (2007). Mago nashi is essential for spermatogenesis in Marsilea. *Mol. Biol. Cell*, *18*(10), 3711-3722.

Vargas, D. Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S. A., *et al*. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell*, *147*(5), 1054-1065.

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial analysis of gene expression. *Science*, *270*(5235), 484-486.

Verheijen, R., Kuijpers, H., Vooijs, P., Van Venrooij, W., & Ramaekers, F. (1986). Distribution of the 70K U1 RNA-associated protein during interphase and mitosis. Correlation with other U RNP particles and proteins of the nuclear matrix. *J. Cell Sci., 86*(1), 173-190.

Visa, N., Puvion-Dutilleul, F., Bachellerie, J. P., & Puvion, E. (1993). Intranuclear distribution of U1 and U2 snRNAs visualized by high resolution in situ hybridization: revelation of a novel compartment containing U1 but not U2 snRNA in HeLa cells. *Eur. J. Cell. Biol.*, *60*(2), 308.

Vuohelainen, S., Pirinen, E., Cerrada‐Gimenez, M., Keinänen, T. A., Uimari, A., Pietilä, M., *et al*. (2010). Spermidine is indispensable in differentiation of 3T3‐L1 fibroblasts to adipocytes. *J. Cell. Mol. Med.*, *14*(6b), 1683-1692.

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell*, *136*(4), 701-718.

Wang, Z., Lo, H. S., Yang, H., Gere, S., Hu, Y., Buetow, K. H., Lee, M.P. (2003). Computational Analysis and Experimental Validation of Tumor-associated Alternative RNA Splicing in Human Cancer. *Cancer Res. 63,* 655-657

Wang, B. B., & Brendel, V. (2006). Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci.*, *103*(18), 7175-7180.

Wang, Z., & Burge, C. B. (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, *14*(5), 802-813.

Weber, K., Geisler, N., Plessmann, U., Bremerich, A., Lechtreck, K. F., & Melkonian, M. (1993). SF-assemblin, the structural protein of the 2-nm filaments from striated microtubule associated fibers of algal flagellar roots, forms a segmented coiled coil. *J. Cell Biol.*, *121*(4), 837-845.

Weinmann, R., & Roeder, R. G. (1974). Role of DNA-Dependent RNA Polymerase III in the Transcription of the tRNA and 5S RNA Genes. *Proc. Natl. Acad. Sci., 71*(5), 1790-1794.

Wetterberg, I., Bauren, G., & Wieslander, L. (1996). The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA*, *2*(7), 641.

Whitfield, W. G., Millar, S. E., Saumweber, H., Frasch, M., & Glover, D. M. (1988). Cloning of a gene encoding an antigen associated with the centrosome in Drosophila. *J. Cell Sci., 89*(4), 467-480.

Wierzbicki, A. T., Ream, T. S., Haag, J. R., & Pikaard, C. S. (2009). RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.*, *41*(5), 630-634.

Wilkinson, M. F., & Shyu, A. B. (2002). RNA surveillance by nuclear scanning? *Nat. Cell Biol.*, *4*(6), E144-E147.

Wilt, F. H. (1973). Polyadenylation of maternal RNA of sea urchin eggs after fertilization. *Proc. Natl. Acad. Sci.*, *70*(8), 2345-2349.

Wilusz, C. J., Wang, W., & Peltz, S. W. (2001). Curbing the nonsense: the activation and regulation of mRNA surveillance. *Genes Dev.*, *15*(21), 2781-2785.

Winter, J., Wright, R., Duck, N., Gasser, C., Fraley, R., & Shah, D. (1988). The inhibition of petunia hsp70 mRNA processing during CdCl 2 stress. *Mol. Gen. Genet.*, *211*(2), 315-319.

Witczak, O., Skalhegg, B. S., Keryer, G., Bornens, M., Tasken, K., Jahnsen, T., Orstavik, S. (1998). Cloning and characterization of a cDNA encoding an A-kinase anchoring protein located in the centrosome, AKAP450. *EMBO 18,* 1858 –1868.

Wolfrum, U., & Salisbury, J. L. (1998). Expression of centrin isoforms in the mammalian retina. *Exp. Cell Res.*, *242*(1), 10-17.

Wolniak, S. M., Klink, V. P., Hart, P.E. and Tsai, C. W. (2000). Control of the development and motility in the spermatozoids of lower plants. Grav. *Sp. Biol. Bull. 13,* 85-93.

Wolniak, S. M., van der Weele, C. M., Deeb, F., Boothby, T., & Klink, V. P. (2011). Extremes in rapid cellular morphogenesis: post-transcriptional regulation of spermatogenesis in Marsilea vestita. *Protoplasma*, *248*(3), 457-473.

Wu, L., Wells, D., Tay, J., Mendis, D., Abbott, M. A., Barnitt, A., *et al*. (1998). CPEB-Mediated Cytoplasmic Polyadenylation and the Regulation of Experience-Dependent Translation of -CaMKII mRNA at Synapses. *Neuron, 21*, 1129-1139.

Xing, Y., & Lawrence, J. B. (1993). Nuclear RNA tracks: structural basis for transcription and splicing? *Trends Cell Biol. 3*(10), 346-353.

Xing, Y., Johnson, C. V., Moen, P. T., McNeil, J. A., & Lawrence, J. (1995). Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC-35 domains. *J. Cell Biol.*,*131*(6), 1635-1647.

Yap, K., Lim, Z. Q., Khandelia, P., Friedman, B., & Makeyev, E. V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev.*, *26*(11), 1209-1223.

Yatin, M. (2002). Polyamines in living organisms. *J. Cell Mol. Biol.*, *1*, 57-67.

Yost, H. J., & Lindquist, S. (1986). RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis. *Cell*, *45*(2), 185.

Zhang, Y., Asante, K. S., & Jung, A. (1986). Stage-dependent inhibition of chloroquine on Plasmodium falciparum in vitro. *J. Parasitol.*, 830-836.

Zheng, C. L., Fu, X. D., & Gribskov, M. (2005). Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA*, *11*(12), 1777-1787.

Zhou, Z., Luo, M. J., Straesser, K., Katahira, J., Hurt, E., & Reed, R. (2000). The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature*, *407*(6802), 401-426.