The role of alternative splicing in mammalian preimplantation development

CHRISTOPHER DOUGLAS ROBERT WYATT

TESI DOCTORAL UPF / 2018

Director de la tesi:
Manuel Irimia

Systems Biology,
Centre for Genomic Regulation (CRG)

upf. Universitat Pompeu Fabra Barcelona

This PhD was supported by a grant from “La Caixa” foundation.

ABSTRACT

The transition from maternal to embryonic control is a crucial time in an organism's life cycle, with an overhaul of the epigenetic marks, targeted mRNA degradation and the initiation of transcription. Yet, not only do these well-studied molecular pathways change, but we show that the predominant alternatively spliced isoform can also change.

Using RNA-Seq from human, mouse and cow representing the transition from the oocyte to blastocyst, we identified thousands of exon-skipping events, most of which change at the stage of genome activation. Particularly, we found many changes that led to premature stop codons or non-canonical isoforms, generally through temporary skipping of constitutive exons. Many of these exon-skipping events were in genes related to DNA repair in the three species. Comparisons to RNA binding protein knockdowns, led us to discover a core set of proteins that may be responsible for the inclusion of the skipped exons (include Snrpb and Snrpd2). We could reduce levels of skipping by microinjection of Snrpb/d2 mRNA into mouse zygotes, and show that this increased the levels of etoposide-induced DNA damage response. This suggests that Snrpb/d2 may be required for correct splicing of DNA repair genes at genome activation.

Apart from these temporary disruptive exon-skipping events, we also discovered many conserved events. These were mostly alternative protein isoforms with shift-like dynamics, and enriched in important developmental pathways, such as the Wnt pathway, Hippo signalling and chromatin modifications. Finally, we discuss the impact of these conserved events in early development, detailing domains and predicted function.

RESUM

La transició del control matern al control embrionari és un moment crucial en el cicle vital d'un organisme viu, el qual inclou canvis epigenètics, degradació de certs ARNm i l'inici de la transcripció a càrrec de l'embrió. Tot i ser aquests, el canvis més ben caracteritzats, no són els únics que es donen, de fet, hem vist que les isoformes majoritàries de "splicing" també canvien durant aquesta transició.

Utilitzant dades de seqüenciació de ARN (RNA-seq) provinents de cèl·lules d'humà, ratolí i vaca, d'oòcit a blastocist, hem pogut identificar milers d'exons exclosos (exon-skipping), la majoria dels quals pateixen canvis en l'estadi de l'activació del genoma. Especialment, hem identificat que molts d'aquests canvis donen lloc a l'aparició prematura d'un codó-stop o a isoformes no canòniques i això es produeix degut a la exclusió d'exons constitutius de manera transitòria o temporal. La majoria de fenòmens de "exon-skipping", observats en les tres espècies, es produeixen en gens implicats en la reparació de l'ADN i el control del cicle cel·lular. Comparant dades provinents de experiments de pèrdua de funció (Knokdown) de proteïnes d'unió a l'ARN, hem pogut descobrir un grup de proteïnes que podrien ser les responsables de la inclusió dels exons que s'exclouen (Snrpb i Snrpd2). Mitjançant la microinjecció d'ARNm codificant per Snrpb/d2 en zigots de ratolí, hem estat capaços de reduir el nombre d'exons exclosos i hem vist un increment en la resposta al dany en l'ADN quan les cèl·lules han estat tractades amb etopòsid. Aquests resultats suggereixen que Snrpb/d2 podria ser necessari per a l'adequat funcionament de l'splicing alternatiu dels gens implicats en la reparació de l'ADN.

A més a més, també hem descobert molts casos d'exclusió d'exons que estan conservats. Aquests corresponen, en la majoria dels casos, a isoformes alternatives de proteïnes amb canvis molt dinàmics, que es troben enriquides en vies de senyalització essencials durant el desenvolupament embrionari com; Wnt , Hippo, i modificacions de la cromatina.Finalment posem en discussió l'impacte d'aquests fenòmens conservats en el desenvolupament primerenc i detallem els dominis i les prediccions de la funció de les proteïnes implicades.

ABBREVIATIONS

DNA-Deoxyribonucleic acid
RNA- Ribonucleic acid
mRNA- Messenger RNA
RNA-Seq- RNA sequencing
RBP- RNA binding protein
shRNA- Short hairpin RNA
snRNA- Small nuclear RNA
snRNP- Small nuclear ribonucleoprotein
AS- Alternative splicing
IR- Intron retention
TE- Trophectoderm (never transposon)
ICM-Inner cell mass
PrE-Primitive endoderm
ESC-Embryonic stem cell
iPSC- Induced pluripotent stem cell
Hsa- Human (*Homo sapiens*)
Mmu- Mouse (*Mus musculus*)
Bta- Cow (*Bos taurus*)
HMC (Human, mouse and cow)
PTC- Premature termination codon
PSC- Premature stop codon
PSI- Percent spliced in
ΔPSI- Delta (change in) PSI
ORF- Open reading frame
qPCR- Quantitative PCR
Ribo-seq- Ribsome profiling sequencing
C-Cytodine/Cytosine
G-Guanine/Guanosine
T-Thymine/Thymidine
A-Adenine/Adenosine
GO- Gene ontology
UTR- Untranslated region.
CDS- Coding sequence
3'ss- 3 prime splice site
5'ss- 5 prime splice site
hpf- Hours post fertilisation

KEYWORDS

Alternative splicing
Embryo development
DNA damage

DATA

For data (raw data), figures too large to print, scripts and supplementary tables, please see look in the hard disk that accompanies this thesis. Each have a separate folders.

PREFACE

Where does it all begin?

The search for mechanisms and rules that govern the early stages in an organisms life have been studied since time immemorial. Yet, for most of this time we could not investigate the inner workings of the cell, let alone explore the tiny molecular machines within each cell and their regulation. Over the past decade, with the development of high-throughput technologies we are now able to read (or sequence) systematically the DNA of single cells, to explore the main biological dictionary of the cell. We can detect modifications to the DNA, such as changes in methylation and histone marks. Further, we can see which genes in the DNA dictionary are actually being expressed in the form of RNA, and finally we can detect protein abundances and their modifications present in embryos. This has provided a wealth of information from multiple species, and from all the "omics" levels (Transcript-; Gen-; Epigen-). This has allowed researchers to propose rules about the roles of particular genes to a particular function, or epigenetic marks to expression regulation, but overall we are still far from understanding all the molecular signatures that define the embryological system.

This thesis explores the patterns, regulation and function of alternative splicing in the embryo, taking advantage of the abundant published datasets, allowing a comprehensive analysis of changes across development and between species.

"Abandon the urge to simplify everything, to look for formulas and easy answers, and to begin to think multi-dimensionally, to glory in the mystery and paradoxes of life, not to be dismayed by the multitude of causes and consequences that are inherent in each experience- to appreciate the fact that life is complex"

Morgan Scott Peck.

TABLE OF CONTENTS

ABSTRACT ..... **iii**

RESUM ..... **v**

ABBREVIATIONS ..... **vii**

PREFACE ..... **ix**

TABLE OF CONTENTS ..... **xiii**

INTRODUCTION **1**

1. Early development **2**

Oogenesis 3

Spermiogenesis 4

Fertilisation 5

Cell cleavage 7

Compaction 9

Lineage differentiation and pluripotency 10

Blastocoel formation, Hatching and Implantation 11

2. Molecular regulation of the cell **13**

Chromatin 15

Transcription 15

Splicing 16

Alternative splicing 19

5' capping and Polyadenylation 22

Non-sense mediated decay and microRNAs 22

Protein translation 23

3. Molecular change in the embryo **25**

Epigenetic changes in the embryo 25

Zygotic genome activation (ZGA) 28

Mammalian transcriptional activation 32

Maternal mRNA degradation 33

Splicing and the embryo 36

Polyadenylation in the embryo 43
Molecular regulation of lineage differentiation. 44
Protein levels in the embryo 45
Transposable elements 46
Telomere lengthening 48
4. DNA damage and repair 49
Sources of DNA damage 50
Repair mechanisms 51
Global response to DNA damage and downstream pathways 55
DNA damage during early mammalian development 57
OBJECTIVES 67
RESULTS 69
Chapter 1. Alternative splicing dynamics in the embryo. 71
1.1 Measuring splicing during early development. 72
1.2 Splicing changes during early development 76
1.3 Transcriptomic diversity at genome activation 81
1.4 Relationship between alternative splicing and transcription 83
1.5 Single cell variation. 85
Chapter 2. AS effect on the open reading frame and protein synthesis 87
2.1 Changes to the ORF 87
2.2 Disrupted isoforms do not lead to NMD 90
2.3 Disrupting isoforms less bound by polysomes. 92
Chapter 3. Regulation of splicing 95
3.1 CIS-regulatory and kmer motif enrichment 95
3.2 Novel motif and transposon enrichment. 98
3.3 Single cell RBP/PSI correlations 100
3.4 RBP knockdown/overexpression correlations 102
3.5 Knockout phenotypes 104
3.6 Conclusion 104

Chapter 4. SNRPs and DNA repair.....107
4.1 Snrpb/d1 may regulate exon skipping in mouse.....107
4.2 Snrpb/d2 expression and protein levels.....108
4.3 Microinjection of SNRPB can prevent exon skipping. 110
4.4 DNA damage levels affected by SNRPB/D2 levels. 111
4.5 DNA repair gene expression.....112
4.6 Conclusion.....112
Chapter 5. Conservation of AS in core pathways 113
5.1 Conserved AS events.....113
5.2 Species-specific events.....118
Chapter 6. Lineage fates and pluripotency.....121
6.1 Lineage specific exons.....121
6.2 ESC specific exons.....125
6.3 Pan-pluripotent alternative splicing.....128
DISCUSSION.....131
CONCLUSIONS.....145
SUPPLEMENTARY DISCUSSION 147
METHODS.....155
Supplementary Methods (Barbara Pernaute- BP) 166
SUPPLEMENTARY FIGURES 169
ANNEX.....213
Annex 1. Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals. 215
Annex 2: Deconstructing Superorganisms and Societies to Address Big Questions in Biology. 233
Annex 3. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. 247
Annex 4. The bladed Bangiales (Rhodophyta) of the South Eastern Pacific: Molecular species delimitation reveals extensive diversity 257

Annex 5. The alternative splicing factor Nova2 regulates vascular development and lumen formation 273

Annex 6: Amphioxus functional genomics and the origins of vertebrate gene 291

BIBLIOGRAPHY 305

CONTRIBUTIONS 333

ACKNOWLEDGEMENTS 335

INTRODUCTION

In this thesis, I first introduce the biological system under study, from the progenitor cells through gamete formation, fertilisation and cleavage stages until the implantation of the blastocyst (Section 1). This overview explains the complexity of the early mammalian cells in terms of morphology and function, mainly building from knowledge acquired before the advent of next generation sequencing. Second, I briefly introduce the molecular dogma of biology, with special emphasis on splicing, specifically addressed in this thesis (Section 2). Third, I introduce the fundamental molecular changes that occur during early development, from the rewiring of the epigenetic marks, global de- and re-methylation, transcriptional activation and protein synthesis with a more detailed explanation of what is currently known about splicing during embryo development (Section 3). Then finally, I introduce what is known of DNA damage in the embryo and the links to splicing and transcription (Section 4), to better understand why splicing could be affecting the DNA damage pathway in early embryos.

1. Early development

Embryo development has been studied since the beginning of our documented history, thousands of years before the discovery of DNA, RNA or even cells. The first known records are from Hippocrates and Aristotle, who described the morphological progression of embryos, and proposed the 'preformatory' and 'epigenesis' developmental models. The 'preformatory' model suggested early embryos were fully formed adults in miniature (Figure I.1.i) that need to be unfolded and expanded, whereas the 'epigenesis' model suggested that organisms arise from unformed material and develop gradually over time (Callebaut 2008).

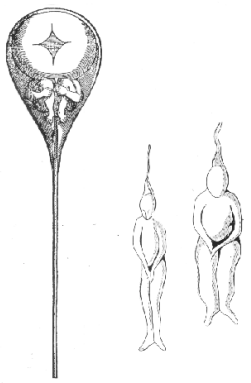


Figure I.1.i. The 'homunculus'. Drawn by Nicolaas Hartsoeker in 1695. Envisioned a miniature human curled up into the sperm, as men were considered the source of life and women merely the incubator.

It took until the 18th century however, and with the development of cellular theory through the advancement of microscopy technologies, that the preformatory model was rejected and scientists (notably by Karl Ernst von Baer, "the founding father of embryology") conducted more descriptive analyses in multiple vertebrates, discovering the blastocyst, notochord and germ layers to name a few (Abzhanov 2013). These studies could show that life begins in a single maternal egg cell (or oocyte) and paternal sperm, that fuse, then divide, forming a mass of single cells that self organise to form the primitive lineages of cells, that will develop slowly into all adult organs and tissues.

In this section, I outline the main features of the developing embryo in mammals, from the formation of eggs and sperm (gametogenesis) to embryo implantation into the uterus (Figure I.1.ii), mainly focusing on the morphological and physiological

changes, a lot of which has been known since the 1800s (For molecular regulation- go to Chapter 3). It is also important to note that there are many differences between mammals, especially in the timing of events, for example a mouse oocyte takes around 5 days until implantation, whereas a human oocyte takes 7 and cow 8-9, yet all contain generally the same main feature of which are introduced here.

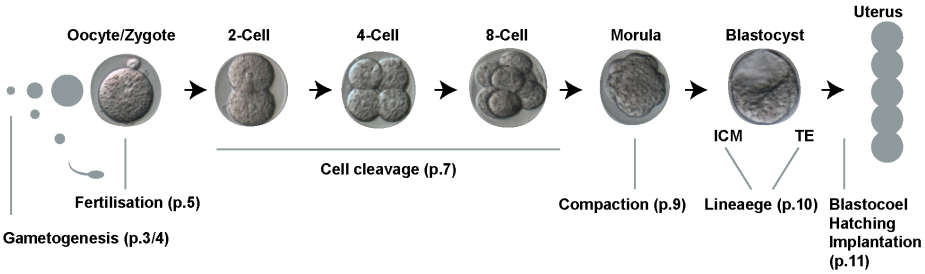


Figure I.1.ii Generalised overview of early mammalian pre-implantation development. Starting on the left, progenitor cells go through the process of gametogenesis (p17/18) to form an egg and sperm, these fuse at the stage of fertilisation (p19) and triggers the initiation of the cell cycle (p22), where cells divide without growth to the morula stage where the embryo can be seen to compact (p24). At the blastocyst stage (right), we have a round ball of cells with a fluid filled cavity or blastocoel (p26) and the emergence of lineages (p24), either trophectoderm (around the surface) or inner cell mass (ICM; within one side of the embryo), before the embryo hatches and implants/attaches to the uterine wall (represented as a grey structure on far right; p26). Embryo images from mouse (with permission, Barbara Pernaute)

Oogenesis

Mammalian oocytes are formed from the primordial germ cells, where before or shortly after birth, oogonia (egg progenitors) are formed, enough to supply a female for life (Hartshorne et al. 2009). These then initiate steps of meiosis I, which is a special type of cell division with crossing-over occurring between non-sister chromatids, exchanging genetic material between homologous partners. At this stage, the cell arrests in the diplotene stage of prophase where they grow larger in size. At puberty, production of follicle stimulating hormone (FSH) in mammals (human monthly-cycles; mouse every ~5 days) causes a subset of primary follicles to develop into a single (or few) egg cells. These can then complete meiosis I to form a haploid secondary oocyte, with two sister chromatids per chromosome (but no homologous pair), along with the small 1st polar body, a tiny cell containing half the genomic material, which is later degenerated (Gilbert 2000). Secondary oocytes then initiate meiosis II, stalling at metaphase II for around 24 hours, where ovulation is initiated

by FSH and Luteinizing Hormone (LH) (Marieb 2013), and the oocyte is released into the fallopian tube, until potential fertilisation. At this point the oocyte has just one chromatid of each chromosome (Gilbert 2000), is surrounded by a proteinaceous extracellular matrix called the 'zona pellucida' and is massive in size (~120 micrometres) with an enormous cytoplasm filled with proteins and mRNAs to sustain its continued development (Trounson, Anderiesz, and Jones 2001). An overview is shown in Figure I.1.iii.

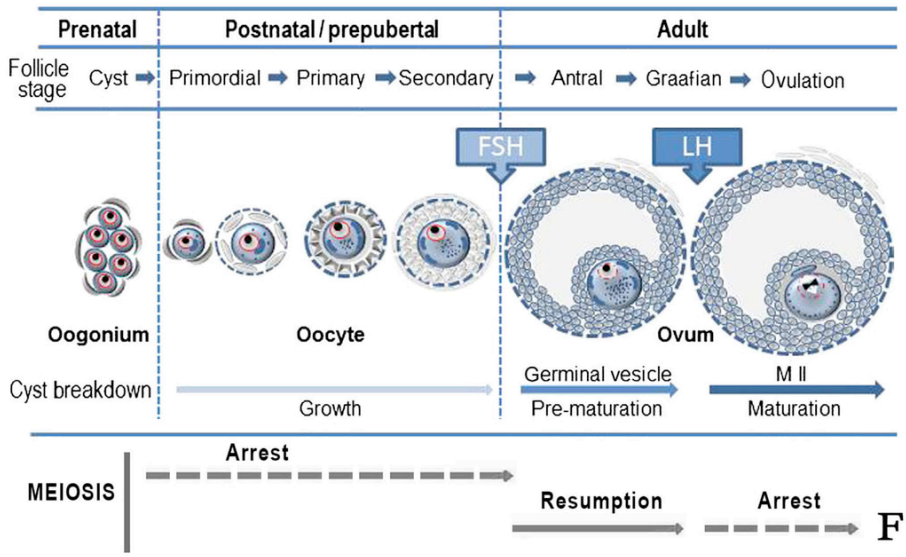


Figure I.1.iii Development of the mammalian oocyte within the ovary. Showing the prenatal, postnatal and adult phases in development and an approximation of the structure and size differences at each stage. (Plant, Zeleznik, and O'Shaughnessy 2015). FSH: Follicle stimulating hormone. LH: Luteinizing Hormone.

Spermiogenesis

Sperm are formed in the seminiferous tubules of the testis, where spermatogonial stem cells divide by mitosis (conventional cell division), producing two daughter cells. One will replace the stem cell progenitor and the other will form a primary spermatocyte. This then divides again by meiosis-I into secondary spermatocytes (again with crossing-over), which further divides by meiosis-II into four haploid spermatids. These radial spermatids then mature into spermatozoa (Figure I.1.iv) with their characteristic polarity, acrosome head (a cap like organelle, that contains digestive enzymes), tail and a condensed genome. Sperm are extremely small with

tiny cytoplasms, which suggests that they may only contribute their genetic material and activate the egg (see next section). Finally, sperm are transported to the epididymis where they become motile before being ejaculated.

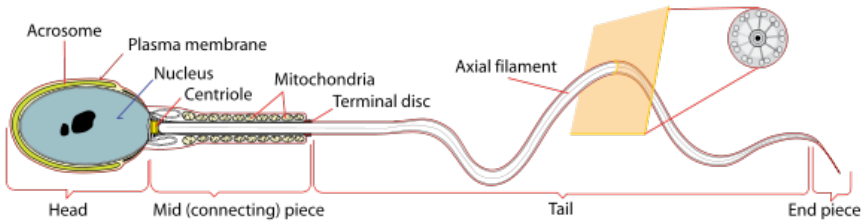


Figure I.1.iv. Structure of sperm, showing characteristic features (Figure from Mariana Ruiz; public domain).

Fertilisation

Once the sperm arrives in the female reproductive tract, it must undergo a few changes before it can fertilise the egg. First the sperm must 'capacitate', where signals from the uterus (including heparin and albumin) change sperm membrane permeability by removing cholesterol (Eliasson 1966), allowing an influx of calcium that leads to an increase in sperm motility and destabilisation of the acrosome head membrane, later allowing it to penetrate the egg (Ikawa et al. 2010).

Next, the sperm travel up the mucus-lined female reproductive tract to reach the oocyte at the top of the fallopian tube. Most sperm do not make the journey, but those that do can initiate the acrosome reaction, where sperm release a cocktail of enzymes to create a hole in the zona pellucida of the egg, which is a proteinaceous layer that surrounds the oocyte (Figure I.1.v).

The sperm can recognise the egg by Izumo1 protein on the cell surface of the sperm (N. Inoue et al. 2005) and the Juno (Folr4) receptor on the egg's surface (Bianchi et al. 2014). Once a single sperm has fused with the oocyte, the egg expels its Juno receptors into extracellular vesicles to prevent multiple entry of sperm or polyspermy (Bianchi et al. 2014). The sperm also initiates the cortical reaction, where cortical granules inside the oocyte fuse with the membrane, releasing enzymes that modify glycoproteins in the zona pellucida forming cross-links, thereby creating an

impermeable boundary to additional sperm (Austin 1956; Bleil, Beall, and Wassarman 1981).

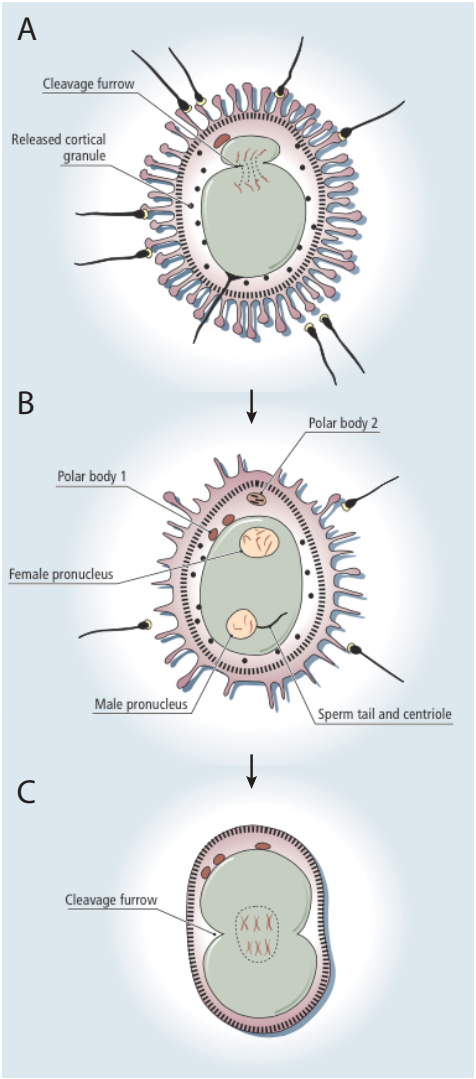


Figure I.1.v. Fertilisation

A: Fertilisation by sperm (black) with acrosome heads (yellow), allowing a single sperm to pass the zona pellucida to enter the egg. This leads to calcium dependant activation of the egg, forcing the second polar body to be extruded and release of cortical granules.

B: The two pronuclear genomes are both present in the egg where their nuclear membranes break down.

C: Chromosomes of the two parents can fuse, ready for duplication and the first mitotic cell division. Edited from (Johnson 2007)

At this stage, in most mammals the sperm fully enters the oocyte. The known exception being hamster, where the midpiece of the sperm that contains the mitochondrion is kept outside of the oocyte (Sato and Sato 2013). Since the majority of mammals have maternal inheritance of mitochondria, it is known at this stage that the paternal mitochondria are targeted for degradation through a number of mechanisms (Sato and Sato 2013). In mouse and cow, sperm mitochondria are

ubiquitinated before fertilisation, leading to their degradation by the proteasome and/or lysosomes by the 4-8 cell stage (Sutovsky et al. 2000), whereas in hamster it has been suggested the midpiece (Figure I.1.iv.) may accumulate vesicular bodies / lysosomal vacuoles (Sutovsky et al. 2000).

In addition, sperm entry causes calcium release into the oocyte, that oscillates through the cell, and is triggered by a sperm specific gene PLCZ1/Plcz1 (phospholipase C) (Saunders et al. 2002). This allows the oocyte to exit its metaphase II arrest and complete the second meiotic division, leading to the extrusion of the second polar body and a haploid oocyte (Johnson 2007). At this stage, two nuclear membranes exist in the oocyte, one from the mother and the other from the sperm; called the 'pronucleus' stage. The paternal genome must first open up its chromatin, before each pronucleus can replicate their DNA separately. Finally, the nuclear membranes breakdown during prometaphase, allowing the paring of the homologous pairs of chromosomes, ready for the first cell division (Szabo and O'Day 1983).

Cell cleavage

Cell division precedes though the recognisable steps of mitotic cell division as shown in Figure I.1.vi, yet the timing and length of the cell cycles is one of the main species-specific differences in the early embryo.

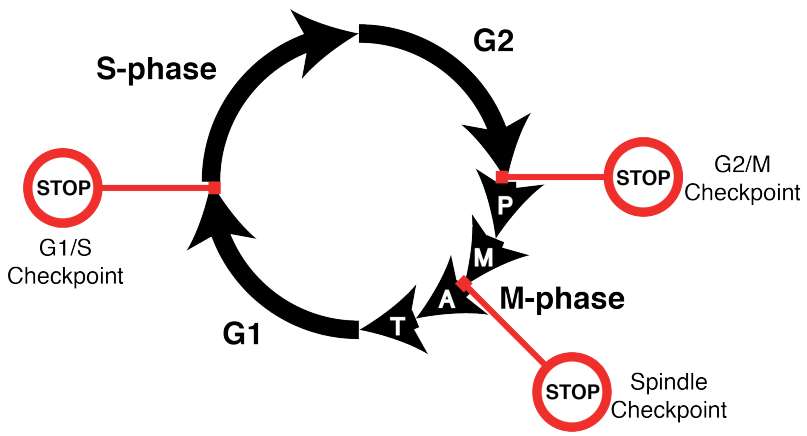


Figure I.1.vi. Mitotic cell cycle. Starting at gap phase 1 (G1) a period of cell growth, before S phase where nuclear DNA is replicated, before another gap phase (G2) of growth, and finally, initiating of cell division in M phase, of which it is split into P=Prophase, M-Metaphase, A-Anaphase and T-Telophase. Cell cycle checkpoints are listed.

In mouse, we know the first two cell divisions take roughly 18-20 hours each (Ciemerych, Maro, and Kubiak 1999), which is approximately comparable to the typical eukaryotic differentiated cell (Cooper and Hausman 2007), however the proportion of time in the four main stages of mitosis differ. In typical cells, G1 may last 11 hours, S for 8 hours, G2 for 4 hours and M for 1 hour, and in the first division these proportions are roughly equivalent, yet in the second division, G1 lasts only 0.5-1.3hours, S for just 1-5 hours, then an extremely long G2 at 12-16 hours and a relatively normal M phase at 70 minutes (Palmer and Kaldis 2016). It is at this long initial G2 phase that the genome is first activated in mouse (Artus and Cohen-Tannoudji 2008). The next cleavage divisions from the 2-Cell embryo to the blastocyst are shorter, averaging just 11 hours per division (Artus and Cohen-Tannoudji 2008) and noticeably have smaller time spent in G1 or G2 (10–14 h; G1: 1–2 h, S:7 h, G2/M: 1–5 h).

In human, much less research has been conducted, but the first division takes 24 hours, followed by a slightly longer cycle to the second division (27 hours), before returning to 24 hours thereafter (Figure I.1.vii). The length of each phase has not been well documented.

-

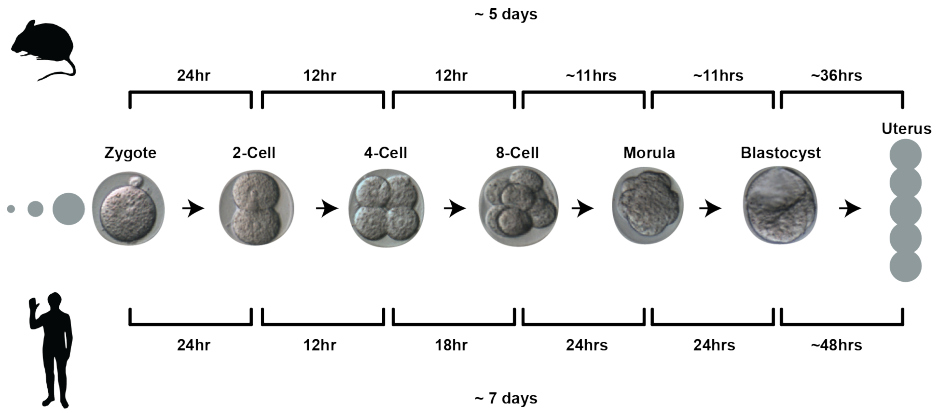


Figure I.1.vii: Mouse and human pre-implantation cell division timings. Starting from left to right, from the fertilised zygote, 2-Cell stage, 4-Cell stage, 8-Cell stage, Morula, Blastocyst and stage of implantation into the uterus. Timings taken from (Wong et al. 2010).

The short G1/G2 growth phases in mouse may be due to the absence of size growth during this time, given the zona pellucida surrounding the embryo prevents nutrient

uptake and hence each cell division leads to a decrease in volume by half. In addition, it is known that increasing the growth phase can induce differentiation into neuroepithelial cells (Calegari 2003), suggesting that cell cycle control may in part regulate pluripotency (see page 10).

At the blastocyst stage, the epiblast (cells that go on to form the embryo proper) are known to divide every ~9hours (in mouse), similar to mouse embryonic stem cells (mESCs; 12 hours) and human embryonic stem cells (hESCs; 15hours), which are derived from the inner cell mass (ICM). This makes them one of the fastest dividing cells in human/mouse (Hindley and Philpott 2013). They are also characterised by their much longer S phase in proportion to G1 (Hindley and Philpott 2013). In the trophectoderm (the cell destined to become the placenta), it is known that the mural trophectoderm (those farther from the ICM) divide much slower than the polar trophectoderm, and this may relate to proximity to the ICM (Artus and Cohen-Tannoudji 2008), this change results in cells of the polar trophectoderm moving around to the mural side to keep equal tension across the layer (Artus and Cohen-Tannoudji 2008).

Compaction

The first salient, multi-cellular, morphological change in the embryo comes at the 8-16-cell stage (mouse), 16-cell (cow, human), 64-cell stage (rabbit) to early blastocyst (pig) (Fleming et al. 1993; Koyama et al. 1994; Reima et al. 1993), where the cells that have up to now been spherical and quite distinct from one another, bind tightly together with the formation of desmosomes and tight junctions, making cells difficult to distinguish (Pratt and Ziomek 1982), as shown in Figure I.1.ii (mouse, 8-cell to morula transition). In addition, before compaction, microvilli are distributed evenly across the surface of the blastomeres, but at compaction they accumulate on the apical surface away from the junctions (Ducibella et al. 1977; C. B. Li et al. 2011).

Lineage differentiation and pluripotency

From the zygote to first few cleavage stages, all blastomeres of the embryo appear equal, and have the capacity to form all later cell types and tissues (Tarkowski and Wróblewska 1967; Condic 2013). These cells are called 'totipotent'. In mouse, from 8-cell onward however, some cells are positioned on the outside or inside of the cluster, which is the trigger for the first restriction in cell fate (Tarkowski and Wróblewska 1967) (and potentially other mammals).
The cells on the outside form the first epithelial layer in mammalian development, the trophoblast or trophectoderm (Wiley 1988), and are polarised with a specific distribution of organelles along the apical-basal axis (Wiley 1988). The trophectoderm allows the embryo to invade (e.g. human) or attach (e.g. mouse) to the uterus, and forms the embryonic section of the placenta. Trophoblast cells gradually lose totipotency, as it has been shown that cells taken from the morula in mouse, can only give rise to trophectoderm lineage cells (Papaioannou 1982). In human however, early blastocyst trophectoderm cells can develop into ICM cells (De Paepe et al. 2013), suggesting differences exist between species of when cells become restricted to their fate.

The inner 'unpolarized' cells congregate at one pole of the blastocyst (inner cell mass; ICM) and are fewer in number compared to trophoblast (Handyside 1978). Similar to TE, these cell become committed to their lineage (Papaioannou 1982; Handyside 1978). These cells are called 'pluripotent' as they give rise to all the cells of the body, but not the TE-derived placental tissues. At a later stage the inner cell mass then splits into two different lineages, known as the epiblast and primitive endoderm (or hypoblast), which segregate into two distinct layers within the embryo. The epiblast (the remaining pluripotent tissue) will produce the embryo proper, and primitive endoderm the amniotic tissues, which surround the growing foetus.

From the inner cell mass we can derive embryonic stem cells (ESCs), which can be grown *in vitro* and be differentiated into adult cell types, by specific culture conditions and induction of specific transcription factors (Mummery 2003). These have the potential for use in regenerative medicine, indeed stem cells are already used in a limited number of clinical situations (Watt et al. 2010). Alternatively, due to the

ethical and practical reasons (availability), the development of induced pluripotent cells (iPSCs) was developed, which are adult somatic cells that have been "reset" back to the pluripotent stage, at which a cell can differentiate into other cell types. In this way, scientists can produce potentially unlimited cells of different types.

Blastocoel formation, Hatching and Implantation

At around the morula stage, sodium pumps in the trophectoderm move sodium from the exterior to the surface (i.e. from non-polar to polar side), which in turn draws water osmotically into a growing cavity (Plachot 2000). This action makes the embryo resemble a hollow ball of cells and increases the overall diameter of the embryo, forming what is called the 'blastocoel', and the whole embryo, a 'blastocyst'. This requires a lot of energy, and explains why there are pronounced differences in metabolism between the inner cell mass and TE (Houghton et al. 2003; Houghton 2006).

Once the blastocyst has reached the uterus, it has reached a size that will no longer fit within the zona pellucida. At this point the trophoblast cells release a trypsin-like protease called strypsin, which creates a hole in the zona from which the embryo can 'hatch'. It is important to hatch out of the zona when near the uterus, as this is the required area of implantation, earlier implantation in the fallopian tube causes tubal pregnancy, which has the potential to cause haemorrhage (Stanger et al. 2001).

Once the blastocyst has emerged, integrins on the cell surface of the trophoblast cells are recognised by the endometrium (lining of the uterus), which contains collagen, fibronetin and laminin, which bind to the embryo. The trophoblast cells of the embryo then release a cocktail of protein digesting enzymes enabling it to fully implant. Implantation allows the embryo to receive nutrients from the mother through the placental tissues, whilst also protecting the embryo from immunological rejection. At implantation, mouse trophectoderm cells are roughly 25-40 micrometres in diameter (Van Den Abbeel et al. 2013), and inner cell mass cells around 10 micrometres, which is among the smallest of any known mouse cell type. Therefore implantation is precisely timed with size possibly acting as a control on the development of the blastocyst.

2. Molecular regulation of the cell

Since the ground-breaking work of Avery–MacLeod–McCarty in the discovery of DNA, Watson and Crick in the discovery of the structure of DNA, Crick and Brenner for the discovery of the genetic "triplet" code, Severo Ochoa for synthesizing RNA in vitro for the first time, and many more important breakthroughs, scientists in mainly the 50s and 60s managed to piece together the flow of information of the cell from its DNA 'dictionary' to its final RNA/protein 'effector molecules'.

At its core, the central dogma of molecular biology first proposed by (Francis Crick 1958) outlines the modes in which information can be transmitted, from DNA to RNA (transcription), RNA to protein (translation), DNA to DNA (replication). In special cases RNA can be reverse transcribed to DNA (e.g. retroviruses) and RNA can self-replicate (e.g. viruses). Finally, DNA can be converted directly to protein in a cell-free artificial environment (McCarthy and Holland 1965). These relationships are shown in Figure I.2.i. (F Crick 1970).

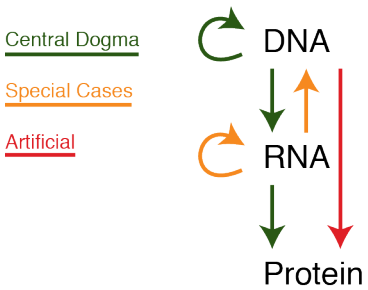


Figure I.2.i. Central dogma of molecular biology. Concept from (F Crick 1970). Colours show the three categories of possible movement between the main polymers of molecular biology.

Since these times, the central dogma has remained true (since the 1970's description), yet the complexity and the additional layers of regulation that have been discovered since have been enormous. In this section, I briefly introduce the main stages in the 'modern' dogma, from chromatin regulation, transcription, splicing, poly-adenylation and protein synthesis, with particular emphasis on splicing regulation (Figure I.2.ii).

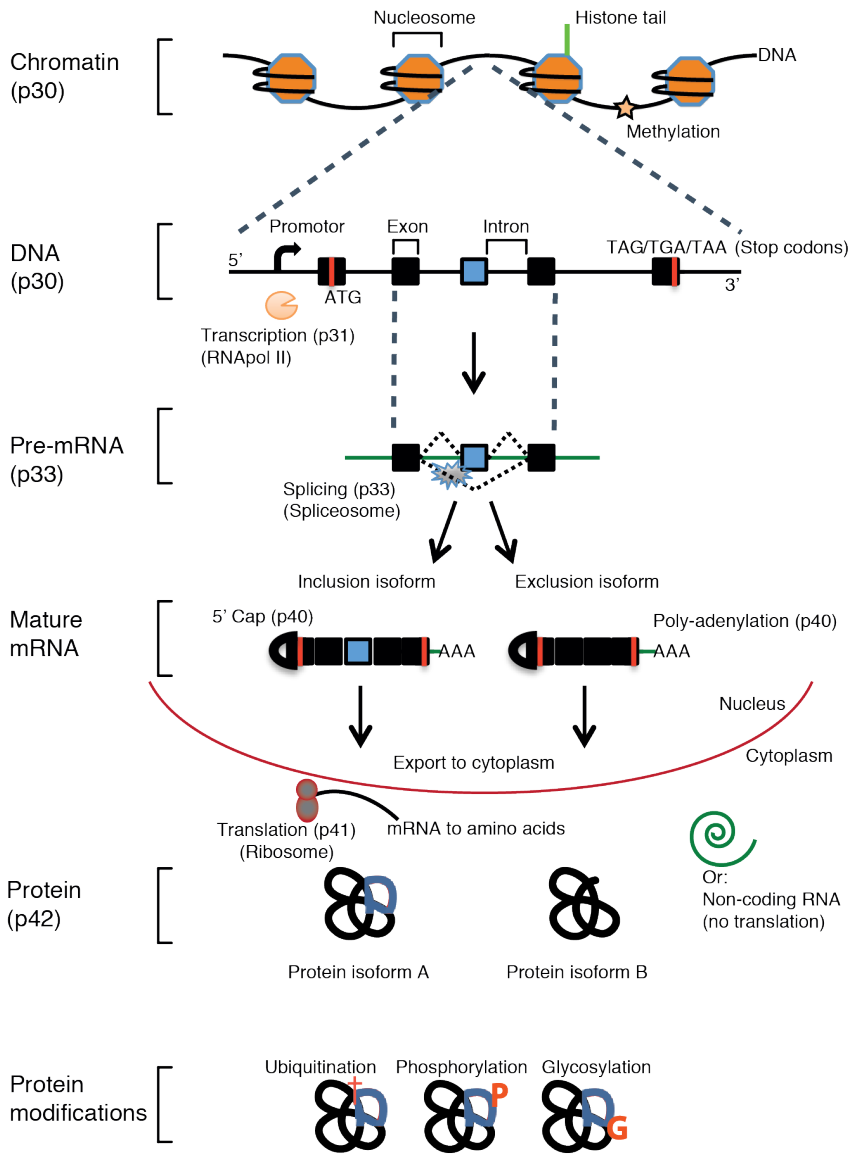


Figure 1.2.ii. Expanded 'modern' dogma, from chromatin, DNA, RNA to protein. Chromatin (1st level) consists of DNA (black line) wrapped around nucleosomes (orange), two major modifications are DNA methylation and histone tail changes. Within the DNA (2nd level) genes are located that contain exons (coding sequence; black boxes) and introns (lines), start and stop codons are listed (red lines show location), with a promoter at the 5' end (arrow). Pre-messenger RNA (3rd level; pre-mRNA) shows a part of the transcribed product of the above gene, RNA is represented as a green line with three exons. The pre-mRNA can then be spliced differently into two different forms (with the blue exon (included) or without (skipped) (4th level). These can then be polyadenylated and 5' capped before export into the cytoplasm and translation into proteins (5th level) by a ribosome. Some mRNAs do not need to be translated, are 'non-coding' and have a functional three-dimensional structure (e.g. U1 RNA). Finally, proteins can acquire different post-translational modifications (6th level, a few listed) and be folded into different 3-dimensional shapes.

Chromatin

The chromatin consists of DNA, histones and additional proteins, which are folded into the 20, 23 or 30 pairs of chromosomes in mouse, human and cow respectively. The histones and DNA are organised into a series of 'nucleosomes', which are a collection of 8 histone units (two of H2A, H2B, H3, H4), are wrapped around which 146bp of DNA (Figure I.2.ii). It is important where these nucleosomes are positioned for accessibility of trans-acting factors to promote or repress transcription (M. J. Liu et al. 2015), and this can depend on the epigenetic modifications, which are heritable modifications that do not change the base nucleotides.

The histones can be changed in a wide variety of ways (including acetylation, methylation and phosphorylation) on the N terminal tails. In addition, nucleosomes and their modifications can directly recruit trans-acting factors to the chromatin, so their identity and state can determine how transcriptionally active a particular gene is. For instance, histone acetyltransferases (HATs) can acetylate lysines 9, 14 and 28 in the histone tail of H3 opening up the chromatin and histone deacetyltransferases can reverse this reaction, their balance being key to accessibility.

In addition, DNA itself can be methylated by the addition of CH_3 group to cytosine residues in the DNA specifically at CpG (cytosine and guanine) sites (ones where G follows C in the linear sequence). Methylation in the promoters of genes universally leads to lower expression or silencing (Zemach et al. 2010). It is also used to repress transposable elements (Huff and Zilberman 2014) and for the inactivation of the X chromosome in females to ensure just one is actively transcribed (Sharp et al. 2011).

Transcription

Transcription is the production of RNA using a DNA template and is performed by RNA polymerase II for messenger RNAs (mRNAs; standard protein coding), most small nuclear RNAs (snRNAs; [including the non-coding splicing core RNAs, U1/U2/U4-6, see later]) and microRNAs (non-coding; involved in RNA silencing). Given accessibility of the gene by changes in the chromatin, the next requirement is for a trans-acting factor (or transcription factor) to have access and bind to specific regulatory-elements/enhancers (specific regions of DNA that can be thousands of base pairs away from the target genes) and promoters (located directly upstream from

the target gene; right angle in Figure I.2.ii). This complex of enhancer and promoter bound in a particular conformation allows the RNA polymerase to bind and use its helicase activity to "unwind" the DNA 14 base pairs. Using the single stranded DNA template the polymerase can recruit the four RNA bases (AGCU; where Thymine 'T' is replaced with Uracil 'U'), in the process of 'transcription elongation'.

The presence and activity of specific transcription factors is key to understand how particular sets of genes (or programs) are regulated, and why in developmental time and cell type, each has a particular transcriptomic identity.

Finally, transcription itself does not solely dictate the levels of overall mRNA abundance, as decay rates need to be considered. This can happen passively (e.g. through differential polyadenylation decay rates [See: p.22]), actively (e.g. microRNAs p.22) or in response to premature stop codons (See: nonsense mediated decay: p.22). In mammals a transcript can last from several minutes to a few days (Sharova et al. 2009). There are multiple, functionally redundant endo-ribonucleases and exo-ribonucleases (conserved back to bacteria), that allow transcripts to be degraded that are no longer needed, preventing a build up of non-functional RNA (Houseley and Tollervey 2009).

Splicing

In the vast majority of cases, a transcribed protein-coding precursor mRNA (pre-mRNA) cannot be immediately exported from the nucleus and translated, as the pre-mRNA is interspersed with non-coding nucleotides called 'introns' between the protein-coding 'exons' (Figure I.2.ii). Splicing is the process by which these exons are joined and introns removed, producing the mature mRNA. By removing the introns the mRNA can be considered 'in frame' (protein-coding isoforms only), with the exonic nucleotides in the correct trinucleotide format to produce a viable protein sequence, this is because introns change the correct trinucleotide frame and/or contain stop codons (TAA/TAG/TGA), which inform the protein producing ribosome to detach from the mRNA and end the protein. Around 95% of protein coding genes have introns (Hubé and Francastel 2015) and these usually must be removed in order to have a functional full-length protein, although some truncated proteins can have function, many of which act through self-repression of its own full length protein (Coers, Ranft, and Skoda 2004; Karimi et al. 2014; Tatsumi et al. 2015). In a similar

way, mutations can lead to truncated proteins, and again many often have dominant negative roles to the full-length protein (Candeias, Hagiwara, and Matsuda 2016).

In order for splicing to proceed, exon-intron boundaries are recognised by the presence of specific sequences (5'splice site: AGGTRAG and 3' splice site: YAG, Figure I.2.iii, Top) by the spliceosome, which is composed of around 90 core proteins, the SM ring and 5 snRNAs (U1, U2, U4-6; Figure I.2.iv). These work in a series of complexes to cleave the introns and join two opposing exons together (Will and Lührmann 2006). The spliceosome reaction happens in a highly dynamic series of stages, which I very briefly outline in Figure I.2.iii.

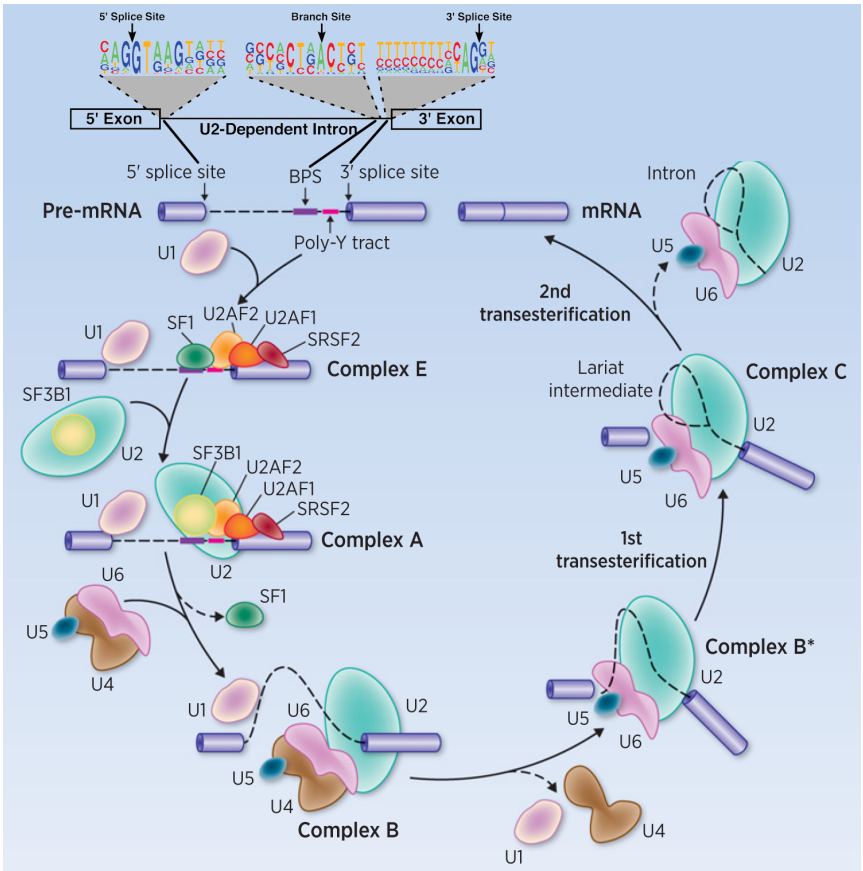


Figure I.2.iii Splicing dynamics. **TOP:** U2-spliceosomal introns are recognised by specific components of the spliceosome. The 5' splice site is recognised by U1 that binds at a 'AGGU' motif, the branch point site 'A' (BPS; 18-40 nucleotides from the 3' splice site) by U2, the poly-pyrimidine tract (Poly-Y; U rich) by U2AF2 and the 3' splice site 'NAG' motif by U2AF1 to form the E complex. **BOTTOM:** The presence of U2AF1/2 allows the U2 complex (which contains SF3B1) to bind to the branch-point-site (or BPS), removing SF1, and forming the A-complex. Next, the tri-snRNP complex of U4-6 is recruited, the spliceosome is remodelled and the U1/U4 complexes are released to create the catalytically active B-

complex which initiates the first transesterfication reaction, where the intron is cleaved from the 5' intronic GU sequence forming a loop/lariat connecting the cleaved end to the branch-point. Now the spliceosome is in the C-complex and the lariat intermediate before the second transesterification reaction occurs, removing the lariat and fusion of the two exons. At around 20-24 nucleotides from the 5' end of the splice junction the exon junction complex (EJC) is deposited to form the mature ribonucleoprotein (mRNP), which remains bound through export into the cytoplasm. (Figure adapted from (Yoshimi and Abdel-Wahab 2017) and (Padgett 2012).

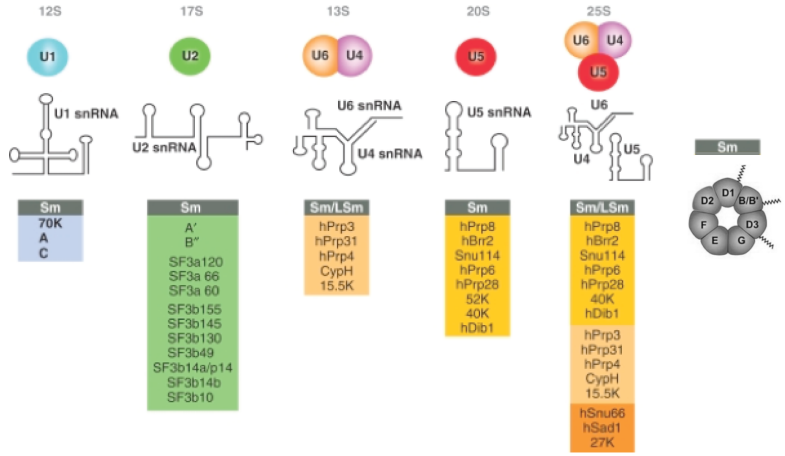


Figure I.2iv. SnRNAs in each spliceosomal complex. For each small nuclear RNA in a complex (top), each has a specific secondary structure, and each is associated with an Sm or LSm (Sm-like) ring required for biogenesis and metabolic stability of the snRNPs (snRNA and protein complexes) (Urlaub et al. 2001; Saltzman et al. 2011). The Sm ring (right) is a heteroheptameric complex comprising SNRPB, SNRPD1, SNRPD2, SNRPD3, SNRPBE, SNRPF and SNRPG (Will & Lührmann 2006). Apart from its role in the core spliceosome, SM ring proteins have been linked to other functions in the cell such as chromosome telomere replication (Seto et al. 1999), histone mRNA 3' end formation (Schümperli & Pillai 2004). Figure adapted from (Will and Lührmann 2006). Structure of the standard SM ring (Schümperli and Pillai 2004)

Finally, splicing is generally considered to be co-transcriptional, in that the spliceosome is known to bind the nascent pre-mRNA in the nucleus (Beyer, Bouton, and Miller 1981) and begin the splicing process before the RNA polymerase terminates (Merkhofer, Hu, and Johnson 2014). This discovery led to a better understanding of the link between splicing and transcription, and indeed this was subsequently shown in several key papers. (Carrillo Oesterreich, Preibisch, and Neugebauer 2010) could show in Saccharomyces cerevisiae that the RNA polymerase pauses before the terminal exon in order for it to be spliced, (G. Roberts 1998) could

show that transcription elongation rates can regulate alternative splicing, (E. Kim et al. 1997) that RNA polymerase II could recruit splicing factors promoting co-transcriptional splicing and (Muñoz et al. 2009) could show that ultraviolet irradiation could affect co-transcriptional splicing, through hyper-phosphorylation of RNA polymerase II, leading to slow down of the polymerase and altered exon inclusion (in many downstream apoptosis genes). In addition, it has been observed that chromatin structure can affect splicing, directing the spliceosome to exon-exon junctions (Tilgner et al. 2009), and can affect the speed of the polymerase in turn, changing inclusion levels (Hnilicová et al. 2011).

Alternative splicing

In itself, the process of splicing that has just been described sounds an expensive and wasteful task, to produce long mRNA sequences with large amounts of discarded material (introns). However, alternative splicing (to which most of this thesis is dedicated), provides a clear, adaptive, evolutionary advantage to the necessity of intron removal. Through alternative inclusion of different exons from a single pre-mRNA, multiple viable protein sequences can be synthesized (Figure I.2.v; left) or isoforms with frame-shifts (Figure I.2.v; right). Frame-shift isoforms can lead to truncation or non-sense mediated decay (NMD), which can change protein abundance (if the truncated isoform is unable to be translated or actively degraded). Alternative splicing of in-frame exons, can include/exclude specific domains of the protein, that can affect the protein's location in the cell (Martin and Ephrussi 2009), activity (Mosley and Keri 2006) or its interactions with other protein partners (Narykov, Johnson, and Korkin 2018). If the exon is in a RNA/DNA binding domain, it can change the protein's specificity and therefore its target gene's expression levels (Gabut et al. 2011). Finally, alternative isoforms can be synthesized in different tissues or at different developmental times, allowing a single gene encoded in the genome to have different biological activities depending on the cell type or stage (Barbosa-Morais et al. 2012). This allows evolution to specialise a single protein to multiple functions without affecting the proteins original function in another cell type (Bush et al. 2017).

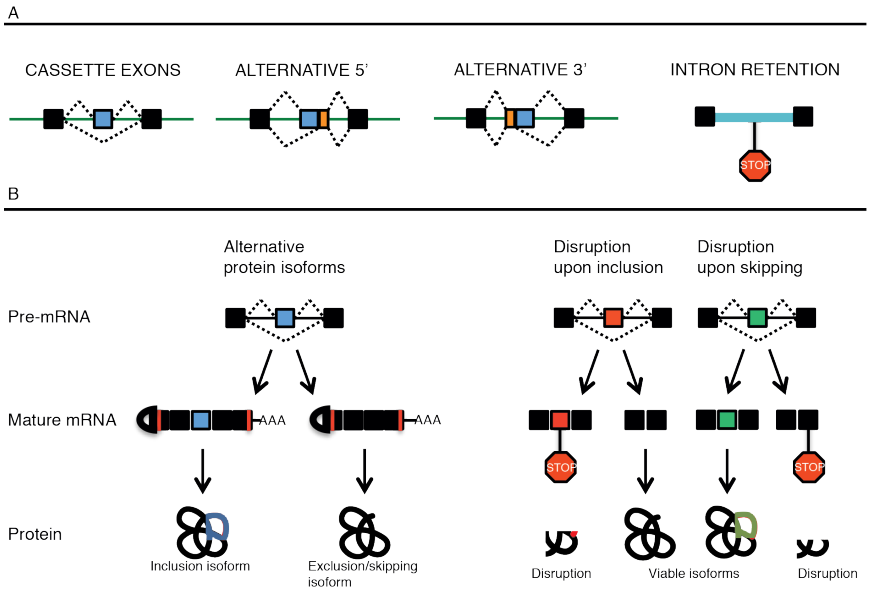


Figure I.2.v. **Alternative splicing types (A) and possible outcomes from cassette exons (B)**. **A** Four types of splicing dealt with in this thesis. **B**: Flow chart showing the possible outcomes of splicing, using a simple model pre-mRNA (three focal exons). **LEFT**: shows mRNA two isoforms being produced from a single gene, leading to translation of two alternatively spliced protein isoforms, with or without the blue exon. This difference can lead to different functions, localisations, interactions or activity. **RIGHT:** shows simplified alternative scenarios, disruption upon inclusion/skipping, where differential splicing can lead to premature stop codons (PSC) being included into the mature mRNA or skipped. For example, if the red exon is included into the final transcript it will introduce a PSC in the middle of the transcript, whereas skipping will allow the transcript to be "in frame" (able to produce a full length protein). In contrast the green exon will have the opposite effect, where inclusion is required to prevent a premature stop codon being included. The premature stop codon can lead to a truncated protein or to an NMD target, which could lower the amount of total available transcript for translation.

Alternative splicing is regulated by trans-acting RNA binding proteins (or 'splicing factors'), which normally have an RNA binding domain that allows the protein to bind to exonic/intronic enhancers/silencers (ESE, ESS, ISE, ISS) in the pre-mRNA leading to inclusion or exclusion of an exon/intron into the final mRNA (Figure I.2.vi). For example, SRRM4 (also called nSR100) is specifically expressed in the neurons of vertebrates, and permits the inclusion of microexons (considered those between 3 and 27 nucleotides) (Irimia et al. 2014). SRRM4 acts by binding to UGC motifs located directly upstream of its target exons and this affects a program of genes that are involved in neuronal differentiation, synapse, vesicle and ion channels (Irimia et al. 2014).

Around 95% of multi-exonic genes can be alternatively spliced in human tissues (Pan et al. 2008), which shows the potential of splicing to regulate many processes and there are greater than 200 RNA binding proteins that could play a role in tissue-specific splicing programs (Bebee et al. 2015). Although many of these RNA binding proteins could have a role in mRNA targeted degradation, RNA post-processing, and do not have to be involved in splicing alone, or at all.

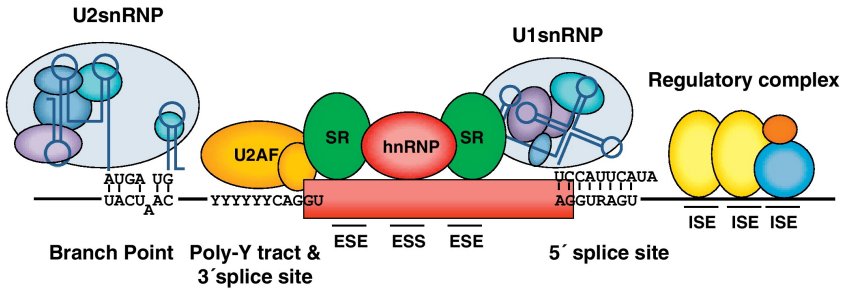


Figure I.2vi Alternative splicing regulators. Schematic showing positions of the core snRNP complexes U1 and U2, with hypothetical positive acting regulators (e.g. SR- Serine/Arginine-repeat containing) bound to specific sequences in the exonic splicing enhancer (ESE) and negative acting regulators (hnRNP) bound to exonic splicing silencers (ESS). In addition/instead RBPs can bind to intronic enhancer or silencer regions (ISE, ISS). Depending on the assembled RBPs, more skipping or more inclusion can be attained. Figure adapted from (Irimia and Blencowe 2012).

Alternative splicing can be measured by in a variety of ways, for instance, Tophat/Cufflinks (D. Kim et al. 2013) maps short RNA-Seq reads to the genome, in a splice aware mode (taking account that reads may map across from one exon to the next), then uses the counts to estimate the number of reads mapping to each isoform, whereas vast-tools (Irimia et al. 2014), Sanjuan (Valcarcel lab, CRG), use exon-exon, exon-intron junctions (created using multiple tissues for each species), and maps RNA-Seq reads to them to determine exon inclusion directly from the ratio of reads spanning the inclusion and exclusion junctions. This strategy results in percent-spliced-in values (or PSIs), which span from 0 to 100 % inclusion of an exon, and percent-intron-retention (or PIR), which also span from 0 to 100% inclusion of an intron.

5' capping and Polyadenylation

5' capping is required in most mRNAs to stabilise the transcript from degradation (initiating as soon as the pre-mRNA is being transcribed), capping is at the 5' end and consists of a guanine nucleotide connected to the rest of the mRNA via a 5' to 5' triphosphate bridge and is further methylated to form a 7-methylguanylate cap. This capping prevents degradation by exonucleases, can regulate nuclear export and promote translation.

As soon as the RNA polymerase has reached the end of the gene, the newly made RNA is cleaved by CPSF (Cleavage and poly-adenylation specificity factor [genes]), which then recruits polynucleotide adenylyltransferases that add adenosines to the 3' end of the transcript. The position at which a poly-A tail is added depends on the potential usage of multiple sites containing a polyadenlyation signal (AAUAAA), in this way transcripts can be alternatively poly-adenylated. Poly-adenylation can change the stability (translational ability) and nuclear export to the cytoplasm, these differences can be used to regulate final mRNA levels and cell localisation (Di Giammartino, Nishida, and Manley 2011).

Non-sense mediated decay and microRNAs

In addition to passive mRNA degradation, an active translation-dependant RNA surveillance pathway can be used, called NMD (or Non-sense mediated decay), that targets transcripts with premature stop codons (PSC), derived from either mutations or isoform variants (Kurosaki and Maquat 2016). Discovered in both yeast and humans initially (Chang and Kan 1979; Maquat et al. 1981), the pathway is now known to be conserved to the last eukaryotic common ancestor (Causier et al. 2017). There are different proteins involved in different species, with the conserved core being UPF1/2/3 (UP-Frameshift) genes, two of which are associated with the exon-exon junction complex in mammals, and the SMG (1,5,6 and 7) genes, which generally act through phosphorylation of UPF1. PSCs are detected if the ribosome gets stalled at a stop codon before the next exon junction complex (EJC) (Lindeboom, Supek, and Lehner 2016), where they then recruit UPF1 to initiate the degradation process. This process is sensitive to the location of the PSC, it must exists at least 50

nucleotides upstream of the last EJC to activate this decay pathway (Thermann et al. 1998).

MicroRNAs are short (~22 nucleotides), non-coding RNAs that can regulate gene expression by complementary base pairing to target mRNAs. They are transcribed by RNA polymerase II, can be spliced, then folded into a hairpin loop. This is then bound by various proteins, specifically on the strand that contains the microRNA, and the proteins DROSHA/DGCRB trim around 11nt from the microRNA 3'/5' tail region, leaving a two nucleotide 3' overhang. The microRNA can then be transported out of the nucleus, where it is further processed by the protein DICER then cuts off the hairpin loop, to produce a microRNA duplex, which finally is bound by RISC (RNA-induced silencing complex). This complex contains some Argonaute protein members, that finally have a single stranded RNA template that can be recruited to specific targets, labelling them for degradation. The sequence similarity may not be 100% complementary in sequence to its target, and microRNAs are known to undergo various post-transcriptional modifications (such as RNA editing).
MicroRNAs also have additional roles in the cell and even outside of it. In the nucleus, they can act as enhancer regulators known as 'RNA activation' (RNAa), and are linked to changes in histone marks, related with gene activity (Xiao et al. 2017). They can also be transported out of the cell in exosomes, where they can act as signalling molecules (Squadrito et al. 2014).

Protein translation

Once the mRNA has been transported to the cytoplasm, a ribosome can then attach to the translation start site at the trinucleotide ATG (producing the first amino acid-methionine). Depending on a protein's function and/or required post-translation processing or folding, the mRNA can be recruited to specific areas of the cell, such as the endoplasmic reticulum (ER), which has ribosomes bound to the membrane. The ER is a special environment where the amino acids can be folded correctly to produce a viable protein.
Protein levels can be measured in various ways, including mass-spectrometry, which involves the degradation of proteins (with trypsin) to be able to measure the weight of

constituent peptides to quantify the relative abundances of specific proteins. Protein translation can be estimated using polysome profiling, which separates cell lysate into the mRNAs attached to single or multiple ribosomes, with RNA-Seq of each fraction (Piccirillo et al. 2014).

3. Molecular change in the embryo

The following section deals with the molecular changes of the embryo. Starting in order of the envisioned molecular dogma hierarchy (Introduction section 2; p.13) and the expected organisation within the embryo from the chromatin to protein synthesis, I review what is known about these processes in the developing embryo with emphasis to connections to alternative splicing.

Epigenetic changes in the embryo

One of the most striking molecular changes in the embryo is the rewiring in epigenome, where specific marks are considered completely reset during this time in order for the embryo to be able to differentiate into all the cell types of the adult organism. This is most clear in the total de-methylation of the genome, first in the paternal genome at fertilisation and again at the blastocyst stage (in mouse), and the re-methylation toward differentiation into somatic cells and gametogenesis (Figure I.3.a). De-methylation is linked to open chromatin and transcriptional activation.

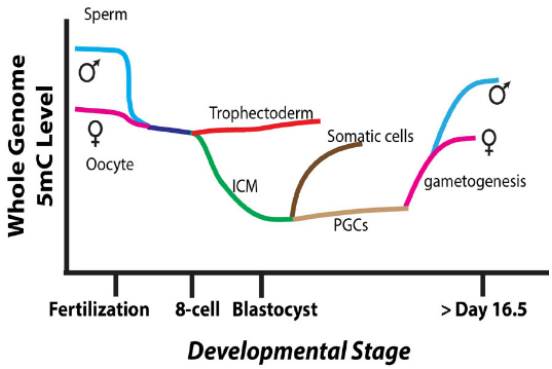


Figure I.3.a. General trend in genome 5'-methylcytosine levels during mouse early development (O'Neill 2015). Showing de-methylation in sperm and the inner cell mass (ICM) and re-methylation in the somatic cells and during gametogenesis. Based on data from (L. Wang et al. 2014)

In the sperm, the genome is especially condensed due to the replacement of histones with protamines, which allow DNA to be packed more densely. This process allows sperm to be more stream-lined and protects the DNA from damage (W S Ward, Coffey, and Ward 1991). It also renders the cell transcriptionally silent, although in humans (and some other mammals), 2 to 15 % of the haploid genome is occupied by histones instead of protamines (Gatewood et al. 1987; W. Steven Ward 2009), with some evidence that specific genes are left transcriptionally active (Hammoud et al.

2009). At fertilisation, the protamines must be removed in order for the two genomes to fuse.

This change from closed to open chromatin can also be visualised by studying topologically associated domains (TADs), which are self-interacting regions of the genome that are revealed by Hi-C (Dixon et al. 2012), which cross links DNA, before digestion and religation of DNA strands within cross-linked positions, to reveal long distance interactions. This has been well defined in mouse, and show that sperm have highly defined chromatin domains with extra-long-range interactions due to its protamine organisation (Ke et al. 2017) and mature oocytes (MII) have a lower degree of compartmentalisation (Flyamer et al. 2017). At the zygote stage, there are very weak boundaries, but these are quickly strengthened by the genome activation stage (Figure I.3.b, Late 2-Cell in mouse). Cohesin, which is critical for formation of domains around CTCF sites in the early embryo (Gassler et al. 2017), but also RNA pol II sites and active transcription has been suggested to potentially establish insulation between TADs (Hug and Vaquerizas 2018). Given RNApolII is central to genome activation, experiments were conducted to inhibit transcription in Drosophila and mice embryos. This showed that TADs are not dependant on transcription but it is important for retention of boundaries (Hug et al. 2017; Ke et al. 2017). Further, DNA replication has been linked to the establishment of the boundaries in 2-Cell mouse embryos, and this has been attributed to the accessibility of the chromatin during this transition, allowing higher-order chromatin organisation (Ke et al. 2017). How these high order chromatin changes directly influence the development and transcription programming of the early embryo is still under intensive study (Hug and Vaquerizas 2018).

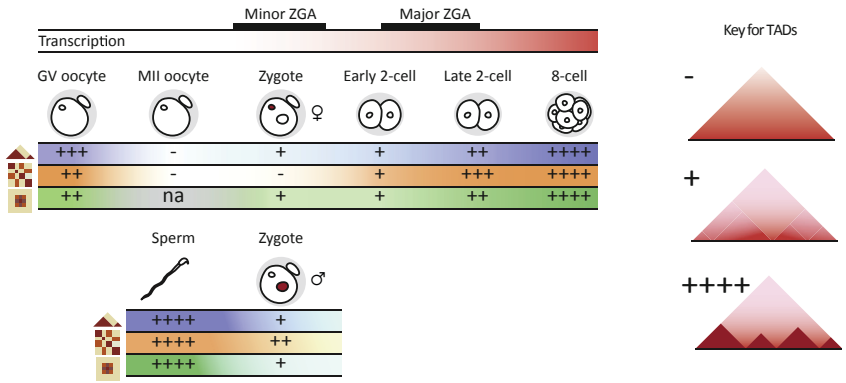


Figure I.3.b. Chromatin structure during mouse early development. Levels of topological associated domains (TADs; blue), compartments (orange) and focal loops (green), in Mouse Germinal vesicle oocytes, Metaphase 2 oocytes, pronuclear genomes (male/female), sperm and some initial cleavage stages before and after genome activation. Key to show differences in TAD intensity, from negative (-) to string (++++). Modified from (Hug and Vaquerizas 2018).

With the development of high-throughput techniques, we also have detailed datasets on histone modifications changing during oocyte maturation and early development. Lysine acetylation has dynamic fluctuations in intensity from germinal vesicle to MII oocytes, in mouse (H3K9/14 and H4K5/8/12/16), cow (H4K5/8/12/16), pig (H3K9/14 and H4K5/8/12/16) and sheep (H3K9 and H4K5/12) (Gu, Wang, and Sun 2010), with many becoming de-acetylated during meiosis and the acetylated by/at zygotic division and somatic cells, indeed somatic cell nuclear transfer to oocytes, leads to de-acetylation of the donor nucleus and this appears to be regulated by HDAC1 within the oocyte (J. M. Kim et al. 2003).

Further, two papers independently published genome-wide profiles of the activating H3K4me3 marks during pre-implantation development (Dahl et al. 2016; Xiaoyu Liu et al. 2016), which came out at a similar time with seemingly contradictory conclusions, Dahl *et al*. 2016 found that H3K4me3 signal was in broad domains (extended region of continuous signal, >5kbp) in the transcriptionally-silent oocyte (covering 22% of the genome), but at the 2–cell stage was confined mostly to the transcriptional start site (TSS) around 1-2 kilobases in size, in line with the onset of zygotic transcription in mouse. Liu et al. did not focus on the ‘non-canonical’ flat

H3K4me3 domains, but instead used more stringent calling of peaks/domains, and found that domain numbers increased from oocyte to 2-cell, around the TSS, and that broad domains actually increased from the oocyte to the blastocyst. Also, as previously suggested (Benayoun et al. 2014), broad domains led to increased transcriptional activity compared to narrow H3K4me3 domains (Xiaoyu Liu et al. 2016), and this may be due to the capacity of broad domains to form a complex with super-enhancers (F. Cao et al. 2017). Both studies noted that these domains were anti-correlated with DNA methylation, and that these broad domains are removed by KDM5A/B, (Xiaoyu Liu et al. 2016).

(Xiaoyu Liu et al. 2016) could also show that repressive H3K27me3 marks also get re-established on the chromatin, but at a slower rate than H3K4me3. (Dahl et al. 2016) could show that H3K27ac coverage increases from oocyte to 2-cell, and is a mark of active cis-regulatory elements, by investigating the regions with H3K27ac in 2-cell and 8-cell they could show they were enriched for ZGA genes and were enriched for specific transcription factor binding motifs, including RARG, RLR1, NR5A2, cMYC and CRX.

Finally, X-chromosome inactivation is another key epigenetic change that occurs at this time (in females), where the paternal X is inactivated at the 4 to 8-cell stage (mouse), with the X being re-activated at around the blastocyst stage, and further random X-inactivation in the ICM at the late blastocyst stage (Payer 2016).

Zygotic genome activation (ZGA)

Normally called Zygotic genome activation (ZGA) or embryonic genome activation (EGA), this process describes the first time in development when genes are expressed. In mouse this occurs between the zygote to 2-cell stage, and in human/cow at the 4 to 8-cell stage.

The transition from maternal to embryonic/zygotic control has been studied long before the advent of molecular biology. How and when the embryo takes control of its own development has been a core question in developmental biology. Here, I outline some key experiments that helped understand this transition.

In 1936, Ethel Browne Harvey discovered that de-nucleated sea urchin embryos (genus: Arbacia) could continue to divide until the 500 cell stage with similar

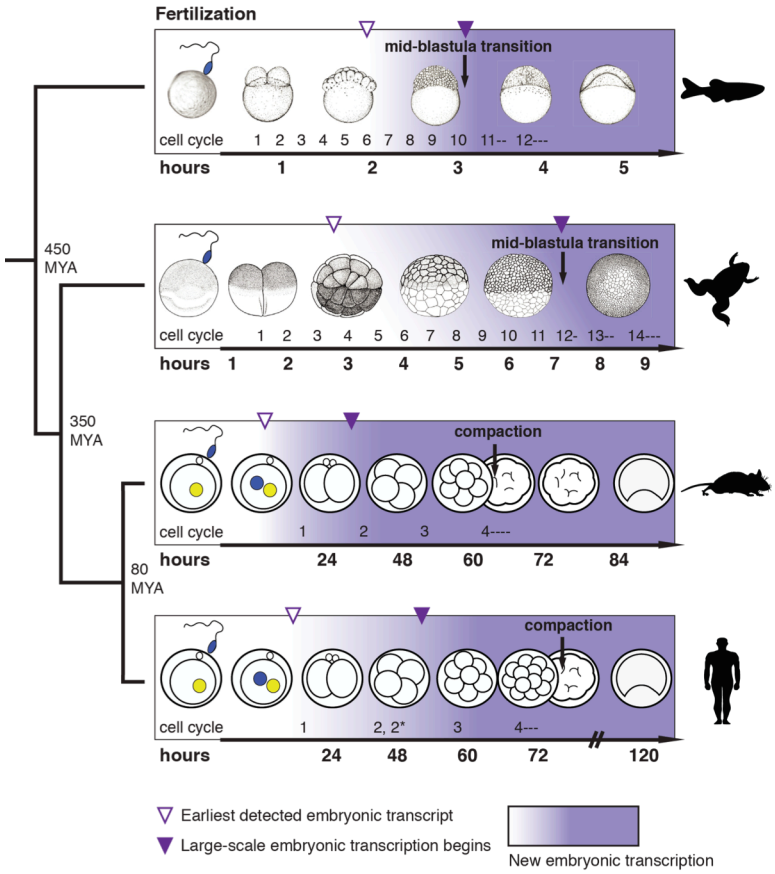
morphological characteristics to nucleated cells (Browne Harvey 1936), suggesting the eggs cytoplasm has the capacity to control its own development and cell cycle during the initial stages using maternal factors.

In the 60s, and with the knowledge of molecular biology, Sir John Bertrand Gurdon then discovered that somatic nuclear transfer to a de-nucleated oocyte could develop into a new frog, suggesting that the oocyte's cytoplasm alone is sufficient to transform a terminally differentiated nucleus into a chromatin reset version (Gurdon 1962).

Early work using inhibitors of RNA synthesis could show exactly when de novo transcription was required in the embryo (Brachet, Denis, and de Vitry 1964). Use of actinomycin D (which binds DNA, preventing transcription) could show that micro-injected xenopus wild-type oocytes would continue to divide until the blastula stage of development, where they would arrest and show sign of chromosomal failures (Brachet, Denis, and de Vitry 1964).

These studies showed that the oocyte's cytoplasm was competent to direct its own development but eventually transcription must be required for further development. How these RNAs are first activated and which mechanisms are present to direct degradation and chromatin remodelling not completely known. Box 1 gives a summary of the ZGA in non-mammalian systems.

Box 1. ZGA in non-mammals (in brief)



Zebrafish has rapid maternally coordinated cell divisions until the 10th cycle, when along with cell cycle lengthening (S and gap phases), the first zygotic transcripts are expressed. Inhibition of transcription at this time does not affect cell cycle lengthening, so the two processes are somewhat independent. Nucleocytoplasmic ratio is known to control the onset of cycle lengthening and not the cell division number, as haploid embryos transition at the 11th cell division and tetraploid embryos at the 8th/9th (Kane and Kimmel 1993).

Box 1. (Cont.)

Micro RNAs are known to be expressed even earlier, at the 6th cell division, including miRNA-430, which are involved in the maternal degradation of its target maternal mRNAs (Giraldez et al. 2006).

Many of the early genes are short, suggesting insufficient time in interphase to complete long transcripts (Heyn et al. 2014).

Xenopus also have fast dividing early embryos, with a minor-wave of ZGA at the 8-Cell stage and major at the 12th division, coordinated with cell cycle lengthening (S-phase) and asynchronous cell divisions. At this stage the DNA damage checkpoints also become activated with the addition of gap phases. miRNA-427 is expressed at the minor-wave and is believed to clear maternal transcripts (Giraldez et al. 2006)(Owens et al. 2016). Finally, some transcription factors (*bix4*, and *sox17α*) and nodal—related ligands are known to be expressed which could prepare the embryo for later gene expression (Skirkanich et al. 2011).

Drosophila early development is controlled by maternal mRNAs/proteins, until cell cycle 14 (2-3 hours after fertilisation), where the protein Smaug (SMG in human; those involved in NMD as well) is translated upon egg activation, binds to specific UTR regions on target maternal mRNAs leading to their targeted removal in association with specific miRNAs (micro RNAs) such as miR-309 (responsible for destabilisation of 410 maternal mRNAs, for example). (Tadros et al. 2007; Green et al. 2003)

Overall, these non-mammal model systems are defined by their external embryo development, their much faster cell cycle after fertilisation, use of targeted maternal degradation and strong control of the nuclear to cytoplasmic ratio to initiate the onset of transcription.

Mammalian transcriptional activation

In mammals genome activation has been studied in a number of species, and suggested to fall into two main phases, minor and major, which describe the two initial "waves" of expression detectable in the embryo. The minor wave was discovered by using reporter gene arrays (Ram and Schultz 1993) and incorporation of BrUPT (5-Bromouridine 5′-triphosphate; a traceable nucleotide) in mouse (Aoki, Worrad, and Schultz 1997). This wave of expression is somewhat insensitive to RNApolII-blocker alpha-amanitin, and is characterised by low levels of expression across the genome, even in intergenic regions (Abe et al. 2015). In mouse this begins at the zygote stage (Aoki, Worrad, and Schultz 1997), and in human at least by the 3-4 cell stage (Taylor et al. 1997).

In contrast, the major wave of genome activation is more specific to exonic regions, involves hundred-thousands genes and occurs at the Zygote-2 cell stage in mouse (Schultz 1993), 4-8 cell stage in human (Braude, Bolton, and Moore 1988) and 8-12 cell stage in cow (Camous et al. 2004; Telford, Watson, and Schultz 1990) (Also see box 1). Alpha-amanitin stalls development at the 2-Cell stage in mouse, suggesting that some genes expressed at this time are required (Goddard and Pratt 1983). The extent and role of this burst of expression is not well understood, indeed, many speculate that the lack of embryonic lethality of many gene knock-down or knock-outs during preimplantation development may mean that many genes are not required (Artus and Cohen-Tannoudji 2008), of 10688 mouse knock-outs in the Jackson database (www.informatics.jax.org; as of 20 February 2019), 753 (7.0 %) are listed as 'embryonic lethal', 'lethal before or at implantation' or 'embryonic growth arrest', suggesting that the maternal contribution of proteins may allow the embryo to function perfectly normally, without the need for gene expression at this stage (Supp. Table E1), for at least some genes. There is also a lack of evolutionary conservation of the genes being expressed at this time (Ko 2016; Heyn et al. 2014), with only 8% of ZGA genes shared between human, cow and mouse, in comparison to 49% of maternally stored genes (Xie et al. 2010). This suggests that either, the 'second wave' could be required for species-specific differences or may not be so critical, showing somewhat promiscuous effects (similar to the first wave).

In mammals, several ZGA genes have been shown to be essential for progression, including Zscan4 (Falco et al. 2007), Zar1 (Wu et al. 2003) and Hira (C. J. Lin et al.

2014). In addition several transcription factors have been linked to this burst of expression, such pioneer factors including DUX4, which is a homeodomain transcription factor, which up-regulates similar genes at ZGA and in facio-scapulo-humeral dystrophy, where DUX4 is also aberrantly expressed (Lemmers, Miller, and van der Maarel 1993). In addition, we have shown that multiple homeobox genes are activated at genome activation, along with their targets (Maeso et al. 2016)(Box 1.4; Annex 1). The root cause of this transcriptional activation in mammals is still under debate, with chromatin remodelling (Jukam, Shariati, and Skotheim 2017) changes responsible in part for opening up the DNA to trans-factors, that together lead to the initial burst.

Maternal mRNA degradation

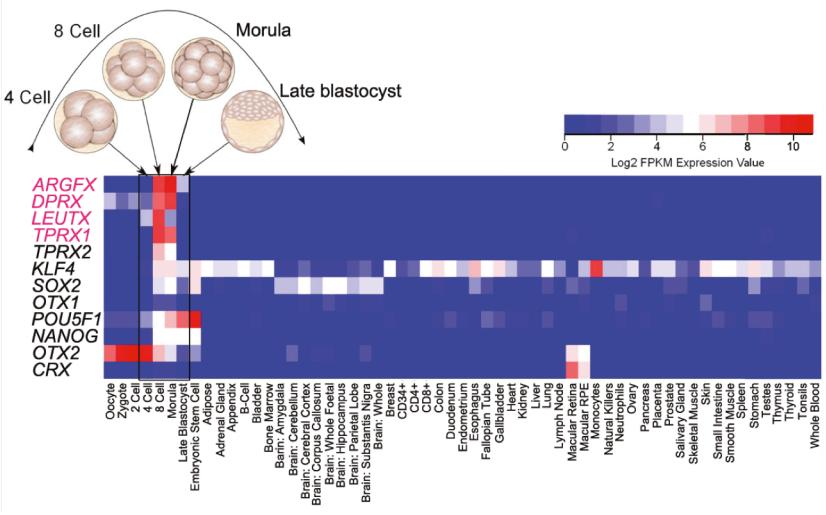
Finally, maternal mRNA clearance is also important during this stage, with much research focusing on miRNAs. In non-mammals, there are many examples such as miR-430 in zebrafish (Giraldez et al. 2006) and miR-309 in drosophila (Tadros et al. 2007), which are actively transcribed at genome activation and lead to decay of specific maternal mRNA targets. In mammals, there are a few studies on miRNAs, but they do not have such clear roles as in many non-mammals. In mouse, oocytes injected with sperm from Drosha KO mice (key to miRNA biogenesis) show reduced development, that could be rescued by injection of miRNAs from wild-type sperm (Yuan et al. 2016). 14 miRNAs were shown to be present in sperm (Yuan et al. 2016) but absent in unfertilised oocytes, one of which miR-34c has been shown to delay DNA synthesis and supress the first cleavage event (W.-M. Liu et al. 2012). At fertilisation, miRNAs (in mouse) are reduced at the two cell stage (genome activation; 60% down-regulated), and a few apparently selectively targeted for degradation (F. Tang et al. 2007). In overall levels, miRNAs have a reduction at genome activation that is recovered to oocyte total levels by the blastocyst (F. Tang et al. 2007). The first embryonic miRNAs to be detected are mir-290 and mir-295 in mouse (Houbaviy, Murray, and Sharp 2003), and along with miR302 in cow, have also been linked to degradation, based on their shared sequence (AAGUGC) with invertebrate miRNAs involved in degradation (Svoboda and Flemr 2010; Hwang et al. 2018). Examination of the core genes involved in miRNA biogenesis revealed that many are down-

regulated at zygote and subsequent stages (in mouse), potentially meaning early embryos cannot generate/process miRNAs efficiently (García-López and del Mazo 2012), and interestingly it has been proposed that some miRNAs are protected in the early embryo by post-transcriptional modifications that allow them to be reactivated and may bypass the need for biogenesis (Yang et al. 2016). Finally, Dicer1 and Drosha knockouts are known to be embryonic lethal, and show that miRNAs may have an important impact during pre-implantation development (Bernstein et al. 2003; Y. Wang et al. 2007).

Box 2. Evolutionary origin and functional divergence of totipotent cell homeobox genes (Full publication: Annex 1).

The function of the numerous genes specifically activated at genome activation is unknown and is characterised by low levels of conservation in the genes becoming expressed. To further understand the role of genes during genome activation, we looked specifically at the role of homeobox genes, which are short transcription factors known to be involved in development and embryonic patterning.

Here, we could show that the PRD class of homeobox genes (ARGFX, DPRX, LEUTX and two TPRX genes in human) arose by tandem duplication from the retinal-expressed CRX gene, followed by functional divergence of the resultant genes, and restricted expression to the genome activation stage of human, mouse and cow, and therefore presumably across eutherian mammals. Below you can see the log2 FPKM (fragments per kilobase per million reads) values for the five PRD genes expressed at 8-morula stage in human only and not in any differentiated tissue (along with various selected lineage markers), whereas CRX is expressed only in the retina.



Box 2 (cont.)

Ectopic expression of the ETCH-box genes in human primary fibroblasts could show that ARGFX, LEUTX and TPRX1 could drive extensive transcriptional remodelling. LEUTX and TPRX1 had a high level of overlap, suggesting some functional redundancy between the two, yet ARGFX showed an antagonistic pattern to the others. This we suggest may explain the rapid turnover of their target genes at this period in development. Downstream targets included histone genes (e.g. HIST1H2BD) and RLB, a transcription factor in the NFκB.

Splicing and the embryo

Little is known about the dynamics, functionality and consequences of splicing in pre-implantation embryos, hence the purpose of this thesis, yet a few have explored some specific events. Further, splicing has been much better characterised in ESCs and in the immature oocyte.

In the oocyte, maternal Srsf3 (a well described RNA-binding protein and pre-mRNA splicing factor) is known to be critical, where conditional deletion of Srsf3 in germinal vesicle oocytes largely fail to breakdown the vesicle and progress through meiosis (Do et al. 2018). This was linked to loss of correct target spliced isoforms and also to de-repression of B2 SINE transposable elements (Do et al. 2018). Additionally YTHDC1, which detects N^6-methyladenosines, a common post-transcriptional modification, is required during both sperm and oocyte development, and could be shown to lead to splicing defects in a mutant, potentially due to its role in recruiting splicing factors (including CPSF6, SRSF3 and SRSF7) to the pre-mRNA (Kasowitz et al. 2018).

During the first wave of genome activation, it was found that mice lack of correct splicing, denoted by high rates of intron retention (Abe et al. 2015) and even

transcription of intergenic regions. We also know that mRNAs do not appear to be stored as pre-mRNAs for later processing (splicing), as Poly-A negative data (those sequencing non-poly-A RNA species) show that the maternal mRNAs are already spliced (Abe et al. 2015).

In cleavage stage embryos, very little is known. The majority of publications with RNA-Seq at these times (i.e. the raw data of this thesis) do not address genome-wide splicing variability during early development in human and mouse. Some display specific examples, e.g. Foxp1 (Yan et al. 2013). However, (Chen et al. 2017) did explore isoform change, noting a lack of conservation between human and mouse. They also found that many events that were exon-skipping in human were listed as constitutive-exons in mouse. Overall this study did not explore in any more detail how splicing affects the functioning of the embryo. In addition, from the MGI (Mouse genome informatics, Jackson lab), 8 RNA-binding proteins (RBPs) are listed as embryonic lethal before the implantation stage, Sf3b1, Sf3a1, Srsf3, Snrnp200, Hnrnpl, Prpf3/19, Rnpc3, Qk, Tra2b and Rbm19 (Supp. Table E1). For example, Srsf3 mutants die by the morula stage (Jumaa, Wei, and Nielsen 1999). This could suggest that splicing is important during the pre-implantation stage, yet RBPs can have disparate roles in the cell, form RNA editing to RNA localisation, so we cannot distinguish their alternative splicing roles solely from these data.

At the blastocyst stage, (W. Huang and Khatib 2010) used RNA-Seq from normal and degenerated blastocysts to show that many exons were alternatively spliced (with minimum of 2 reads per expected inclusion/exclusion junctions), 4,426 (exon skipping, alternative 5' splice site and alternative 3' splice site) events found in 2,032 genes. Yet the authors only found two significant splicing events different between normal and degenerate (with multiple testing correction and FDR of 1%), with only one listed (MYL6). The low number of differences were attributed to a lack of exon-exon junction reads in their dataset, highlighting the difficulty measuring splicing in these small cells.

In the post-implantation embryo (Revil et al. 2010) conducted a microarray study of splicing during mouse embryonic stages day 8.5, 9.5 and 11.5, including placental samples, and could show that 9% of their splicing probes were alternative between the

embryo and placenta and 2% between embryonic stages, many in genes relating to development, cell adhesion and signal transduction. This suggested that splicing may be functionally relevant at this stage in development and they could specifically link the expression of a splicing factor Fox2 to particular target exons that changed in inclusion. (X. Lu et al. 2018) using RNA-Seq data could further show that alternative splicing and promoter usage is abundant in E7.5 mouse embryos, and are enriched in developmental (including Wnt Signalling), adhesion, chromatin splicing, cell-cycle, DNA damage related GO terms. They could even show that there are splicing differences between the three germ layers: mesoderm, epiblast and endoderm.

Box 3. Specific examples of splicing in early development

The role of a few specific gene isoforms has been documented, many of which provide a reference to our RNA-Seq analysis. These include: cow: §OPN/§STAT5 (Laporta, Driver, and Khatib 2011), mouse: Bcl (Perumalsamy et al. 2010) and human: HPRT (Daniels et al. 1998) examples. Here, I detail a few more interesting examples.

Xbp is a protein that responds to ER stress, and mRNA is unusually spliced by a cytoplasmic endoribonuclease (Inositol requiring enzyme-1 [IRE1]) removing a 26-nt intron. Removal of the intron creates a frame-shift, the resultant protein of which can enter the nucleus acting as a transcription factor, up-regulating unfolded protein response genes. This reaction is known in cow (Sung et al. 2006) and mouse (J. Y. Zhang et al. 2012)**,** where the long isoform is found from the oocyte to blastocyst, yet the small isoform is not found in the one-cell stage in mouse. In zebrafish, the short isoform is present first then disappears by 9 hours post-fertilisation (J. Li et al. 2015). It is also important for xenopus pancreas development (Y. Li et al. 2016).

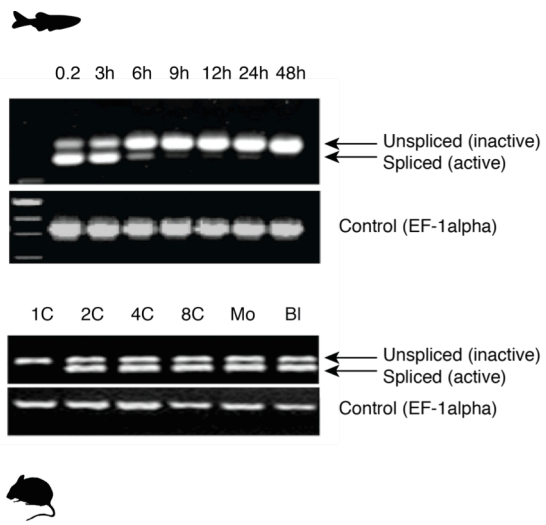


Figure. B3. RT-PCR of Xbp long (un-spliced) and short (spliced) isoforms in zebrafish (hours post fertilisation) and mouse. Adapted from (J. Y. Zhang et al. 2012) (J. Li et al. 2015).

Box 3. Cont. Specific examples of splicing in early development

HLA-G isoforms have been shown to be alternatively included during human preimplantation development (Yao, Barlow, and Sargent 2005). HLA-G isoforms 1-4 are membrane bound, and 5-6 are soluble in the cytoplasm and the protein is linked to ability of the embryo to implant successfully, as HLA-G is required to prevent the maternal immune system destroying the embryo. Most isoforms increase in abundance from 2-Cell to blastocyst, but HLA-G5/G6 are not detectable until the morula/blastocyst stage, where only 20%-33% of embryo had detection (Yao, Barlow, and Sargent 2005).

Finally, the role of splicing in stem cells is well known, in that a program of 181 human and 103 mouse ESC-specific exons were found (Han et al. 2013) that are negatively regulated by MBNL1/2. Of these events, a mutually exclusive event in FOXP1 stood out, as the exon was in the DNA-binding domain of this transcription factor (Gabut et al. 2011). The two isoforms were found to have different binding specificities to their target genes, with the ESC-specific isoform affecting the transcription of early developmental and pluripotency genes and the differentiated isoform driving the expression of gene involved in differentiation (Figure I.3.d).

Studies reprogramming mouse embryonic fibroblasts (MEFs) into stem cells, found specific RNA-binding proteins were up-regulated, such as Esrp1 and Mbnl1/2 and Rbm47 (Cieply et al. 2016). It was also shown that ectopic expression of Esrp1 could modulate the splicing of Grhl1 (an epithelial transcription factor). Further, (Kanitz et al. 2019) could show that the dynamics of Esrp1 and 2 have a conserved role in reprogramming, changing the isoform landscape of many genes involved in the mesenchymal to epithelial transition. Therefore we would assume these same dynamics are involved in the early embryo.

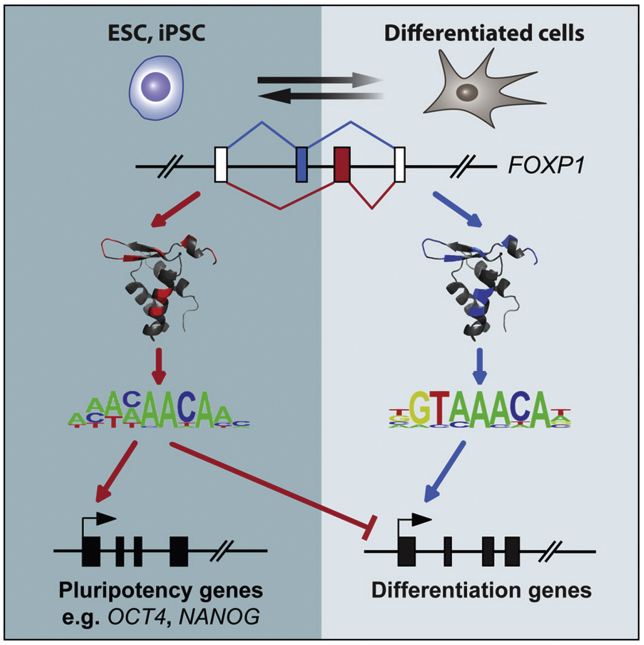


Figure I.3.d. FOXP1 splicing in ESCs (Embryonic stem cells)/iPSC (induced pluripotent stem cells) and differentiated tissues. Differential splicing of Foxp1, results in two different isoforms, that change the binding specificity of the transcription factor, driving pluripotency-related of differentiation-related genes. Figure from (Gabut et al. 2011).

Box 4. Splicing Zebrafish/Drosophila embryos

In zebrafish, Aanes *et al.* (2013) could show that 2,160 primary transcripts change in splice isoform composition, between pre- and post- zygotic genome activation. Most showing subtle changes, though a few had a switch like change. They also discovered many 'novel' isoforms due to exon-skipping at genome activation, that had a shorter coding sequence (i.e. they are truncated by premature stop codons introduced). They specifically mention *foxh1*, *hdac5*, *dnmt1*, *dnmt3*, *gsk3a and f11r, but in most cases only speculate as to their function. Dnmt1 was truncated due to the skipping of 19 exons pre activation in zebrafish, removing most its functional domains, which corresponds to its inactivation in xenopus before the mid-blastula transition. At later stages the genes has a full coding sequence and is functional. Finally,* Aanes *et al.* (2013) *found that exon skipping events were enriched in H3K36me3 and H3K4me3 marks, compared to a random selection of exons, whereas H3K27me3 marks were not enriched. This enrichment was found specifically in the exon and upstream intron, and the author's suggest that these may recruit specific splicing factors.*

In Drosophila, Guilgur *et al.* (2014) could show the gene fandango (Xab2 in mouse), was essential for splicing of early zygotic but not maternal pre-mRNAs. Fandango mutants had blastoderm defects, which could be rescued by ectopic maternal expression of early zygote pre-mRNA. In addition, small early zygotic transcripts with multiple introns were often poorly spliced. Guilgur *et al.* (2014) surmised that, due to the fast cell divisions in drosophila embryos, only short genes could be spliced in time the time between each cycle, as mitosis inhibits both expression and splicing (Guilgur *et al.* 2014).

In summary, apart from confirming the splicing machinery must exist at least by genome activation stage and that splicing factors are crucial for correct development, we know very little about splicing dynamics and regulation between the oocyte to blastocyst.

Despite many high-profile publications of early-embryo transcriptomes, none have systematically explored the global splicing dynamics. Several papers highlighted the enrichment of RNA processing genes changing in expression between stages (Yan et al. 2013; Graf et al. 2014), and a few gave specific examples to show splicing occurred (Yan et al. 2013). This apparent lack of interest could derive from the difficult to detect splicing accurately given the low levels of RNA and coverage (See Discussion) or that this layer of regulation was overlooked (most focusing of genome activation and/or pluripotency).

Polyadenylation in the embryo

Many studies have investigated polyadenylation tail length as a mechanism for transcriptional control in the early embryo. Most studies thus far have been in *Xenopus* (Radford, Meijer, and de Moor 2008), *C.elegans* (Boag et al. 2018) and *Drosophila* (J. Cui et al. 2013; Lim et al. 2016). They have shown that mRNAs in the oocyte can be stored in a protein complex or have short poly(A)-tails (~20-40 nucleotides) that prevent translation, until they are activated by extending the poly(A) tail (to 80-250 nucleotides), at later time points (Radford, Meijer, and de Moor 2008), in the cytoplasm. Long poly(A)-tails are thought to affect translation by attracting poly(A) binding proteins that recruit initiation factors, forming a closed-loop complex that stabilises translation. Cytoplasmic polyadenylation in the early embryo of *Xenopus* has been linked to characteristics in the mRNA sequence, such as cytosine-rich and uracil-rich cytoplasmic polyadenylation elements, that can be bound by CPEB, PCBP2 and ElavA (Radford, Meijer, and de Moor 2008). The best-known element is CPE (cytoplasmic polyadenylation element), roughly equivalent to UUUUUAU/UUUUUAAU, was found in *Xenopus* mRNA that has a strong change in polyadenylation. Cytoplasmic polyadenylation element binding protein (CPEB) is required for the recognition of these sequences (Stebbins-Boaz, Hake, and Richter 1996; Fox et al. 1992).

Interestingly, poly(A)-tail length was shown to strongly correlate with translational efficiency (as determined from ribosome footprint profiling) before genome activation but has no correlation by gastrulation stage, in zebrafish and *Xenopus,* nor in most later stages of development (Jalkanen, Coleman, and Wilusz 2014). However recent

research has highlighted that the length is still important in adult tissues (Lima et al. 2017).

In mammals, not much in known. Ji et al. (2009), could show that 3' UTRs generally increased, due to usage of a more distal site and Vassalli et al. (1989) could show that translation was in part regulated by polyadenylation during oogenesis.

Molecular regulation of lineage differentiation.

The first sets of cell lineages to arise are the trophectoderm and the inner cell mass (ICM), and in mouse, this simple distinction is brought about through polarisation of cells that fall on the outside or inside of the developing ball of cells, where naturally as the cells divide, some cells will be located in the middle and others on the periphery. This positioning effect results in differences in Hippo pathway signalling, where the polarised outer membrane forms an apical domain containing proteins aPKC (atypical protein kinase C) and polarity protein Par3, whereas the inner cells do not form this domain (Rossant and Tam 2009). The polarised trophectoderm leads to Amot localisation to the apical domain, where it is bound to actin making it inactive. In contrast, in the inner cells, Amot is found throughout the membrane near adheren junctions where it is active and can be phosphorylated by Lats, turning on hippo signalling. Active hippo signalling, means Yap is phosphorylated and sent for degradation, whereas in inactive Hippo signalling, Yap is not phosphorylated can enter the nucleus and bind to Tead4 which is the transcription factor responsible for turning on trophectoderm genes such as Cdx2, Id2, Eomes and Gata3. The inner cells express pluripotent-related transcription factors (Sox2, Oct3/4 and Nanog). An overview is shown in Figure I.3.e. In other mammals, the exact role and timing of the hippo pathway during lineage differentiation is different, though it is assumed to have a conserved role in other species, for instance, YAP is found in the nucleus of the inner cell mass in human and involved in hippo signalling and pluripotency (Qin et al. 2016).

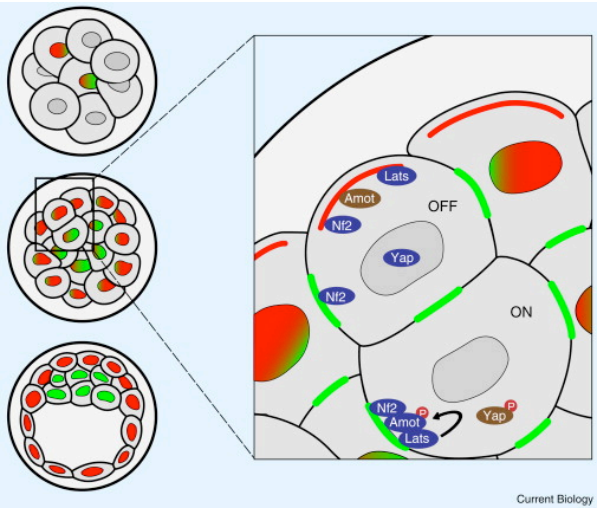


Figure I.3.e. **Hippo pathway proteins during mouse 8-cell to blastocyst.** Inner and outer cells have differences in the apical membrane polarization (red=polarized, green=not), which leads to hippo pathway activation/deactivation (Manzanares and Rodriguez 2013).

Protein levels in the embryo

Despite intense study of pre-translational activity in the cells, mainly due to the technological/cost reasons, protein level information is often missing. Given that transcription output does not correspond very well with protein levels (Vogel and Marcotte 2012), this is a serious caveat to the study of embryos. However, a few studies have been performed.

The earliest comprehensive study (to my knowledge) measured mouse protein synthesis rates (Latham et al. 1991). They showed that 60% and 85% of analysed proteins, at the one-cell and 2-cell stages respectively, had a >2-fold difference, but from the 2-cell stage onwards, there was little change in protein synthesis rates. Specifically, during the 2-cell stage a third of analysed proteins increased by 5 fold, another third decreased 7-fold and 10% had a variable change, in terms of synthesis rates. They suggested the transcription of new mRNAs is directly linked to this increase in protein synthesis (Latham et al. 1991).

(Wilmut et al. 2010) used semi-quantitative mass-spectrometry in mouse oocytes and zygotes and could show that zygotes have less than half the number of detected proteins, presumably due to maternal protein degradation. However, of the proteins

that were enriched in the zygote compared to the oocyte, protein ubiquitination proteins were found, suggesting that potentially newly synthesized proteins of the zygote help with its maternal clearance. Sadly, the authors did not look at post-zygotic genome activation stages. (Deutsch et al. 2014) similarly looked at pre-ZGA stages only (oocyte and 4-cell stage; ZGA in cow is at 8-16 cell stage), and concluded that only 87 proteins significantly changed in abundance.

It has also been shown that in mouse, translation is not active at the 1-cell stage, but becomes activated around the same time as genome activation, at the 2-cell stage (Schultz 2002).

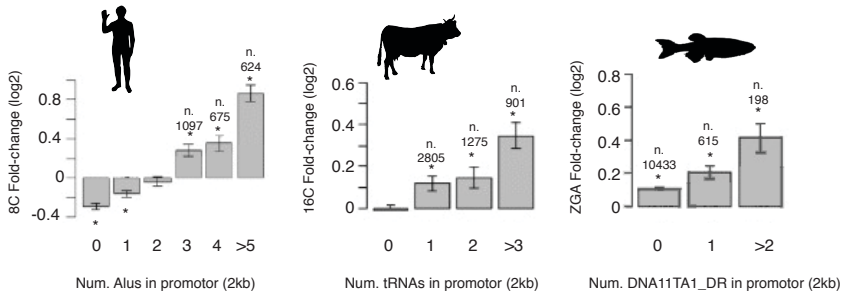
Finally, (Gao et al. 2017) used quantitative mass spectrometry to elucidate protein abundances from zygote to blastocyst, using pools of 8000 mouse embryos per stage. They found that most proteins from the zygote to morula stage have no real abundance changes, contrary to what is found at the expression level, where expression changes dramatically at ZGA (2-cell). However, in their RNA-Seq analysis they erroneously detect activation of the genome at the 8-cell stage (Gao et al 2017: Figure 6B), which brings into doubt some of the conclusions that come from this paper. Searching for public western blot figures (to explore if the protein data are correct) of the 3767 proteins they detect, I found only 3 comparable proteins. mTor, and Pou5f1 (OCT4 in human), show the expected protein levels (Pou5f1 beginning at the 8-cell stage; mTor with consistent levels), whereas Eomes should have high protein levels at least by the morula stage (Ralston and Rossant 2008), yet is present at the blastocyst stage mostly in Gao et al. 2017. The discrepancy between mRNA (starting at the 2-cell stage) and protein (blastocyst-in this paper), could be due to a biological delay in protein translation, or due to technical issues. Also, they only detect a small fraction of the total number of proteins (3767/~20,000; ~18.5%), so we cannot fully compare known protein levels to ensure this data is accurate.

Transposable elements

It has long been held that, given global demethylation in mammalian embryos, transposons would be more transcriptionally active, given that methylation can repress them (Goll and Bestor 2005). Therefore, it may be this point in development that novel TE insertions are likely to be possible, helping them proliferate (Walter et al. 2016). Though some research has contested this belief, and show that there may be

more complex silencing mechanisms available to the cell, given that Tet3-mediated methylation does not significantly contribute to TE-activation (A. Inoue, Matoba, and Zhang 2012; Walter et al. 2016). It is also know, that despite "global demethylation", some TEs, exons, promoters or even splice junctions can remain methylated (Smith et al. 2014; Smith et al. 2012).

Whatever the cause, specific families of transposons do become active during early mammalian development, such as LINE-1 (A. Inoue, Matoba, and Zhang 2012; Marchetto et al. 2013; Ge 2017) , ERVL (A. Inoue, Matoba, and Zhang 2012) elements and SINEs, which have correlation with gene expression when located in



their promoters (Figure I.3.f).

Figure I.3.f. Barplots showing the number of SINE elements in the promoter and average log2 fold change at genome activation stage. A. Human Alu propensity in promoter correlates with increase in gene expression at genome activation. B. Bovine SINE tRNA family have similar density to gene expression change association, C. Zebrafish DNA11TA1 family transposons significantly change in gene expression. *Stars (*)* indicate significant difference from zero based on a *t*-test. Adapted from (Ge 2017).

Telomere lengthening

Telomeres, the protective caps at the ends of chromosomes, need to be lengthened during pre-implantation development (Kalmbach et al. 2014). However, telomerase (the conventional somatic telomere lengthening protein) is undetectable at early mouse stages, and potentially a separate group of recombination-related proteins are believed to add to telomere length at this time (L. Liu et al. 2007). At the blastocyst stage, telomeres are seen to increase in length, and the activity of the conventional telomerase is detected (Ozturk, Sozen, and Demir 2014). Parthenogenetically activated oocytes can also lengthen their telomeres, suggesting the ability to add telomeres is maternally inherited (L. Liu et al. 2007). The dynamics of telomere length are shown in Figure I.3.g.

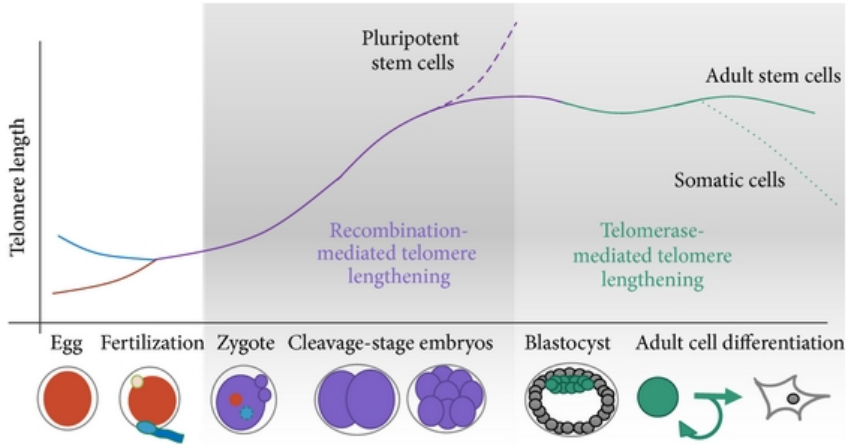


Figure I.3.g. Dynamics of telomere length in mouse early development. Showing rapid increase in telomere length after fertilisation, from a low telomere length in the oocyte. At the blastocyst stage and in later adult stages the conventional telomerase protein takes control of telomere length. Figure from (Kalmbach et al. 2014).

4. DNA damage and repair

So far, I have introduced the core molecular components at play within the embryo that have been studied at great depth, yet one component that has not been well studied is DNA damage.

DNA damage can be defined as: "a deleterious change in the chemical structure of DNA"; **DNA repair** is the ability to reverse some of these anomalies and **DNA damage tolerance** is the ability to overcome damage (e.g. through by-passing replication fork-stalled DNA lesions), although this can lead to mutations. **Mutations** (changes from normal base nucleotides to another) are considered slightly differently to DNA damage, as they can become unrecognisable to repair mechanisms if both strands of the DNA have changed (i.e. with correct pairing A-T, G-C).

DNA damage is also linked to other important downstream effectors. For example, the cell cycle, which contains DNA damage checkpoints at specific stages, which if activated, delay progression until the issue is resolved (Murray and Carr 2018) and chromatin remodelling, whereby chromatin regulators are known to be directly involved with opening up damaged DNA (Price and D'Andrea 2013).

Typical human cells are exposed to many deleterious changes each day, with an estimated 10,000 oxidative lesions (Ames, Shigenaga, and Hagen 1993), 55,000 single-strand breaks (Tubbs and Nussenzweig 2017) and 10-50 double-stranded DNA breaks (Haber 1999; Vilenchik and Knudson 2003) that occur per cell per day. Similarly, embryos are exposed to DNA damage, during oocyte storage (lasting >30years) and preimplantation development, which must be detected and repaired, to prevent the proliferating of mutated cells. However, the extent and activity of the DNA damage pathway is not well understood, with most research being conducted in non-mammal systems, showing that DNA repair may be turned off for some of the preliminary stages. This section introduces DNA damage sources, repair mechanisms and what is currently known during early development.

Sources of DNA damage

DNA damage sources can be split into three main categories, 1: Endogenous, those that arise from the normal homeostasis of the cells (e.g. oxidative/replicative stress); 2: Exogenous, those that arise from external natural forces, that are not derived from the cells themselves (e.g. UV radiation) and 3: Human, those that arise from exposure to synthetic chemicals or products derived from human activity (e.g. Cancer drugs-Etoposide, aflatoxins). Some of these sources can be considered in multiple categories, such as UV, which can be both natural or a product of human activity (i.e. sunbeds). Here I briefly outline the main sources.

Endogenous sources are all metabolism derived. For instance, energy metabolism from normal or pathogenic mitochondria, leak electrons, which results in reactive oxygen species (ROS) and reactive nitrogen species (RNS) free in the cell (Muller 2000). A build up of ROS (e.g. H2O2, O2-) and/or decrease in antioxidants within the cell can lead to oxidative stress. ROS acts by oxidising nucleotide (usually guanine) bases (Dizdaroglu 1992), at a rate of about 2400 lesions per cell per day (Swenberg et al. 2011). ROS can also oxidise lipids, proteins or RNAs and is associated with many pathologies, such as male infertility and cancer (Swenberg et al. 2011).

Deamination forms another internal source of damage, believed to be spontaneous, although dependant on temperature and pH (Lepper et al. 2018; Lindahl and Nyberg 1974), where the loss of the amine group (NH2), converts Cytosine to Uracil, Adenine to Hypoxanthine, Guanine to Xanthine and 5-Methylcytosine to Thymine. These spontaneous changes need to be reversed to the original nucleotides, else can lead to replication fork stalling. However, deamination to thymine is not recognised by the cell as erroneous, forming a mutation (Duncan and Miller 1980).

Further, DNA can be subjected to naturally occurring depurinations and depyrimidinations, which again can increase upon external stimulus (Lindahl and Nyberg 1974).

Finally, replication itself can be the source of error, with the wrong nucleotide/s being incorporated at a rate of roughly 1 per 10^4 to 10^5 of correct bases for most polymerases (McCulloch and Kunkel 2008).

Exogenous sources comprise ultra-violet (UV) light, ionizing radiation (IR) and x-rays. UV damages DNA by directly absorbing a UVB photon, instead of releasing it as heat, which in turn leads to thymine (or sometimes cytosine) bases that become covalently bound in a pyrimidine dimer (Crespo-Hernández, De La Harpe, and Kohler 2008). These dimers then inhibit DNA polymerases and arrest replication and so must be removed. Ionising radiation can come from cosmic radiation to which we are all exposed, or regional sources, such as radon exposure in granite rich regions, to medical radiation used for cancer treatment. Radiation can cause radiolysis (the dissociation of molecules), which can break water into ROS, or mitochondrial ROS production by up-regulation of the electron transport chain (Yamamori et al. 2012).

Human sources can be split into those relating to i. human behaviours/activities, such as sun beds, smoking and burnt food ingestion, ii. Natural toxins (e.g. Aflatoxin) or iii. Intercalators and Cancer drugs (Natural or synthetic; e.g. Camptocthecin). Smoking (type i), for example, introduces a wide variety of chemicals into the lungs, some of which are known to lead to DNA damage, the most well documented chemical is Benzo[α]pyrene (BaP) (also found in burnt foods), that can be broken down into benzo[α]pyrene diol epoxide, that reacts with guanine nucleotides in the DNA, forming an adduct. This can prevent transcription, cause mutations or disrupt the cell cycle (W. Li et al. 2017; Boysen et al. 2009). Cancer drugs (type iii), are a broad category of chemicals including alkylating agents (Kondo et al. 2010), antimetabolites (Parker 2009), topoisomerase inhibitors (Nitiss 1994) and others. They generally work by creating DNA damage and disrupting DNA replication/cell cycle, which normal cells should overcome, but weakened, fast-dividing cancer cells cannot. In addition, cancer cells may have lost some DNA repair genes, which makes them unable to fix lesions, leading to cell cycle arrest. This ideally leads to apoptosis in cancer cells and little effect on normal cells, at a dose that “kills the cancer before it kills the patient” (Corrie 2008). Many of the drugs are known intercalators of DNA, in that they fit within the DNA molecule, change its structure and ultimately inhibit DNA replication (Richards and Rodger 2007).

Repair mechanisms

Given DNA damage in the cell (Figure I.4.i), the next step is to detect the location of the problem, then recruit specific factors to repair the issue. The main pathways include: direct-repair (multiple mechanisms/pathways), mismatch repair (MMR), base excision repair (BER), nucleotide excision repair (NER), homologous recombination (HR) and non-homologous end joining (NHEJ) (see Figure I.4.i). For a more thorough review of the respective pathways, please refer to (Ciccia and Elledge 2010).

Direct repair involves the chemical reversal of DNA damage in a base or bases, normally involving one gene, and without the need of a template to recover the original state. A classic case is demethylation of 6-O-methylguanosine to the correct guanine base, by the protein MGMT (Graves, Li, and Swann 1989). Another case is the Alkb genes in mouse, which have been shown to remove alkylation damage from 1-methyl adenosines/guanines and 3-methyl cytosines/thymines (Ringvoll et al. 2006).

Mismatch repair is generally used to repair errors introduced during DNA replication and recombination, where insertions, deletions and base substitution can occur; this is considered a “proofreading” step (Schaaper 1993). The cell is able to determine which strand is original, and which is the daughter. If a mismatched Watson-Crick base (e.g. G/T) is found, the DNA is removed (up to 100bp) by an endonuclease, before re-synthesis of the strand to recover the correct base pairing (Bebenek, Pedersen, and Kunkel 2011). MLH1 is the most well known eukaryotic protein involved in this pathway, and is frequently mutated in hereditary non-polyposis colon cancer (Froggatt et al. 1996). Mlh1 null-mutant mice show meiotic arrest and are infertile with reduced levels of chiasmata (Edelmann et al. 1996).

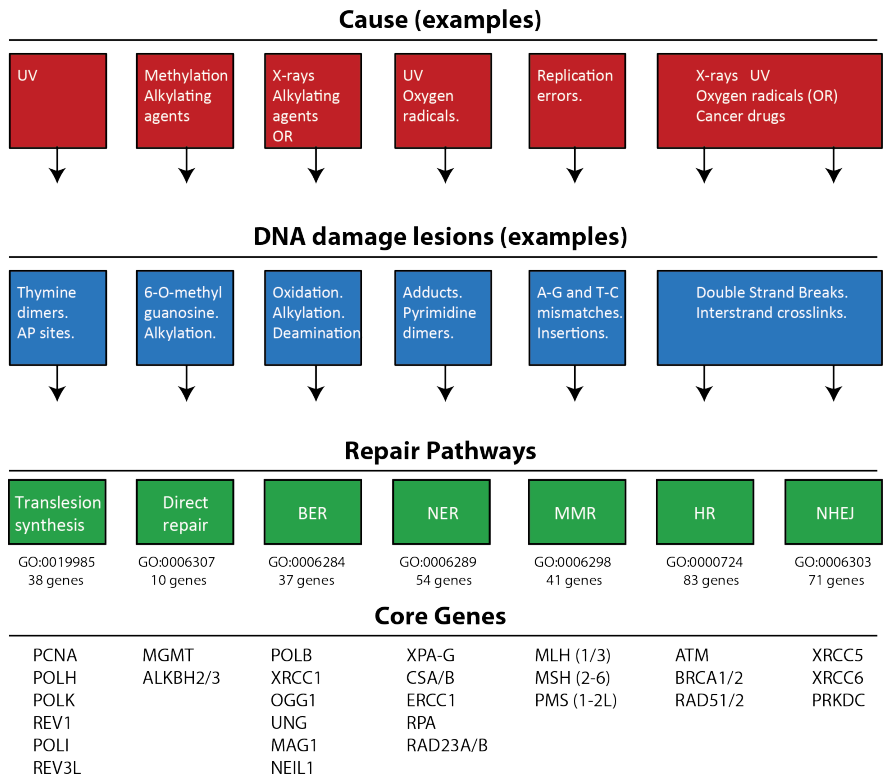


Figure I.4.i DNA repair from cause to repair. For several DNA damage causes, there are specific lesions that can occur. Each type of damage is dealt with by a separate pathway/gene either: BER/NER: Base/Nucleotide excision repair, MMR: Mismatch repair, HR: Homologous recombination and NJEH: Non-homologous end joining. Names of the main genes known to be involved in each pathway are listed alongside their GO term names and total number of genes. Translesion synthesis (not discussed in the text) is a DNA damage tolerance pathway that allows the DNA replication machinery to bypass specific lesions. This only represents a fraction of all known pathways dedicated to DNA repair in the cell, but represent the most studied examples.

Base excision repair is a pathway that can remove non-helix distorting, single stranded lesions (from a single to 10 nucleotides) in the DNA, in order to prevent mis-pairing of nucleotides (causing mutations) or DNA breaks. It can occur throughout the cell cycle, and repairs oxidised, alkylated and deaminated bases, by first recognition by glycosylases to specific lesions, which remove the damage base/s forming an apurinic/apyrimidinic (AP) site. These sites are then cut by an AP endonuclease before being replaced with newly synthesized nucleotides. In addition AP-sites can be formed due to ionizing radiation, where hydroxyl radicals can break the glycosidic bonds of the DNA.

The nucleotide excision repair pathway (NER), is somewhat similar to BER, but can also remove bulky, helix distorting, single stranded lesions, such as thymine dimers caused by UV light. There are two variants of the pathway, global genomic and transcription coupled, which differ in how the lesion is detected. The first can detect these distorting lesions in the DNA, which again are constantly scanned during the cell cycle, in transcribed and non-transcribed genes. This is mainly carried out by DNA damage binding proteins (DDB1, DDB2, XPC). The second repairs active transcribed strands preferentially, and is initiated by a stalled RNA polymerase, and therefore does not require the same detection proteins. Both pathways then recruit TFIIH, ERCC (2, 5, and 6) and other proteins to excise the damaged bases, using the correct strand as a template to synthesise new DNA and then relegate the lesion.

Homologous recombination is where genetic material is exchanged between two sister chromatids of the homologous pairs of chromosomes. This process is used to repair double stranded DNA breaks, which are acutely dangerous to the cell as they can result in genome rearrangements. In addition, it is also involved in meiosis, where it allows crossing-over of the two parental genomes. The MRN complex (Human) initiates the pathway, holding together the two broken double stranded ends, exonucleases then strip back the DNA to leave two single stranded 3' tails (or overhangs), which can then be used to "search" for similar sequences in a process called "strand invasion" (involving DMC1 and RAD51). Next, with double strand break repair (DSBR) or strand-displacement annealing (SDSA) can repair the missing regions. DSBR involves both 3' overhangs invading the homologous chromosome/sister chromatid, forming a double holliday junction, before extension and re-ligation. SDSA involves extension of the 3' invading strand along the recipient DNA by a DNA polymerase (Resnick 1976), which can slide along, preventing the need for holliday junctions, before DNA synthesis using the template.

Non-homologous end joining (NHEJ) is also used to repair double stranded breaks, but uses the (micro)homology of the single-stranded ends of the DNA, only when the sequences are completely complementary (Moore and Haber 1996). This provides the most common mammalian repair strategy for double stranded breaks (Guirouilh-Barbat et al. 2004). Other than DNA repair, it is also involved in V(D)J recombination

during the development of lymphocytes, that results in the highly diverse antibodies and T cell receptors required for adaptive immunity (Rooney et al. 2003).

In summary, there are distinct pathways of repair that the cell can use to remove specific lesions from the DNA, involving the recruitment of well-characterised proteins and cascades. However, DNA damage can have multiplying/pleiotropic effects, and be involved in various separate/linked pathways. This makes the study of DNA damage difficult to assign cause and effect, due to the complex interactions of molecular mechanisms.

Global response to DNA damage and downstream pathways

When cells are subjected to high levels of DNA damage, they can activate additional downstream pathways that can help give more time to repair lesions (by cell cycle arrest), or lead to apoptosis when a lesion cannot be repaired. The best examples are the MRN/ATM and ATR pathways.

In the first pathway, MRN (a complex comprising, MRE11, RAD50 and NBS1) recruits ATM (Ataxia-telangiectasia-mutated; a serine/threonine kinase) to double-stranded DNA breaks, holding the two ends together. ATM then auto-phosphorylates serine-1981, which promotes its activity by dissociation of ATM from inactive multi-mers/dimers to active monomers. Its activity also phosphorylates H2AX within minutes (Ayoub et al. 2008). It has also been shown that CK2 (Casein kinase 2), which is activated by DNA damage, can phosphorylate Thr51 in HP1 (also known as CBX1; a chromatin-bound protein), allowing greater mobility that in turn is important for the phosphorylation of H2AX (Ayoub et al. 2008). The phosphorylation of H2AX is widely used to measure the level of DNA damage in the cell, and despite being a downstream response (not a direct measure of DNA damage), it is considered very sensitive and correlates well with foci detection (Ivashkevich et al. 2011). Finally, ATM then can phosphorylate BRCA1, p53, Chk1/2, E2F1 and other important proteins, that in turn can lead to repair, cell-cycle arrest, chromatin remodelling and apoptosis as shown in Figure I.4.ii.

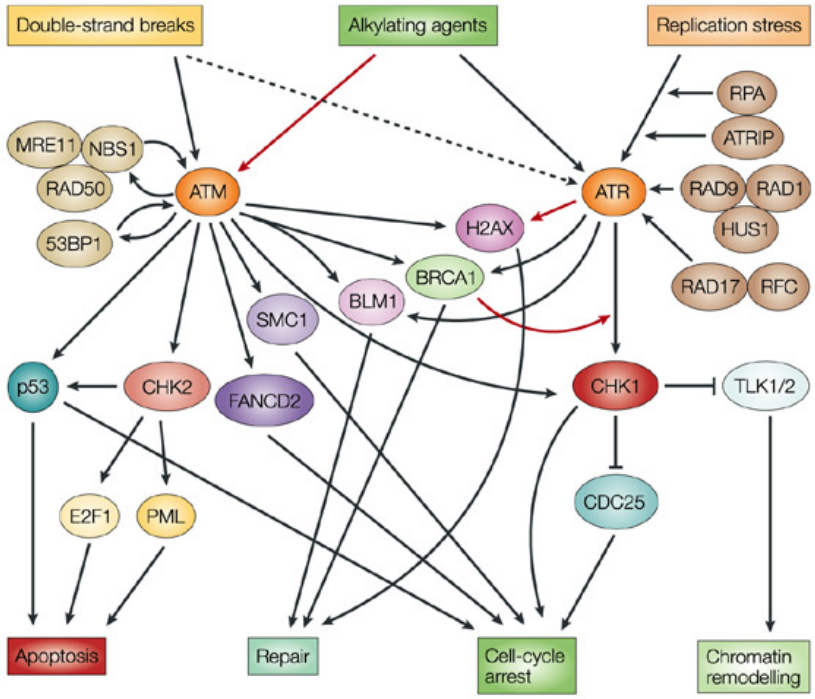


Figure I.4.ii. ATM and ATR (activation and downstream effects). For the three listed types of DNA related stress/damage, each lead to activation of ATM/ATR, leading to a cascade of protein activations, that ultimately affects apoptosis, repair, cell-cycle and chromatin-related functions. Arrows direction of activation. (B. B. S. Zhou and Bartek 2004)

ATR (ataxia telangiectasia and Rad3-related), another serine/threonine kinase, is a protein similar to ATM, but is mainly invoked during replication stress and single strand DNA breaks. ATR recognises single stranded DNA through an associated protein ATRIP, which activates at single stranded damage or during repair where single stranded DNA is formed (e.g. including in double-strand break repair). Similarly to ATM, it can phosphorylate downstream effectors leading to recruitment of repair proteins, arrest the cell cycle or chromatin remodelling.

Finally, the cell cycle is also linked to the DNA damage machinery, where at specific checkpoints, the cell can arrest if the cell detects damage that requires more time to be fixed, or even stop completely and go toward apoptosis, if the lesion cannot be fixed. Cell cycle is also important for the activation of specific DNA repair pathways, for example double-strand break repair through chromosome homologous recombination, must occur during or shortly after DNA replication in the S/G2 phase, but before

mitosis (Alberts et al. 2008). NJEH on the other hand is more predominant in the G1 (growth) phase, but can occur throughout the whole cell cycle (Rothkamm et al. 2003; Mao et al. 2008). Consistent with this, H2AFX staining of double-strand DNA breaks is much more apparent at the S/G2 phase of the cell cycle (Pilzecker et al. 2017), where the histone mark is most strongly phosphorylated by the ATM cascade during the cell cycle checkpoint (G2 to M) (Ichijima et al. 2005).

DNA damage during early mammalian development

During early mammalian development, we know relatively little about the exact DNA damage pathways that are active, yet expect that external DNA damage exposure must be similar to other internal cell types, and that any damage must be removed to ensure genomic integrity. However, gametes and early embryos are not like typical adult somatic cells, they have their own internal metabolism, suspended development during oogenesis, active DNA breaks during meiosis, rapid proliferation in sperm, fusion of two haploid genomes as well as total epigenetic remodelling and genome activation in the early embryo (to name just a few), which may contribute an extra 'load' to the system. Here, I describe a few examples of what is currently known about DNA damage and repair during this time.

Progenitors and Meiosis

During gametogenesis in both sperm and egg progenitors, we have the first stage at which the embryo is subjected to damage. In this case deliberate, in that the paired homologous chromosomes are fragmented by introducing double-stranded DNA breaks, and then subsequently repaired to fully unite the pairs during meiosis (Scott and Pandita 2006). These breaks also allow "crossing-over" of sister chromatids, the process in which segments of the homologous regions can swap branch of origin. Therefore double-strand break repair must be active during meiosis in both female and male progenitors.

The egg itself then has to wait for reactivation at ovulation stage in a semi-quiescent state, paused at meiosis I, until potential fertilisation, this can last 40 years or more (Carroll and Marangos 2013), at which time many oocytes are believed to undergo

apoptosis, presumably due to un-repairable DNA damage (Suh et al. 2006). Here, it is known that the primary oocytes are particularly sensitive to DNA damage, with 4-10 double-strand breaks being enough to lead to apoptosis (Suh et al. 2006), and this sensitivity is dependant on the concentration and activity of a isoform of p63 (a p53-homolog). Suh et al. 2006 found, that when they generated a mouse line without exons 2 and 3 of p63 (covering its transactivation domain), oocytes were more resistant to ionising radiation than wild-type oocytes. This mechanism provides a way for oocytes to be removed from the pool and helps explain why oocytes appear to endure the lowest levels of mutation/damage compared to somatic cells (Tichy and Stambrook 2008), as damaged cells are removed from the population.

At the germinal vesicle stage we begin to see changes in the sensitivity to damage, etoposide treatment results in higher intensities of gamma H2AX and RAD51 foci in mouse, yet have 3 to 6 fold less fluorescence intensity in rhesus monkey embryos (X. Wang et al. 2017), suggesting a difference in DNA repair ability between species (at least for HR repair).

Once a fully mature egg is activated for ovulation, it appears that mouse embryos too lose their sensitivity to DNA damage. Etoposide treatment at ovulation stage does not lead to the G2/M checkpoint activation (Marangos and Carroll 2012; Marangos et al. 2015; H. Wang et al. 2016) and levels of ATM/p63 are low (Marangos and Carroll 2012), but instead the spindle assembly checkpoint can be activated and arrests development at metaphase of the first meiosis (Marangos et al. 2015).

Sperm commonly have fragmented genomes, due to oxidative stress (J. Aitken and Fisher 1994), protamine imbalance (Ni et al. 2016), and abortive apoptosis (R. J. Aitken and De Iuliis 2009). The folding of the chromatin into toroids (compacted protamine ‘packets’), which reduce the genomic volume 6 times, itself is believed to protect the genome from assaults, however in order to compact it is believed active DNA breaks are need to reposition the chromatin (Laberge and Boissonneault 2005).
Once spermiogenesis is completed with protamination of the chromatin and transcription/translation stop, many of the normal DNA repair pathways are inactive, including HR, NJEH, NER and MMR, although BER appears to be functional (Jaroudi and SenGupta 2007).

Fertilisation, and the fusion of the gametes

Before fertilisation occurs, sperm must navigate and mature in the female reproductive tract. To capacitate (Ref-Spermiogenesis) requires reactive oxygen species (or ROS), which phosphorylates tyrosine, and leads to hyper-activation of the flagellum (increasing motility in order to reach the egg). Consistent with this, inhibition of SOD (superoxide dismutase) can inhibit hyperactivation (de Lamirande and Cagnon 1993). ROS is also required for the acrosome reaction (Ref-Spermiogenesis)(Donà et al. 2011). These processes may help explain why sperm harbour much DNA damage compared to the oocyte, as ROS can lead to oxidative DNA damage.

At fertilisation, sperm's highly damaged DNA needs to be un-compacted (replacing its protamine with histones) and repaired, adding a large source of DNA damage for the egg to resolve (given the sperm lacks cytoplasmic repair proteins) (Ménézo, Dale, and Cohen 2010). In addition, the paternal genome must be actively demethylated, first TET3 modifies 5mc to 5-hydroxymethylcytosine (5hmC), then the base can be excised and replaced by the BER machinery (Ladstätter and Tachibana-Konwalski 2016) or homologous-recombination (HR) repair in tandem. This damage must be removed, as shown by (Konwalski *et al.* 2016), that a conditional knockout of cohesin (an essential HR component), leads to Chk1-dependant checkpoint activation, delaying mitosis. This has been further confirmed by (Gawecka et al. 2013; B. Wang et al. 2013), that showed severe sperm damage does not lead to apoptosis, but does delays paternal DNA replication and therefore cell-cycle.

In addition, both fertilisation and meiosis have been associated with the high frequency of aneuploidies (chromosomal abnormalities, gain or loss of chromosome number) incurred during this time in development, due to mitotic (post-fertilisation) or meiotic non-disjunction (Harton and Tempest 2012) .

Box 5. DNA damage in Non-mammals.

Research into DNA damage/repair in non-mammals is characterised by an extreme insensitivity to DNA damage during the pre-ZGA stage, with many non-mammalian embryos able to endure high levels of damage and continue to divide. Yet, are still sensitive once DNA repair pathways are turned on at a later stage.

In Xenopus, the first 12 cleavage divisions are rapid and are without DNA damage checkpoints, rendering them insensitive to DNA damage insults (e.g. Ionising radiation). After the 12^{th} division the embryo transcribes its own genes for the first time (ZGA), gap phases are introduced into the cell cycle, checkpoints are activated and apoptosis can occur en mass if DNA cannot be repaired (Peng, Lewellyn, and Maller 2008). Research has shown that the DNA to cytoplasm ratio is responsible for checkpoint activation in Xenopus, in that injection of undamaged circular plasmid DNA can activate the checkpoint at the 10^{th} division (Conn, Lewellyn, and Maller 2004).

In zebrafish, the DNA damage checkpoint appears independent to ZGA, is maternally supplied, yet only the Chk2 and not Chk1 pathways can be activated prior to ZGA (M. Zhang et al. 2014).

In *Caenothabditis elegans*, DNA-damage checkpoint response is actively silenced by halting replication fork stalling (requires expression of rad-2, gei-17 and polh-1), allowing damage to be by-passed (Holway et al. 2006). They suggest that this mechanism is in place to allow the fast, uninterrupted cell divisions at this time (Holway et al. 2006).

Genome activation and cleavage stages

Until genome activation, the maternal mRNAs and protein control the fate of the embryo, so it has long been held that the maternal egg must be filled with all the required DNA repair machinery until the embryo is able to produce these proteins itself during zygotic genome activation, in order to preserve genomic integrity

(Jaroudi and SenGupta 2007). However, some studies have claimed DNA repair may be weaker until activation of the genome in mammals (Yukawa et al. 2007; X. Wang et al. 2017). In mouse, it has been shown that 1-2 cell stage mouse embryos are hypersensitive to ionising-radiation and X-ray irradiation due to a lack of DNA damage repair (Yukawa et al. 2007). The authors found that ATR and DNA-protein-kinases were present in the embryo, but could not detect gamma-H2AX phosphorylation (Figure I.4.iii), suggesting that the G2/M checkpoint and DNA repair are weaker at this time (Yukawa et al. 2007). Recently, it has also been shown that Cas9 could induce double-stranded breaks in the early embryo, but this did not activate homology-directed repair from 2- to 4- Cell stage mouse embryos (Grunwald et al. 2019). Finally In human, it has been suggested that specific cell-cycle checkpoint genes may be absent or dysfunctional (Lee and Kiessling 2017), leading to an increase in cell cycle speed.

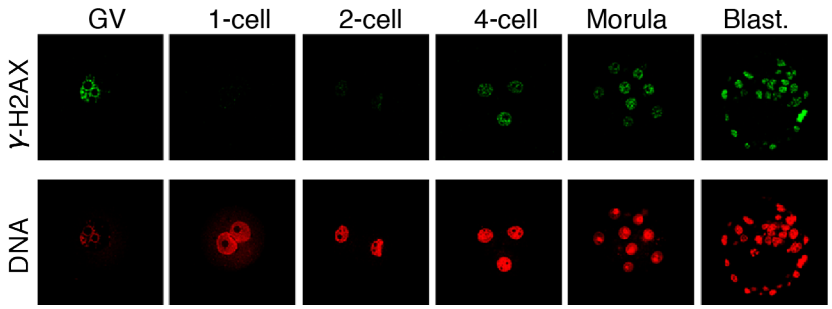


Figure I.4.iii. Lack of γ-H2AX phosphorylation in irradiated early mouse embryos from 1-Cell to 4-Cell. (Yukawa et al. 2007).

In contrast, it has been shown that mouse embryos at cleavage stage do respond to some forms of DNA damage, specifically (Z. W. Wang et al. 2013) could show that a UVA microbeam could induce double-strand DNA breaks in the pronucleus to 8-cell-stage embryos, and activate gamma-H2AX phosphorylation, leading to cell-cycle arrest and apoptosis by the blastocyst stage, with failure of incorporation into the morula. (X. F. Mu et al. 2011) could also show that 2-cell mouse embryos are sensitive to UV and cisplatin induced DNA damage, with increased phosphorylation of gamma-H2AX and arrest of the cell cycle at G2/M for UV and G1/S for cisplatin. Therefore, the ability of the embryo to respond to damage appears different depending on the specific lesion.

In addition to DNA damage, embryos subjected to stress, e.g. hydrogen peroxide (H2O2; Vandaele et al. 2010) or heat (Ozawa, Hirabayashi, and Kanai 2002; Sakatani, Kobayashi, and Takahashi 2004) are known to have different responses in cleavage stage embryos. In cow, early embryos have lower heat tolerance compared to the blastocyst stage (Sakatani et al. 2004), which has been associated with antioxidant levels in the embryo (Figure I.4.iv) (Sakatani et al. 2004). Therefore in cow at least, some stress response pathways may not be activated, and this could relate to expression of specific genes at ZGA (Figure I.4.iv).

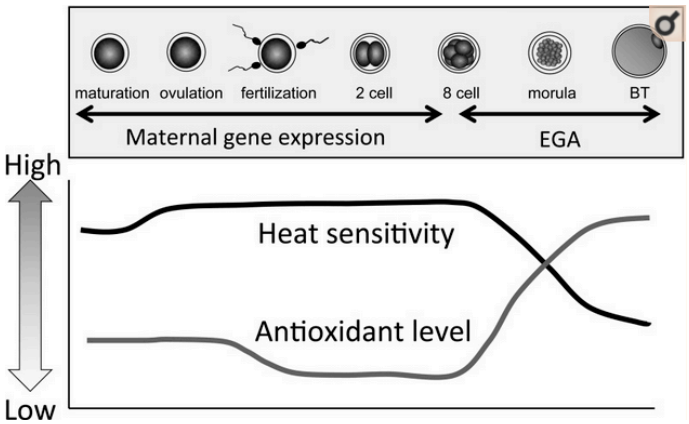


Figure I.4.iv. Levels of antioxidants and heat sensitivity in bovine embryos from oocyte to blastocyst. Figure from (Sakatani 2017)**.**

Finally, the act of genome activation itself has also been linked to double strand breaks in *C.elegans* (Butuči et al. 2015). Inhibition of RNApolII was able to reduce the number of RAD-51 foci, whereas inhibition of translation or splicing did not have the same effect (Butuči et al. 2015). It was also reported in *Drosophila melanogaster*, that transcriptionally engaged DNA triggers ATR/Chk1-dependant checkpoints (Blythe and Wieschaus 2015), and that the reduction of RNApolII recruitment reduced replication fork stalling. Whether this effect is pronounced in the mammalian system is currently unknown, but could suggest genome activation alone may be a source of DNA damage during mammalian development.

Transposable elements and DNA damage

Yet another curious factor potentially responsible for DNA damage in the embryo are transposable elements (transposons); mobile genetic DNA elements that are able to replicate and insert in other places in the genome. It has been shown that particular families of transposons are up-regulated during preimplantation development, including long interspersed nucleotide element (LINES; Kano et al. 2009), endogenous retroviruses class III transposons (ERVL; (Peaston et al. 2004; A. Inoue, Matoba, and Zhang 2012)) and others (Figure I.4.v).

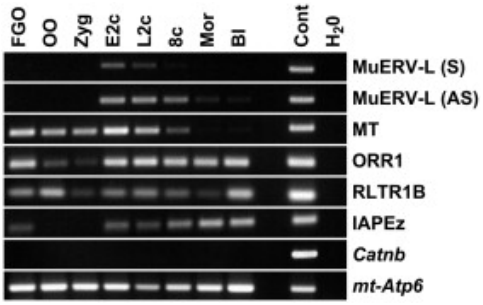


Figure I.4.v. Transposon expression levels during mouse pre-implantation development. Expression of six transposons: MuERV (mouse ERVLs) in sense (S) or antisense (AS) mode; MT (mouse transcript non-autonomous retrotransposon); OOR1 (Origin Region Repeat 1 non-autonomous retrotransposon); RLTR1B (subfamily of ERV1) and IAPEz (Intra-cisternal A-type particle provirus); with two negative and positive controls (Catnb and mtAtp6). From stage: FGO (Full-grown oocyte); OO (ovulated oocyte); Zyg (zygote or 1-cell embryo); E2c (early 2-cell); L2c (late 2-cell); 8c (early 8-cell); Mor (morula); Bl (blastocyst). Figure from (Peaston et al. 2004).

Their up-regulation has been correlated to global demethylation of the genome, suggesting that more transposons are expressed at this time, because they are not being repressed by methylation. This could lead to more transposons inserting into the genome at this time. However, (A. Inoue, Matoba, and Zhang 2012) could show that the maternal and paternal genomes had equal levels of transposons expression (despite methylation differences) and Tet3 (needed for demethylation) siRNAs did not lead to suppression of transposons in preimplantation embryos. Suggesting that methylation may not be the cause for the burst of transposon expression (A. Inoue, Matoba, and Zhang 2012). Whatever the cause, transposon expression could lead to genomic instability or DNA damage, for instance, LINE-1 expression has been linked directly to double-strand DNA breaks (Gasior et al. 2006).

Further, (S. Kim et al. 2014) could show that Prmt5, which normally translocates to the nucleus during demethylation, prevents LINE1 and IAP transposon expression, by regulating H2A/H4R3me2 chromatin marks. Conditional knockdown of Prmt5, led to

higher expression of the transposon families and a DNA-damage response. This suggests that active control of transposons is important during chromatin remodelling and requires specific genes to control genomic stability.

Stem cells

ESCs are particularly well-protected from DNA-damage, and are the most well studied cells in the early developmental time-course (Vitale et al. 2017). Due to their high proliferation rate and shortened G1 cell cycle phase, DNA damage is a concern. Indeed, the G1 checkpoint has been documented to not respond to lesions, yet the G2 phase checkpoint is active (van der Laan et al. 2013). ESCs are believed to be highly resistant to mutation and damage, due to elevated levels of DNA repair activity including, RAD51 (Ahuja et al. 2016; Tichy et al. 2010), KHDC3 (which promotes PARP) (B. Zhao et al. 2015), low levels of ROS due to metabolic rewiring to preferentially use glycolysis instead of oxidative phosphorylation (Xu et al. 2013) and preferential use of the high-fidelity HR repair pathway, which is present at G2 but not G1 checkpoints (Tichy et al. 2010). If damage is irreparable, then stem cells avoid propagation by losing pluripotency, differentiating and undergoing apoptosis (Vitale et al. 2017). DNA lesions such as etoposide, topoisomerase II inhibitors, UV-exposure and γ-irradiation, for example, lead to rapid clearance activation through apoptosis (Filion et al. 2009). Therefore, stem cells have a good ability to repair DNA damage and also have better sensitivity to damage, in order for apoptosis to remove damaged cells.

Aneuploidy

From the clinical side of early development, the main human pathology of early embryos grown *in-vitro* or *in-vivo*, is the presence of additional chromosomes or loss of chromosomes (or aneuploidy). Aneuploidy can occur at meiosis, but is much more frequent in the subsequent mitotic divisions after fertilisation, with an apparent clearance of abnormal cells by the blastocyst stage (Babariya et al. 2017). This has been used as evidence (in human) that the cell cycle may be more relaxed in the first few cell divisions, allowing DNA double-strand breaks to persist (Babariya et al. 2017).

Further, (Bazrgar et al. 2014) could show that poor quality embryos with aneuploidy, had a more active DNA repair pathway, with five genes specifically overexpressed: MSH3, XRCC1, RAD50, LIG1 and CDK7.

Conclusion

In non-mammalian species, the pre-ZGA embryo is characterised by fast cell cleavage cycles, a lack of checkpoint activation and insensitivity to DNA damage. Strikingly, in mammals, there is some evidence that before genome activation, embryos may be insensitive to some DNA damage lesions or stress. This is clearly shown in early mouse embryos, which show reduced phosphorylation of gamma-H2AX (Yukawa et al. 2007), which is a reliable marker of a response to double-stranded DNA breaks. However, there are some inconsistencies in the literature, where some authors detect a response to DNA damage (Z. W. Wang et al. 2013; X. F. Mu et al. 2011). The prevalence of aneuploidy fits well with the suggestion of lower DNA repair at this time. In the oocyte and blastocyst stages however, DNA repair efficiency appears intact. This apparent lack of DNA damage response in the pre-ZGA stages is discussed further on p.140.

OBJECTIVES

To characterise the role of alternative splicing in early mammalian development we began with the following set of objectives:

Section 1: Profile alternative splicing and gene expression changes during pre-implantation development in human, mouse and cow.

Section 2: Characterise the affect of splicing on the open reading frame (ORF) and protein synthesis.

Section 3: Identify splicing regulators responsible for isoform changes detected.

Section 4: Explore the role of the RNA-binding protein 'Snrpb/d2' in early development, and their potential links to DNA damage repair.

Section 5: Determine the extent of splicing conservation and the species-specific differences.

Section 6: Explore the functions and inclusion patterns of embryonic stem specific exon-skipping events in early development

RESULTS

Chapter 1. Alternative splicing dynamics in the embryo.

To explore mammalian pre-implantation splicing dynamics, we took advantage of the abundant RNA-Seq datasets publically available on the Sequence Read Archive (SRA), from oocytes to cleavage and blastocyst stage embryos, from either single-cell or pooled embryos (Figure R.1). Specifically, we focused on human, mouse and cow, as these species had multiple comparable early embryo RNA-Seq datasets that represent phylogenetically disperse taxa, spanning >90 million years (Lucas et al. 2018). This allowed us to make.evolutionary comparisons of functionally equivalent morphogenetics stages.

In order to give an overall impression of the changing transcriptional landscape of the early embryo, this section focuses on the assimilation of RNA-Seq data from the early embryo (1.1), levels of isoform diversity (1.2), global patterns in isoform change (1.3), the relationship between splicing/transcription (1.4) and single cell heterogeneity (1.5).

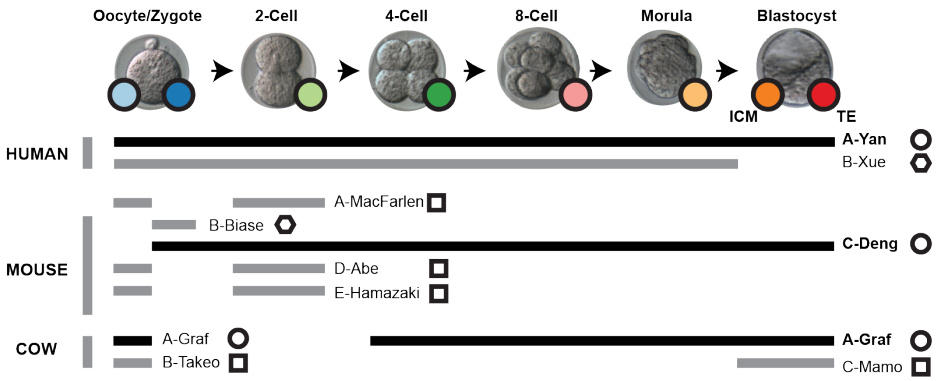


Figure R.1. Overview of the main source of RNA-Seq reads from oocyte to blastocyst in the three species. Each line represents a publically available dataset; the black line represents the main source of reads we used in each species. Colour codes are given for the developmental stages, and shapes are given to each experiment as a key for subsequent plots.

1.1 Measuring splicing during early development.

To explore isoform abundance during early embryo development, we ran a simple pipeline to process all compatible RNA-Seq reads (Figure 1.1.a). See methods, page 155, for more details.

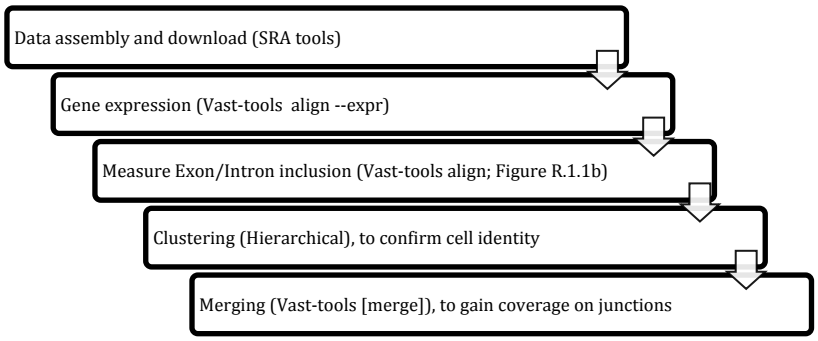


Figure R.1.1a. Main steps in data processing. From RNA-Seq downloaded from the Sequence Read Archive (SRA) to the measurement of exon/intron inclusion for each gene. The main tools are listed at each level, with a full description of scripts and programs used in methods.

Briefly, we started by downloading raw RNA-Seq reads from the Sequence Read Archive (SRA). A summary of samples is shown in Figure R.1 (and Supp. Table 1). Reads had to be from Illumina sequencing platforms, covering the full length mRNA (e.g. SmartSeq2), at least 50 nucleotides long and without 3'bias associated with some sequencing protocols to be comparable in this analysis (See discussion: p.132).

Next, we used vast-tools (Irimia et al. 2014), to calculate gene expression levels in each cell/sample, and to measure percent inclusion of alternative-exons (or cassette exons), 3/5' splice site usage and intron retention events (Types shown in Figure R.1.1b; UPPER). This used events from vast-tool's pre-built database of exon-exon and exon-intron junctions. In addition, we also ran a novel exon discovery pipeline: SUPPA (Alamancos et al. 2014) and custom scripts, using our early pre-implantation data, in order to ensure that all exon-exon junctions were measured (See methods: p.156). For each alternative exon or 3/5' splice-site, we assigned a percent-spliced-in (PSI) value and for each exon, and a percent-intron-retention (PIR) value for each intron (both spanning 0-100% inclusion). For details see Figure R.1.1.b (LOWER).

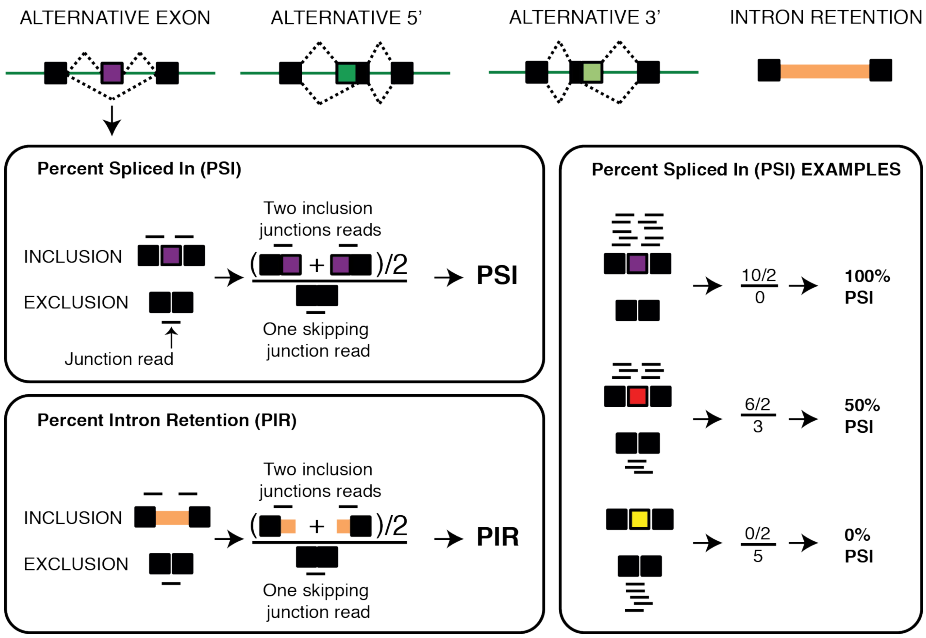


Figure R.1.1b. Types of alternative splicing and calculation of exon/intron inclusion (simplified). TOP: Diagram of four main subtypes of splicing event mentioned in this thesis. BOTTOM LEFT: Percent spliced in (PSI) is measured for alternative exons (cassette exons) and alternative 3/5' splice site, and percent intron retention (PIR) for introns**.** PSI and PIR are calculated by taking half the sum of the two inclusion junction-reads and dividing by the number of skipping junction-reads. However, this is a simplification, vast-tools calculates PSI from all possible inclusion and exclusion junctions, as two alternative exons can be in tandem, see Supp. Figure 1.1.iii**.** BOTTOM RIGHT: Three cassette exon examples (with reads mapping to inclusion/exclusion junctions) and their PSI values are shown.

Next, we used hierarchical clustering (Supp. Figure 1.1.i) of gene expression values to determine the identity of the samples, to the various stages. This was done, for example, to ensure that 2-cell labelled blastomeres cluster with other 2-cells and not with 4-cell stage embryos, despite being from multiple independent experiments. At this stage we also detected anomalous cells that could represent dying, damaged, mis-labelled cells or those taken from incoherent time points (details in column J, Supp. Table 1), these were removed from further study.

Finally, based on our hierarchical clustering (Supp. Figure 1.1.i), we then merged single cells/samples into multiple pools of around 150 million reads, in order to obtain higher coverage on the exon-exon/exon-intron junctions (around 2-8 cells were merged to form a pool with sufficient coverage). High coverage was necessary to

confidently measure inclusion levels for as many exons as possible, given the single cells had low coverage.

Using the finalised groupings, we clustered by Principle Component Analysis (PCA) the gene expression (GE) values (Figure R.1.1.c; TOP) and alternative-exon percent-spliced-in (PSI) values (Figure R.1.1.c; BOTTOM) for the three species. These plots show that cell stages have unique signatures in terms of GE and alternative-exon abundance, and separate by stage and not by experiment. Clustering by intron retention or 3/5' splice site usage could also separate stages, but to a lesser degree (Supp. Data 11). Overall, GE and alternative-exons have similar temporal dynamics (U or V shape), with the largest distance between consecutive stages being between pre- and post- zygotic genome activation (ZGA) in the three species (Figure R.1.1.c); for human between the 4 and 8-cell stage, mouse between zygote and 2-cell stage and cow between 4-cell and morula (with cow activated slighter later than in human, with 8-cell not fully activated). In terms of exon skipping, considerable variation exists between samples at genome activation stage specifically, as represented by the looser clustering of samples at this stage. Finally, expression clustering has the highest difference between oocyte and blastocyst stages (mostly PC1; >40% variance), whereas alternative-exons show a high similarity between these two stages (mostly separated by PC2; <25% variance). This suggests that the oocyte does not change progressively to a blastocyst alternative exon state, but that many events show a transitory state at genome activation stage before returning to the original.

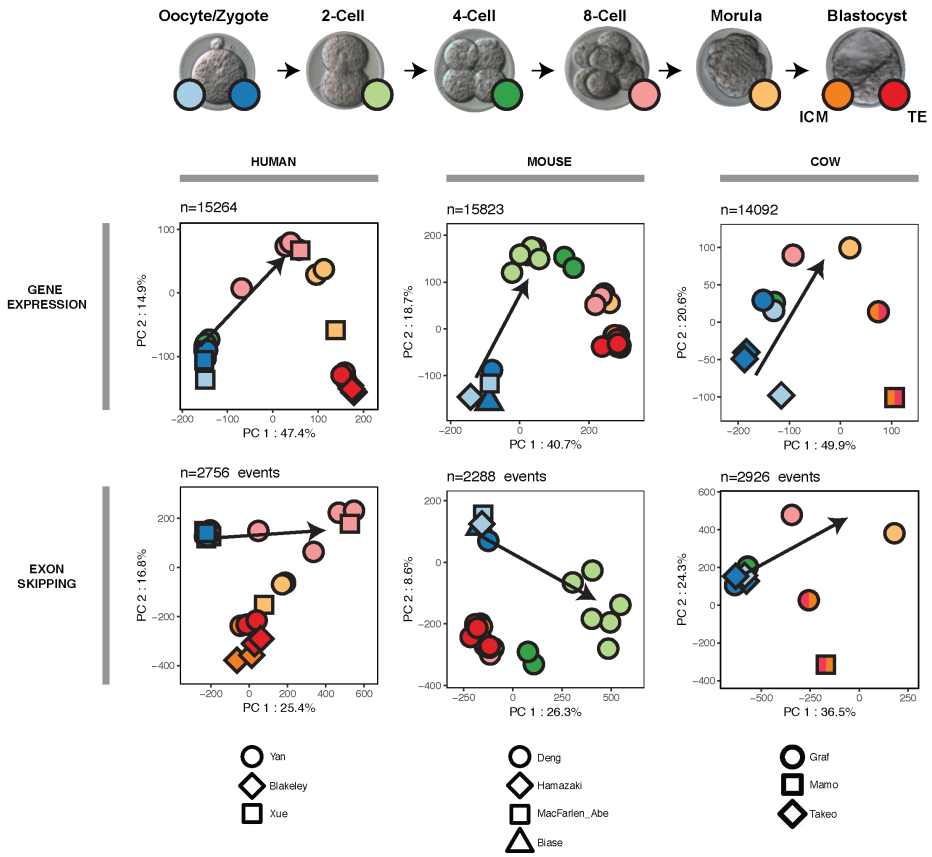


Figure R.1.1.c Principle Component Analysis (PCA) plots for the merged groups, in human, mouse and cow (left-to-right), using either a matrix of gene expression values in cRPKM or alternative exon PSI values (from 0%-100% inclusion), for events that have coverage in 80% of samples. Showing principle components 1 (PC1) and 2 (PC2), and the percent of variation each component represents. Shapes represent experimental origin and colour represents developmental stage, with the black arrow showing point of genome activation.

1.2 Splicing changes during early development

We next measured the number of splicing events that change between each consecutive stage in our time course. For this we required a minimum of two replicates per condition/stage in human/mouse and one in cow (due to the lower number of replicates available), and variable dPSI values (change in PSI between the two condition means), depending on a few conditions (Figure R.1.2.a).

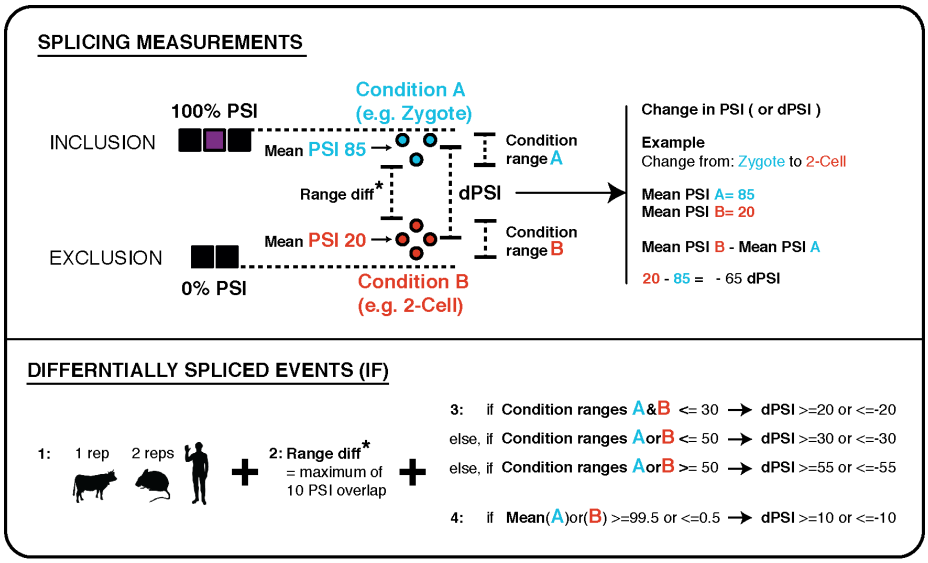


Figure R.1.2.a Calculation of differentially spliced events. For each alternative exon/intron, we can calculate the mean PSI per condition, the condition range (within A and within B), the overlap between ranges (Range diff) and the delta PSI (dPSI; or change in PSI between conditions). This is calculated by subtraction of means in the direction of the transition (in the example, zygote and 2-cell stage). dPSI runs from -100 to 100 between any two conditions. Using these values, we call an event differentially spliced, **IF: 1**. There is at least 1 replicate/condition in cow, and two for human and mouse, **2.** The range diff must not overlap more than 10 PSI, and either **3.** Given condition ranges (of A or B being > or <), different dPSI values (>20/>30/>55) are required to be called differential, or **4**. Events where one condition has a high or low PSI and the other differs (mean diff PSI >10) are also called; these are species cases, as they involve the change from 1 to 2 isoforms between two conditions (or the reverse).

Using these conditions, we found 2711, 1828 and 4748 events (Alternative exons, 3/5' ss and intron retention) changing in any of the consecutive stages (e.g. from oocyte-zygote, or morula-blastocyst), within 2058, 1350, 2735 genes, in human, mouse and cow respectively. For comparison, we also calculated the numbers of

differentially expressed genes (see Methods: Differentially expressed genes; p.158) and found 6545, 8895, 8118 genes showed differential expression in at least one consecutive transition for the three species. The largest number of changes in terms of splicing events and differential expression is at ZGA (Figure R.1.2.b). For example in human, 1130/2711 (41.7%) of AS events with differential regulation and 3826/6545 (58.5%) of genes differentially expressed, change at ZGA (4 to 8-cell stage). Moreover, a clear bias was observed in the direction of regulation for alternative exons and intron retention events in the three species at genome activation: whereas the majority of exons showed increased skipping, most differentially spliced introns had increased retention (Figure R.1.2.b). The differences in numbers (and types) of events between species is difficult to interpret, as each species has very different coverage and number of replicates per stages (Figure R.1.1.c), which leads to significant bias. For instance, the much larger amount of intron retention events detected in cow at ZGA is based on just one replicate for each stage (4 and 8 cell stages), and a mixed oligo-dT and random primer protocol, that can pick up unspliced pre-mRNAs. Also, in mouse for example, we have two replicates for zygote and six for 2-cell stage, meaning that we end up with potentially lower numbers of detected events, because we require two replicates with coverage to be considered, plus adding more samples means the variation within a stage will be greater. However, we can conclude that more AS events are detected at genome activation, at which time there is more exon-skipping of alternative exons and intron retention.

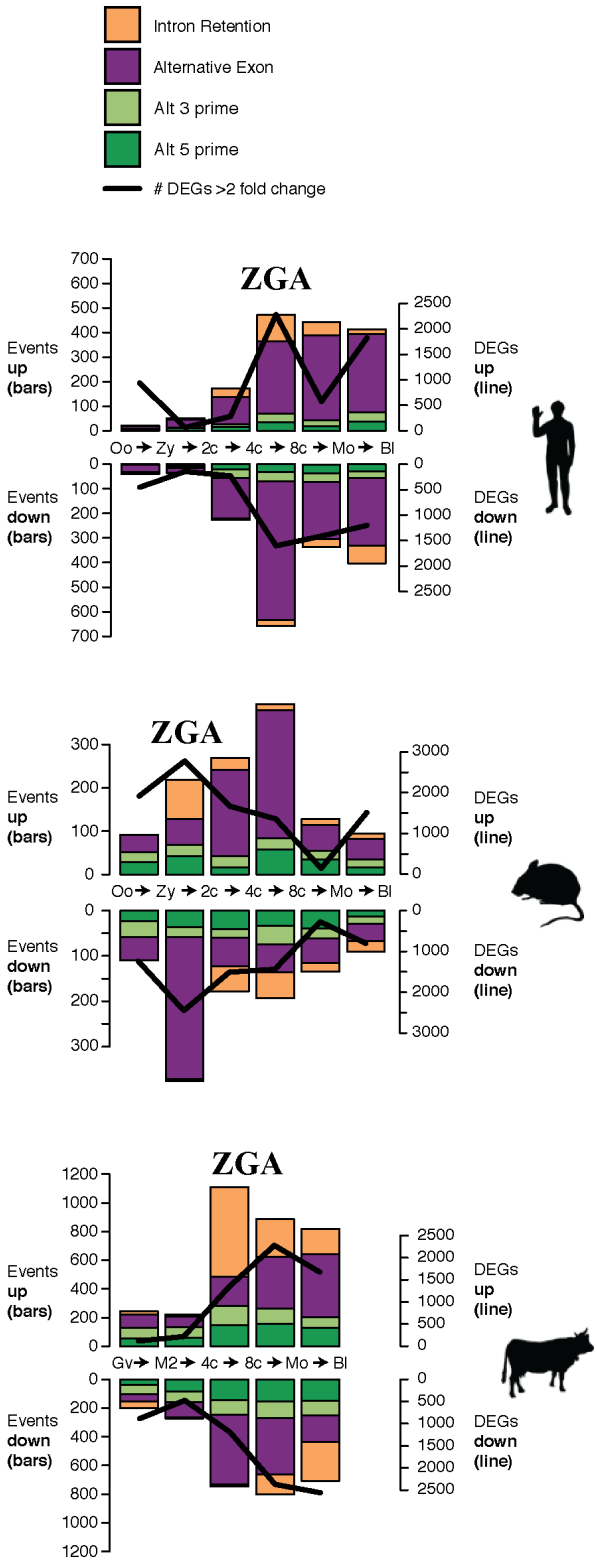


Figure R.1.2.b. Splicing and expression changes per transition. Numbers of alternative exons, alternative 3/5' splice site and intron retention in consecutive pairwise transitions. Number of differentially expressed genes (DGEs) represented as a black line (joint for visualisation). Events and DGEs have been split into increasing (top) or decreasing (bottom), meaning up/down regulated genes or events becoming more included (up) or skipped (down) for alternative splicing events. Stage of initial major genome activation is marked "**ZGA**", 4-8cell (human), Zy-2c (mouse) and 8-cell stage (cow).

To further characterize the temporal dynamics of alternative exons, we used Mfuzz (Kumar and E Futschik 2007) to cluster exons with sufficient read coverage throughout the time course, by their temporal change in PSI (Required a minimum of 5 stages with coverage and imputation, see methods p.159). We obtained 28, 18 and 22 exon clusters in human, mouse and cow, respectively (Supp. Fig 1.2.i). Most of these clusters could be broadly classified into three general patterns: (i) peak-like regulation, in which the exon is highly included or skipped only in a given stage, quickly returning to the initial levels, (ii) shift-like regulation, in which the exon goes from high to low inclusion, or vice versa, at a specific transition, but does not return abruptly to the initial level and (iii) other, those that do not fit the previous descriptions (Figure R.1.2.c; Supplementary Data 1[ClusterType]; for definitions of shift/peak-like, see methods p.159). These patterns were highly validated by RT-PCR using RNA from independent pools of embryos, with 7/9 (78%) of peak-like, 17/18 (94%) of shift-like and 15/17 (88%) other-like alternative exons showing consistent temporal dynamics (Supp. Fig 1.2.ii; Experiments by Barbara Pernaute). Remarkably, most alternative exons showed peak-like behaviour (Figure R.1.2.c; e.g. 42.1% vs 17.3% of shift-like exons in human). This strongly contrasts with the behaviour of gene expression clusters, which are more represented by other or shift-like patterns (Figure R.1.2.c), and suggest that the regulatory principles of both transcriptomic layers are fundamentally different. Furthermore, these results are consistent with the asymmetric patterns observed for exon skipping and intron retention at ZGA, which were inverted in the post-ZGA transition in the three species (Figure 1.2.a), suggesting a large proportion of changes at this time are temporary. Indeed, 427/749 (57%) alternative exons with significant changes at ZGA covered in the “Mfuzz” cluster analysis (Supp. Table 1) show peak-like behaviour.

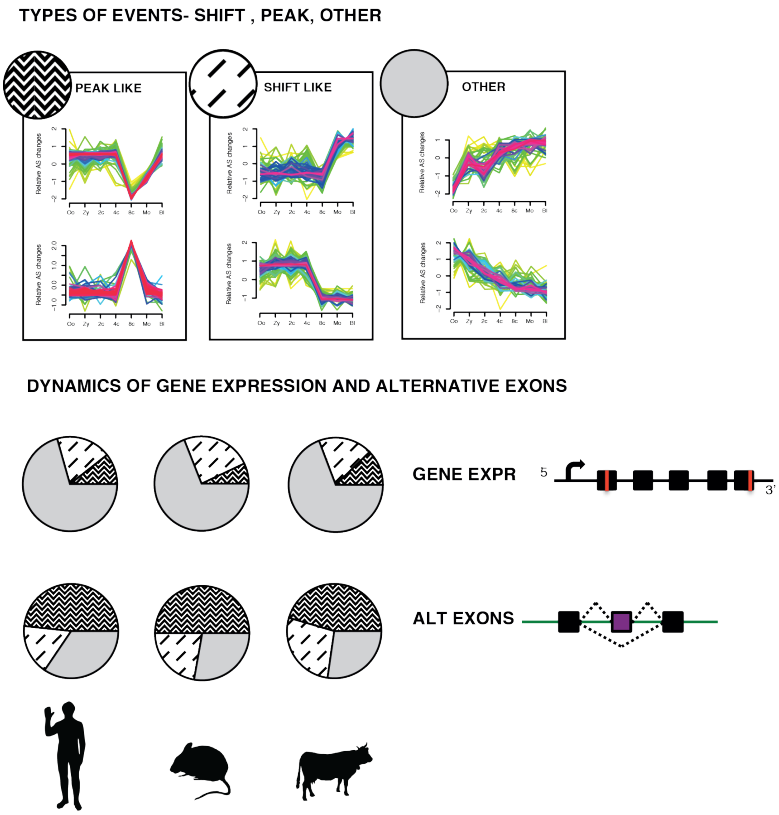


Figure R.1.2.c. Examples of the three main cluster types and proportions within differential genes and alternative exon changes. TOP: Example Mfuzz clusters showing the different types of pattern, which can be described as: Peak-like, Shift-like or Other (Examples from human alternative exon data). x-axes show the stages from oocyte to blastocyst, y-axes show relative PSI change. **BOTTOM:** Proportions of each type of cluster for gene expression clusters (upper) and alternative exons clusters (lower) with coverage in all time-points; for human, mouse and cow.

1.3 Transcriptomic diversity at genome activation

Given thousands of alternative exons were shown to contribute to the difference between stages, especially at genome activation (Figures R.1.1.c and R.1.2.b), we wanted to assess the contribution that AS has to diversify the transcriptome at each developmental stage. For this, we used a simple measure of diversity (Figure 1.3 (A); Full details in Methods p.158), where AS events with sufficient read coverage were classified as either producing one main isoform (PSI ≤ 20 or PSI ≥ 80) or two (20 ≤ PSI ≤ 80), at each stage. Remarkably, the stage immediately after the ZGA distinctly showed the highest levels of isoform diversity in the three studied species (8C in human, 2C in mouse and 8C/Morula in cow), returning to the preceding lower levels in the subsequent stages (Figure 1.3 [blue bars]). Remarkably, the level of diversity of alternative exons upon ZGA was not matched by any differentiated cell or tissue type, including neural, muscle and testis (Figure 1.3 [black bars]; tissue data from Supplementary table S1 of (Tapial et al. 2017)), which are known to have particularly high levels of alternative splicing (Barbosa-Morais et al. 2012; Soumillon et al. 2013). The differentiated tissues are also composed of several cell types, so one may expect higher diversity. We could also show that the increased level of AS among cassette exons was robust to different cut-offs of PSI range and read coverage (Supp. Figure 1.3.i.), only slightly decreased when looking at the single cell level (Supp. Figure 1.3.ii) and was not observed for intron retention or alternative 3'/5' splice sites (Supp. Figure 1.3.i [top]). Altogether, these results reveal that transcriptome diversification driven by changes in alternative exon skipping/inclusion may reach its maximum in mammals for a brief moment during zygotic genome activation.

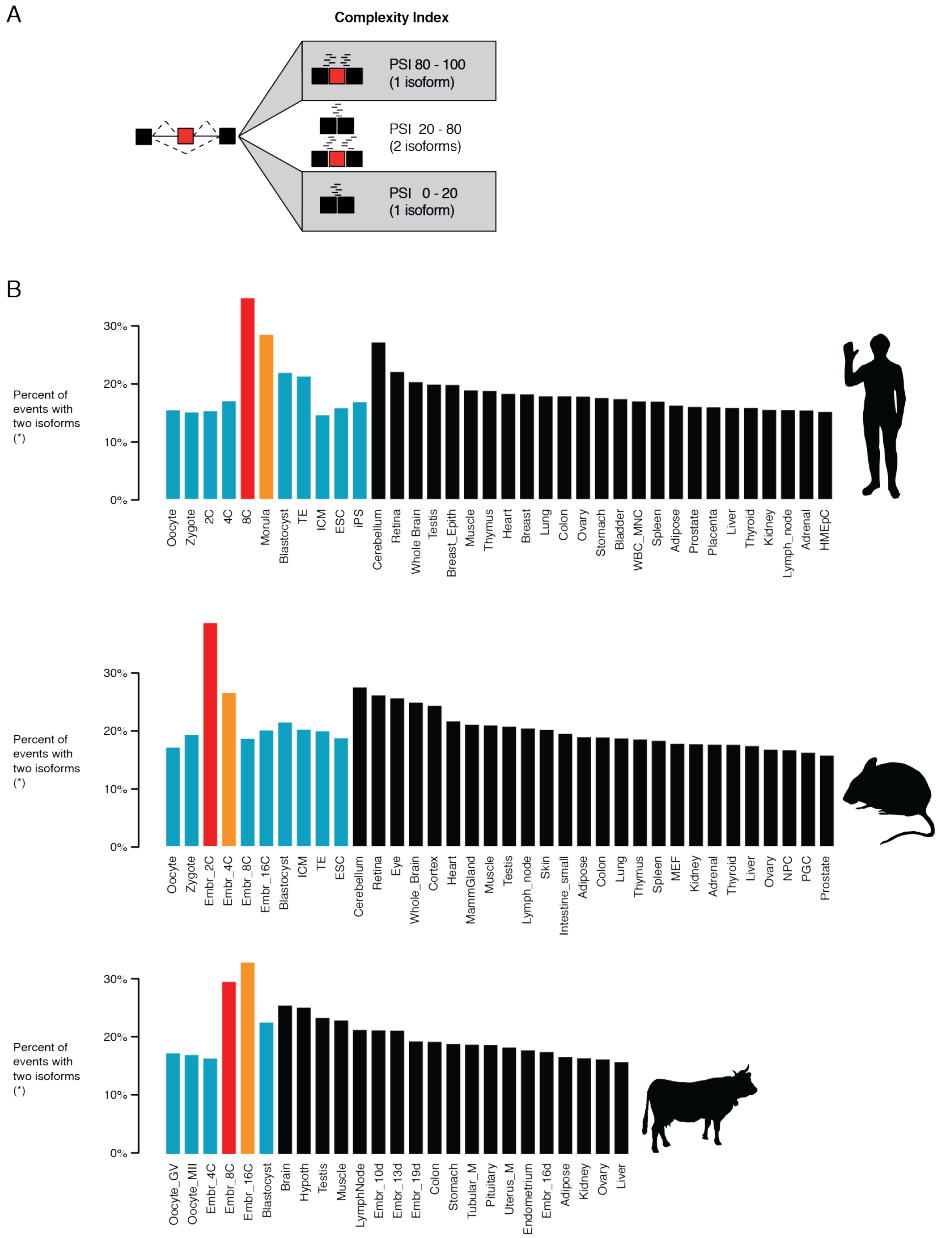


Figure 1.3. **Transcriptome complexity** (A): Given the simple complexity score given, we can calculate the transcript diversity, (B) as measured by % of AS events with two isoforms (with coverage VLOW: see methods), in early development (blue) and in selected differentiated tissues (black). Genome activation stage is highlighted (red), and stage after (orange). Using a PSI range of 20-80 (-2 isoforms) and exon-skipping events only. More tissues/cell lines are shown in Supplementary Figure 1.i.

1.4 Relationship between alternative splicing and transcription

Next, given the majority of differentially expressed genes and differential alternative exons happen at the ZGA stage (Figure 1.2.b), we wanted to determine if this could be linked to RNApolII transcriptional activity. Given splicing is co-transcriptional (Merkhofer, Hu, and Johnson 2014)(Carrillo Oesterreich, Preibisch, and Neugebauer 2010), increased transcriptional activity (such as that at ZGA), could lead to changes in inclusion of exons within the gene.

In our data, we can show that for alternatively exons (Figure R.1.4) and IR (Supp. Figure 1.4.i) at each transition, most events occur in genes that have little or no gene expression change (Figure R.1.4 [light grey]). When calculating the number of events expected by chance (see Figure R.1.4 legend), we do not find any significant enrichment (Figure R.1.4: Triangles), though some stages are depleted. This may suggest that changes in mRNA expression are not likely to play a role in splicing changes. However, the correlation between the total number of differentially expressed genes and total number of splicing events, appears to be linked to the need for transcription to replace the existing mRNA isoform, with the newly transcribed one (i.e. you need transcription [or degradation] in order to see a change isoform abundance).

Interestingly, many events that had a strong switch in isoform at ZGA, showed almost no change in overall gene expression levels. For instance, we identified 77 human genes with an expression fold change of less than 1.5 at ZGA, that had at least one alternative exon (n=84) showing a change in PSI of greater than 50 (Supp. Figure 1.4.ii; Supp. Table 1.4B). This indicates an active switch of the major isoform that is expressed after ZGA, without significant changes in expression. An example is Smg7 (Supp. Figure 1.4.ii), which shows almost no change in expression level, but a clear switch in isoform, which could be validated (Supp. Fig 1.2.ii (Part II)). This finding highlights the importance of measuring isoform changes, not just total gene expression levels. It also means previous estimates of numbers of genes expressed at genome activation may be lower than predicted, given isoform changes suggest the gene must be expressed at this time, even though relative mRNA levels do not change.

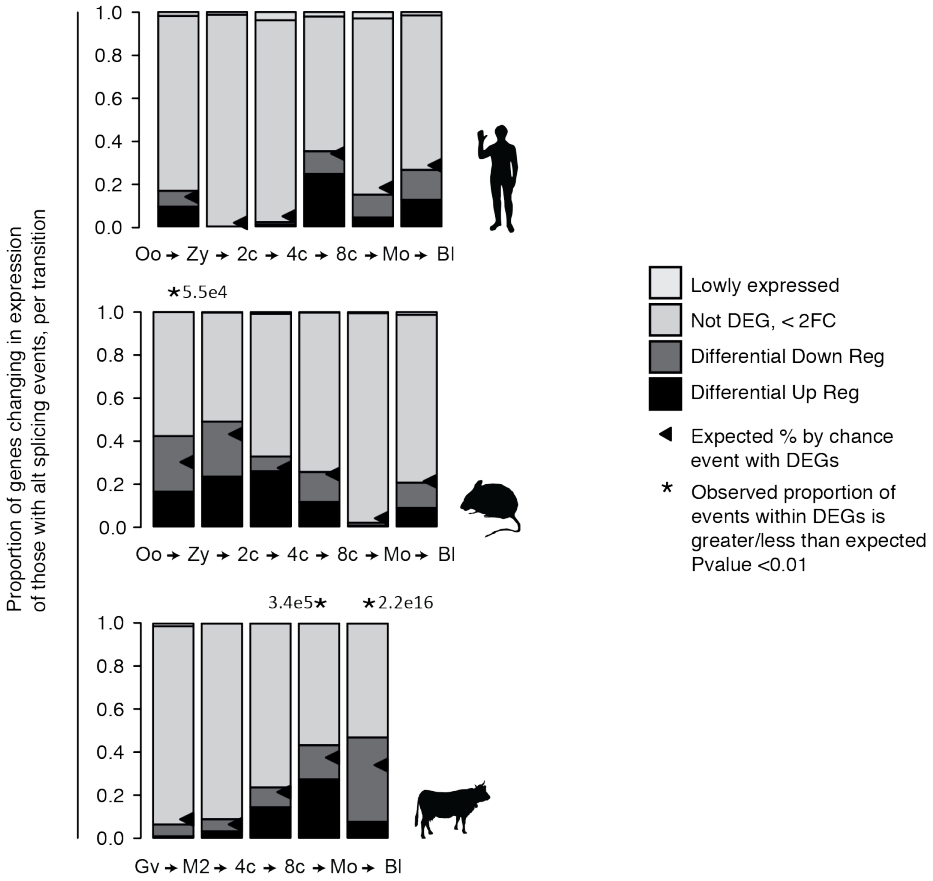


Figure R.1.4. Number of events in differentially expressed genes (DEGs). Showing the proportion of events per transition in different categories of genes. Those in up-regulated genes (black), down-regulated genes (dark grey), genes showing little change (Not DEG; light grey) and in lowly expressed genes (not tested for differential expression [cRPKM < 2 in both stages of transition]; white). Triangles represent the number of events expected by chance to be within differentially expressed genes. This percentage was calculated by taking the number of DEGs over the total number of expressed genes per stage.

1.5 Single cell variation.

Despite the need to merge single cells to acquire sufficient coverage of exon-exon/intron junctions and accurately measure splicing changes (See Methods: p.**Error! Bookmark not defined.**), highly expressed genes could be measured at the single-cell level. Using only single cells from (Yan et al. 2013) (Human) and (Deng et al. 2014) (Mouse), PSI calculations for cells within the same embryo (4Cell to blastocyst) or between embryos (oocyte to 2-Cell) could be used to calculate the bimodality (using Diptest in R) for each splicing event. 237 human and 307 mouse events were found in at least two embryos (Supp. Figure 1.5). An interesting example is shown in Figure R.1.5, where a 48bp exon in Yap, has differential inclusion within cells of the same embryo (c# in Figure R.1.5). Normally, a cell has either the inclusion or the skipping version, but less frequently both. Additional validations (Not published in this thesis) are underway to validate some of the more interesting events (Unpublished work: Barbara Pernaute).

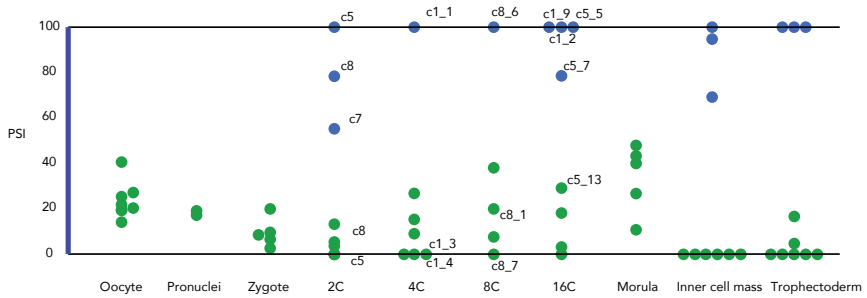


Figure R.1.5. Single cell variation in a 48bp exon-skipping event (MmuEX0052491) in Yap, through mouse embryo stages. Points represent single cells, with some annotated to highlight cell identity where some embryos have both inclusion or exclusion isoforms in different individual cells, but less frequently both.

Chapter 2. AS effect on the open reading frame and protein synthesis

Given thousands of isoforms change during development, we next wanted to explore the effect on the open reading frame (ORF) and protein level. To investigate this, we first documented changes in the coding sequence. Second, assessed whether or not changes could lead to non-sense mediated decay (NMD). Third, tested whether isoform disruption could lead to lower translation.

2.1 Changes to the ORF.

To begin elucidating the potential functional impact of AS changes during early embryogenesis, we investigated the effect that alternative exon inclusion or exclusion has on the open reading frames (ORFs) at each stage. For this, we divided AS events into four categories: (i) those that are predicted to generate alternative protein isoforms upon inclusion/exclusion; (ii) and (iii) those that disrupt the ORF when the alternative sequence is either included or excluded, respectively; and (iv) those overlapping untranslated regions (or UTRs). Then for each transition, we could calculate the proportion of events that change to a frame preserving isoform or a frame disruptive isoform (i.e. either removing or introducing a premature termination codon [PTC]). We found that the vast majority of non-frame preserving AS events (categories ii and iii) disrupt the ORF at the genome activation stage (ZGA) in the three species (Figure R.2.1.a); that is, they were more included at ZGA when the inclusion of the alternative sequence was predicted to disrupt the ORF, and vice versa for exclusion, meaning that many isoforms appear disrupted at genome activation in particular. This pattern is reversed by the blastocyst stage, where non-frame preserving isoforms are replaced with frame-preserving ones (Figure R.2.1.a).

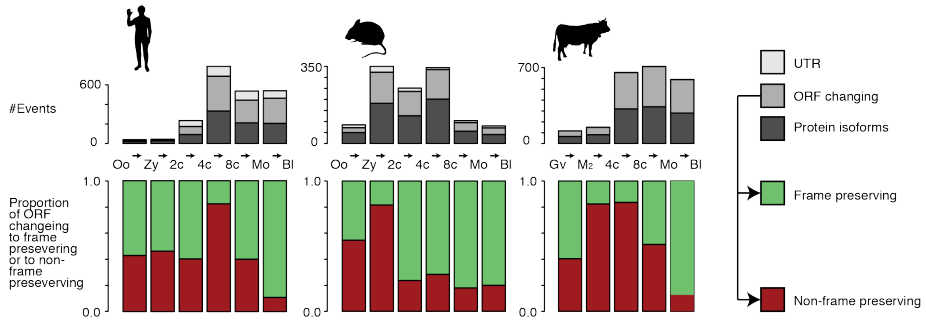


Figure R.2.1.a Effect on the ORF. **TOP**: the number of events that can form alternative protein isoforms preserving the ORF (normally divisible by 3; dark grey), the number that can form disruptive isoforms (ORF changing; light grey) and the number in the untranslated region (UTR; white) at each transition, **BOTTOM**: Of those that are ORF changing (light grey) at each transition, the proportion of events changing to an in-frame version (green) and those to a disruptive version (red; or out of frame; often not divisible by 3) are shown.

Next, we measured the changes within the different Mfuzz clusters (Supp. Figure **1.2.ii.**), looking only at those that changed at genome activation stage (see methods: p160). Peak-like exons usually disrupted the ORF at the peaking stage, whereas shift-like exons more often generated alternative protein isoforms (Figure R.2.1.b, barplots). In addition, peak-down events were usually constitutively included in differentiated tissues (constitutive-like: Human: 91.39%, Mouse: 92.48, Cow: 96.66; Figure R.2.1.b, Box-whisker plots), whereas peak-up events were normally not included in differentiated tissues (cryptic-like, Human: 11.72%, Mouse: 15.96, Cow: 9.06), for both protein-coding and ORF-changing events. A full table (splitting by type) is found in Supp. Table 2.1.a. The change of protein-coding isoforms to a non-canonical form may suggest that even events not introducing a premature stop codon could also be disruptive, as minor (rare) isoforms have been considered dysfunctional (Tress, Abascal, and Valencia 2017). Moreover, some AS events not predicted to disrupt the ORF could form non-functional proteins by the skipping of important domains (Supp. Table 2.1.b).

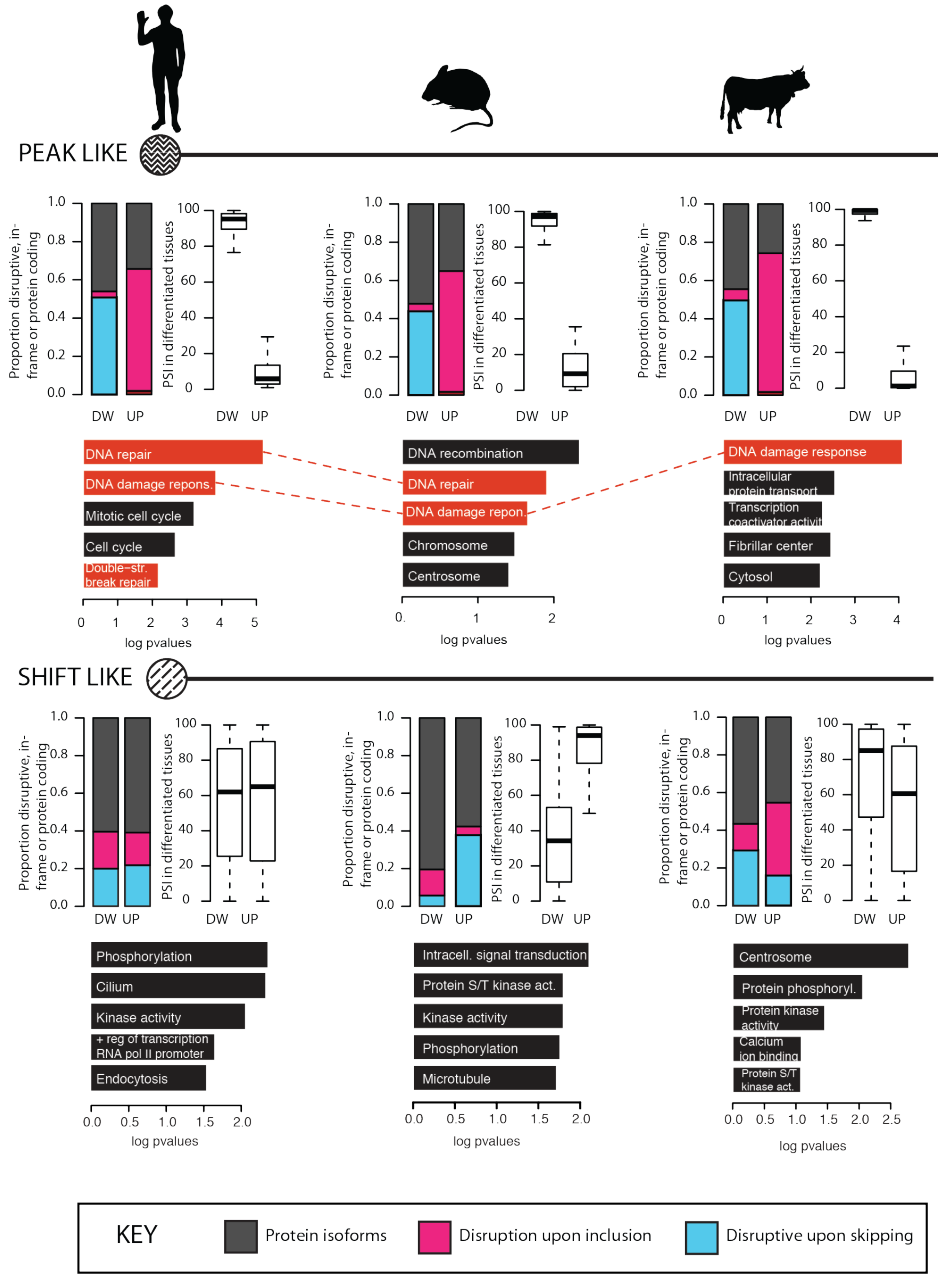


Figure R.2.1.b. ZGA peaks and shifts: ORF prediction, PSI in differentiated tissues and gene ontology. Coloured barplot shows the proportion of exons predicted to be protein isoforms (grey), those disruptive upon inclusion (pink) or exclusion (blue). Box-whiskers show the percent spliced in value for same set of exons, in differentiated tissues. For each, log binomial p.values for the top 5 gene ontology results are given for each category, combining up/down and ORF prediction. DNA damage has been highlighted in orange.

Interestingly, peak-like events were enriched for genes involved in DNA repair and cellular response to DNA-damage in human, mouse and cow (Figure R.2.1.b; [orange]). This contrasts sharply with the gene ontology enrichments for genes with shift-like events. Furthermore, other enriched GO terms may also potentially point to DNA damage related phenotypes. For instance, the centrosome recruits the DNA-repair machinery (Ratsima et al. 2016), cell-cycle is strongly regulated by DNA damage checkpoints (Dasika et al. 1999) and microtubules have been associated with tethering chromosomes during double-stranded DNA break repair (Lawrimore et al. 2017) and DNA-repair protein transport (Poruchynsky et al. 2015). There are also cases where the skipped isoform is in an important DNA damage domain or related protein (e.g. the *DEAD*-helicase in *[DDX52],* caspase domain in CASP3 (Xinjian Liu et al. 2015) and the SNF2 domain in RAD54L (Tanaka et al. 2000), in human), and thus could potentially affect their DNA repair functions. Overall, events were not preferentially in any of the six previously defined DNA damage response categories (Beerman et al. 2014) (Supp. Figure 2.1).

2.2 Disrupted isoforms do not lead to NMD

Given many genes are disrupted at genome activation, this could lead to non-sense mediated decay (NMD), as premature stop codons (PSCs) are introduced, and in turn, this may lead to targeted decay of the mRNA.

We first checked that the NMD genes were expressed in the three species, during preimplantation stages (Supp. Figure 2.2a). This showed that most genes are expressed, with some showing a increase in expression at ZGA stage. However, without knowing protein levels, which may be maternally supplied, it is difficult to determine if all the machinery is present in the early embryo.

We then tested whether ORF changing exon skipping or exon inclusion, leads to gene expression changes. Given PTCs could lead to a reduced total mRNA level (through NMD). However, in human we found no enrichment for disruptive events to be in up or down regulated genes, rather significant depletion (Figure R.2.2). Therefore, exons predicted to disrupt the ORF do not seem to lead to down-regulation of their respective genes. This may suggest that NMD is not particularly active in early

development, that it cannot degrade these specific transcripts or that the NMD machinery is overwhelmed at genome activation (because of this massive burst of expression) and has some delay clearing disruptive transcripts. Given Upf3a is embryonic lethal by mouse embryonic day 4.5-8.5 (Shum et al. 2016) and Smg1 by day 8.0 (McIlwain et al. 2010), this could suggest that NMD is needed to clear these transcripts by the blastocyst stage, although the NMD genes are also involved in other functions, such as telomere lengthening. Interestingly, (McIlwain et al. 2010) noted that Smg1-depleted embryos led to a accumulation of PTC containing splice variants, in genes not normally associated with NMD and an enrichment splicing events in DNA repair genes, cell death and intra-cellular signalling, suggesting there could be some overlap related to PTC-containing isoforms in our data.

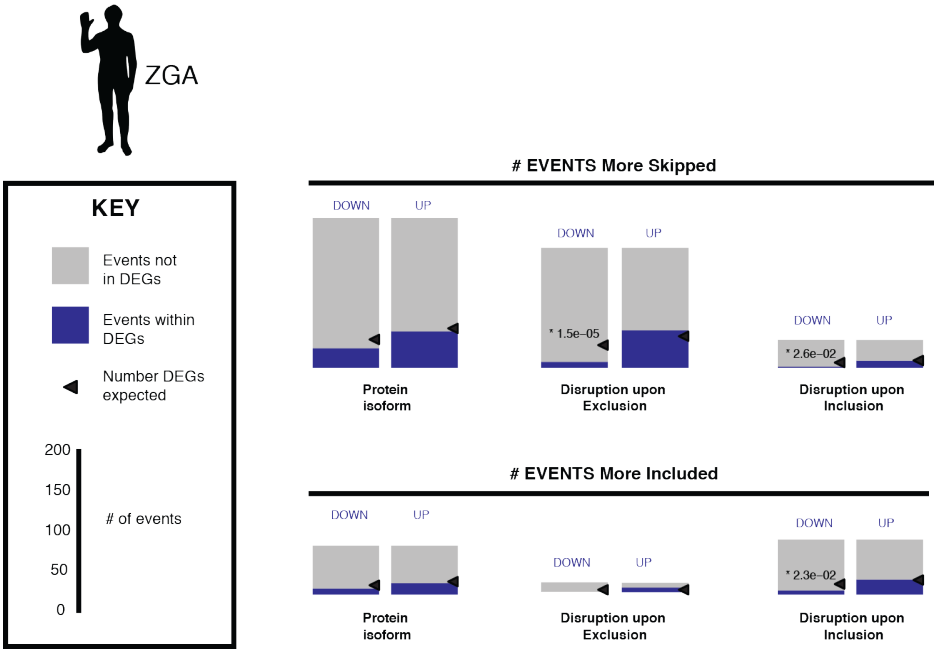


Figure R.2.2. ORF fate predictors and levels of differentially expressed genes (UP or DOWN separately). Showing zygotic genome activation stage (human: 4-Cell to 8-Cell). Bars show numbers of exons becoming more included or more skipped, split into the three predicted ORF types. Blue bars show those events within differentially expressed genes and grey for those in unchanging genes, for up- and down- regulated DEGs separately. Expected number of events in each category within a DEG is denoted by a triangle. Chi-squared p.values are denoted for significant differences from expected. Mouse and Cow are shown in Supp. Figure 2.2b.

2.3 Disrupting isoforms less bound by polysomes.

This disruptive impact on ORFs is likely to be further strengthened by a global differential engagement of isoforms in translating ribosomes. Comparison to "TRiP-Seq" data: RNA-Seq from high and low polysome (cytosolic) fractions, from human ESCs (Blair et al. 2017) showed a strong bias for ORF-changing isoforms (Figure R.2.3). Changes at the 4 to 8 cell stage, showed that disrupted isoforms were less engaged by translating ribosomes (Figure R.2.3), whereas the opposite was true for in-frame isoforms (RESC). The two are significantly different (5.3e-23). No significant bias was observed for alternative exons generating alternative protein isoforms (PROT) or in UTRs. Furthermore, isoforms at the peak stage were found to be strongly depleted for the polysome fraction, irrespective of the predicted impact, indicating that even non-frame shifting events could be less translated (Supp. Figure 2.3).

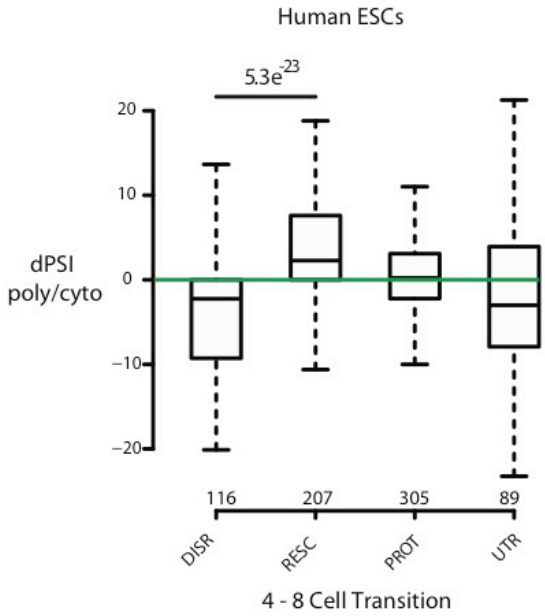


Figure R.2.3. Disrupted isoforms are less bound by polysomes. For human events at genome activation (4-8Cell stage), split into those that change to a more disruptive isoform (DISR), to a more in-frame isoform (RESC), to another protein isoform (PROT; not changing tri-nucleotide order) or in the UTR, the change in PSI between the polysome and cytosolic compartments in embryonic stem cells (data from (Blair et al. 2017)), with a positive delta PSI indicating the more polysome fraction. A significant difference is found between the disrupted (DIST) and rescued (RESC), fishers test.

In summary, these results show that a large number of AS events disrupt the ORF or change to a minor isoform specifically around genome activation stage, and often in peak-like clusters (with temporary disruption). Although, this seems to have no effect on total transcript levels through NMD, we find there may be less protein produced as predicted from polyribosome sequencing (TRiP-Seq). Premature termination codons cause release factors (ETF1 in human) to act, allowing the premature polypeptide chain to be released, and for the mRNA and ribosome complex to be disassociated (Cooper and Hausman 2007). We suggest that it is this phenomenon that we are detecting, that the ribosome is being separated from the stop-containing mRNA. This in turn, should lead to lower protein levels or to truncated proteins isoforms at ZGA. Curiously, we find that DNA damage genes are greatly impacted.

Chapter 3. Regulation of splicing

Having outlined the splicing dynamics of alternative exons throughout pre-implantation development, we next sought to find potential regulators. To address this question we used multiple strategies: **3.1**: Comparison to known-RNA binding motif and kmer enrichment, **3.2**: de novo motif enrichment, **3.3**: single-cell gene-expression correlation analysis and **3.4**: RBP knockdown correlation analysis. Given the diversity of splicing changes, we presume there will be multiple factors that affect exon inclusion changes during oocyte to blastocyst. For this thesis, we decided to focus on the set of exon-skipping events at ZGA in human, mouse and cow. This represent the most extreme phenotype we wanted to explore, yet the supplementary tables and figures document all the transitions (linked in each section).

3.1 CIS-regulatory and kmer motif enrichment

Using both 'Mfuzz clusters' (Supp. Figure 1.2.ii.), and cell stage transitions (e.g. 2-Cell to 4-Cell), genomic sequences could be compiled from the exon body, upstream intronic region and downstream intronic region. We could search these regions for enrichment of known RNA binding protein motifs, downloaded from the CISBP-RNA database (Ray et al. 2013). Motif enrichment can indicate if regulators bind these regions, and potentially induce skipping or inclusion. These analyses were conducted using 'matt' (Gohr and Irimia 2019), which scans the region around alternative exons, comparing the exons of interest (i.e. clusters/transitions), with a background list of events. Full results in Supplementary Data 4.

This strategy finds multiple RNA binding proteins enriched. Often the same RNA binding proteins were enriched/depleted in comparisons of exons becoming more skipped or more included at each transition. Specifically at genome activation, ELAV(1/2) (Figure R.3.1.a.), ORB2, ZC3H14, PTBP1, ZFP36 show very similar patterns and are in the top 10 most enriched for the three species (Full plots in Supp. Data 4). Noticeably, mouse and cow share a similar pattern but human has the opposite for the same motifs. In mouse, exons that become more included are enriched, whereas for human it is the ones that become skipped (at ZGA). At the cluster level, a similar enrichment is apparent (Supp. Data 4).

Despite some interesting candidates, no factor was enriched more than 9% between the targets sequences and the background, however the CISBP-RNA database does not represent all splicing regulators, so this analysis is inconclusive.

Given the RBP motif may be missing, we also checked for enrichment of all possible 2, 3 and 4 mers (k-mers; random AGCT sequences of length 2, 3 or 4), for each transition and cluster. In terms of kmers, we found a strong enrichment for poly-U sequences in the upstream intron for the exon-skipping events at genome activation in human (Supp. Figure 3.1.iii.). Conversely, mouse ZGA skipping events have a depletion of poly-U sequences in the same region and cow shows no difference. Despite the differences between the species, it is notably the upstream intron around the poly-pyrimidine tract is quite different between exons that are skipped and not (at ZGA in the three species). This could mean it is a protein or proteins that bind to this region that are responsible for the skipping, specifically a protein in the U2/U5/U6 spliceosomal complexes that bind here (Figure I.2.iii).

Elav (M116_0.6)
Predicted from Drosophila.

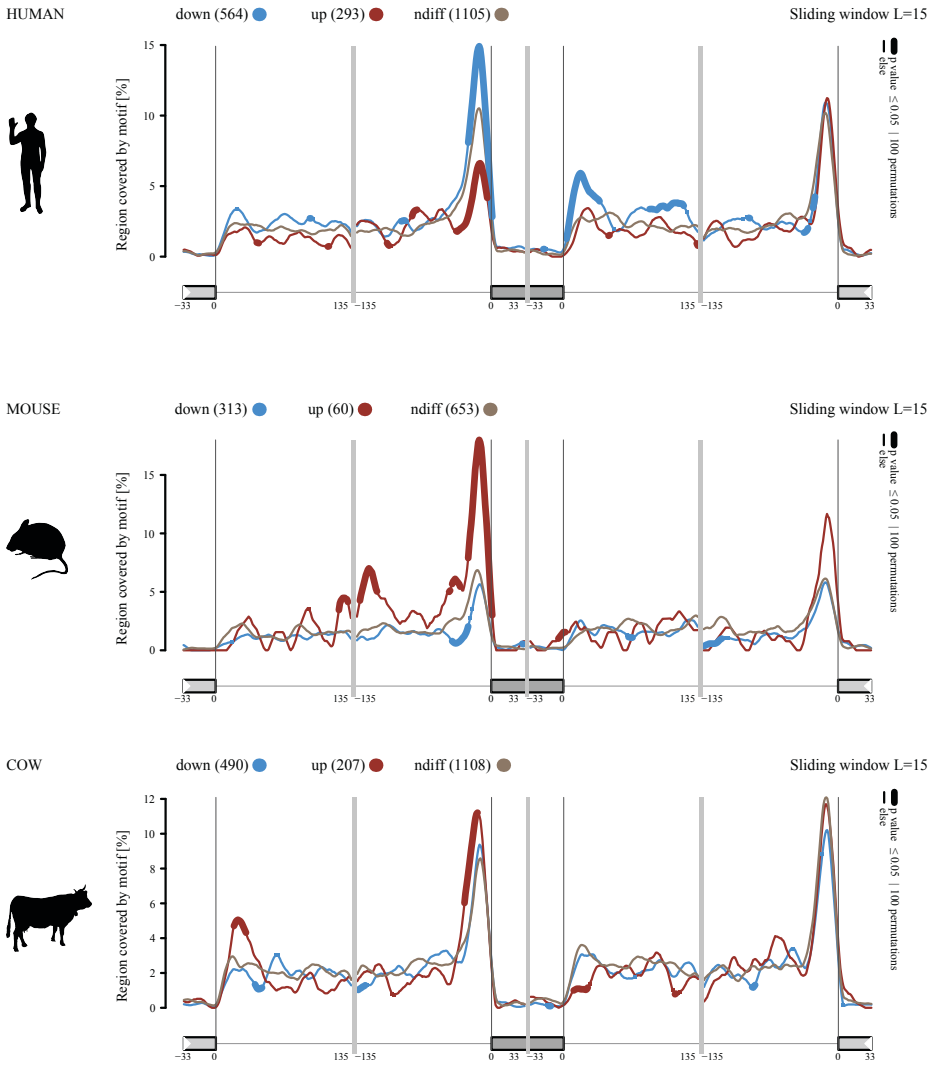


Figure R.3.1. RNA maps (Gohr and Irimia 2019) density plot for ELAV in HMC, showing the sets of exons that increase in inclusion (red), decrease in inclusion (blue) and not change (grey) at genome activation (Zygote-2Cell in mouse, 4-8Cell in human and cow). Top shows the motif, in this case predicted from Drosophila, and the weighted score for each position for AUGC nucleotides.

3.2 Novel motif and transposon enrichment.

We next sought to find novel motifs, both long and short sequences around the alternative exons, using the programs MEME (Bailey and Elkan 1994) and DREME (Bailey 2011) respectively.

Using MEME, we found many motifs enriched in specific clusters and at the transition of zygotic genome activation, for the three species (Supplementary Table 3, Supp Data 6). These sequences however did not appear to belong to RNA binding motifs, as many were extremely long (>40bps). We expect most RBP binding sites to be about 5-12 base pairs long (Dominguez et al. 2018).

Examination of the sequences found that in human, many belonged to Alu repetitive transposable elements (often Alu S family) (Figure R.3.2a). In mouse and cow no noticeable transposon group appeared, except for some repetitive (T)n and (CA)n (cytosine/adenosine) repeats regions, with enrichment in both up and down-stream introns. When masking the genome of repetitive and transposable elements we predictably lose the high enrichment in human, but continue to find the AT and CG rich motifs in human and mouse (Supp. Table 3, tab 4-6).

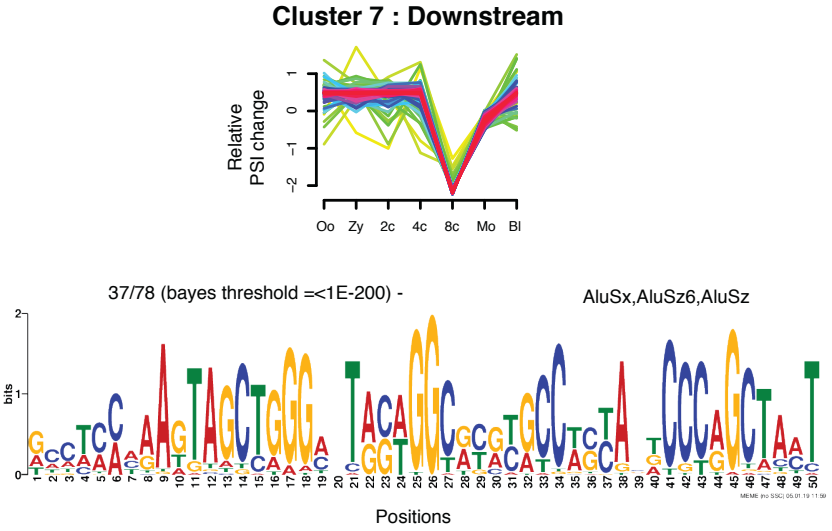


Figure R.3.2.a. Human example, cluster 7, with enrichment in downstream intron of 37/78 exons. This motif is contained within Alu sequences (AluSx,Aluz6,AluSz). The guanosine (G)/ thymine (T)/ adenosine (A) at positions 10-12 are distinctive of right hand isoforms of Alu in sense orientation.

This led us to wonder whether the presence of transposable element sequences in the introns or within the exon of alternative events themselves, may influence exon inclusion. Specific families of transposable elements (or transposons) are transiently activated at ZGA in both mouse and human (Ge 2017), with SINES (short interspersed Nuclear Elements) in the promoters of genes having been linked to higher expression at genome activation in many species (Ge 2017). Moreover, expression of LINE elements after fertilization is important for mouse development, in connection with proper chromatin regulation (Jachowicz et al 2017).

To explore whether transposons are linked to splicing in the early embryo, we calculated the number of transposons in each AS Mfuzz cluster, with known transposon positions from "RepeatMasker" (Smit, Hubley, and Grenn 2013). We found multiple clusters were significantly enriched or depleted for exons overlapping transposons in each species. The enrichment/depletion did not have a clear bias towards any transposon family. Remarkably, there was a very strong association between enriched/depleted clusters and peak-like behaviour in all three species. Exons with down-peaks showed depletion for overlapping transposons, whereas up-peaks showed enrichment when measuring within the exon (Figure R.3.2.b). These patterns suggest that most exons that are temporarily skipped (down-peaks) are constitutive exons, and thus of coding relevance, being void of transposons, and exons with increased peak PSI are largely cryptic, barely encoding functional domains. When looking at the upstream and downstream introns of peak down clusters, they tended to have higher levels of transposons in contrast to the exon itself (Supp. Figure 3.2.iii), which is consistent with the novel exon discovery of Alu elements up and down-stream of these clusters (Figure R.3.2.a). Furthermore, we found a strong association between transposons in exons that disrupt ORF upon inclusion and upon exclusion, within each cluster (Supp. Figure 3.2.iv.). Overall, these data highlight a (likely indirect) association among transposon contribution, propensity to ORF disruption and peak-like regulation of alternative exons during early mammalian development.

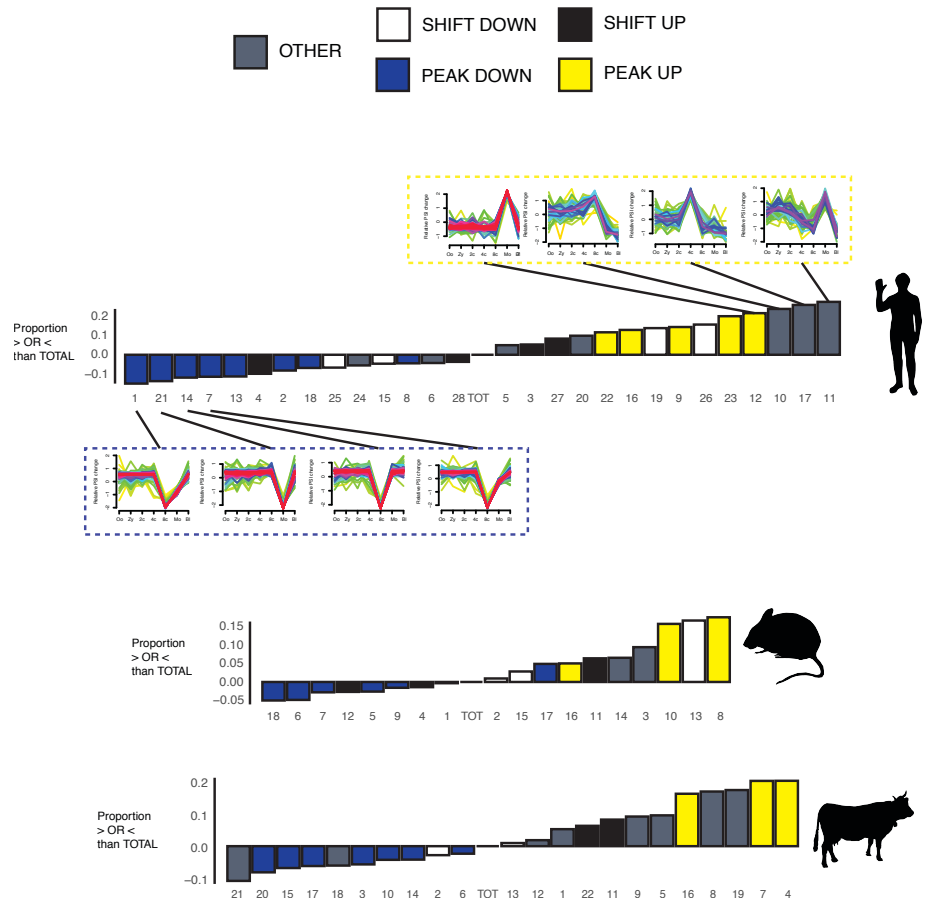


Figure R3.2.b Enrichment of transposable elements within the exon for each Mfuzz cluster of exon skipping events. Showing the proportion of transposons more or less than the background, containing all the events in clusters. (x-axis) Bars are coloured depending on their most common cluster type, either peak up or down (yellow, blue) or shift up or down (black, white) and other (grey). For human, Mfuzz cluster examples are shown, with relative PSI changes across the time course.

3.3 Single cell RBP/PSI correlations

We next correlated the change in expression of RBPs across the time course and the Mfuzz clusters showing the dynamics of exons, in the hope to be able to link changes in expression of a regulator to distinct patterns of inclusion. This found potential candidate RBPs for each Mfuzz cluster (Full result in Supp. Data 9). For example, Figure R.3.3.a show two peak-down like AS clusters (cluster 14 in human and 18 in

mouse), that have positive and negative correlation to several RNA binding protein expression profiles. Interestingly, several RBPs had positive/negative correlation in the three species, SFA3A/Sfa3a was found positively correlated in both human and mouse. This provided a useful list of potential regulators for each cluster type, based solely on correlations.

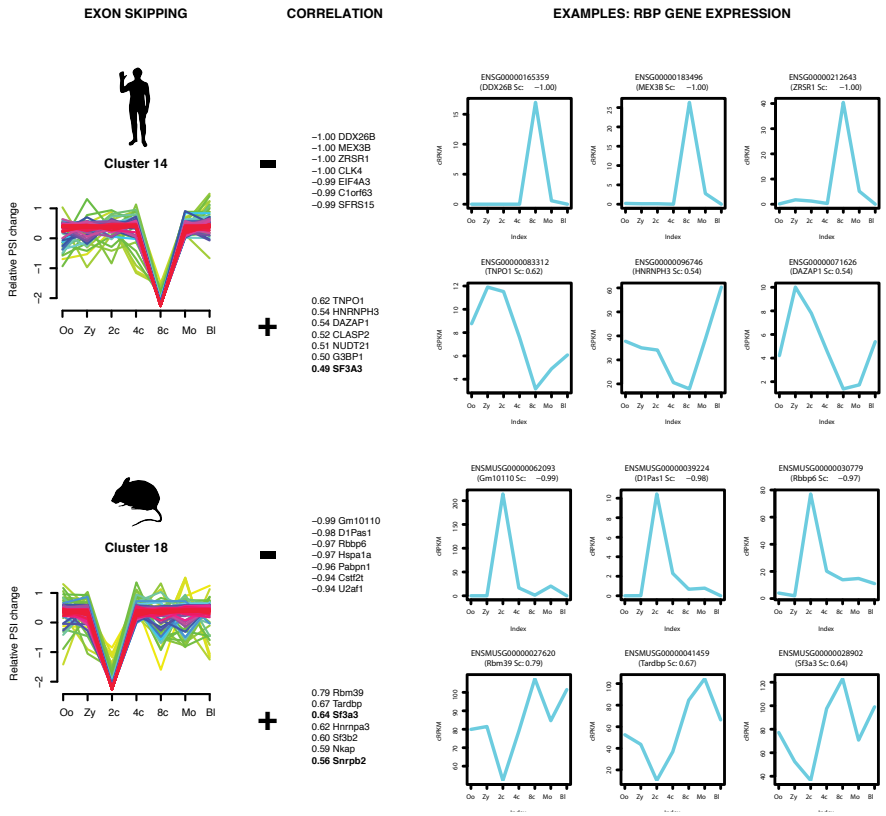


Figure R.3.3.a Example of peak down cluster at genome activation in human and mouse, and the top positive and negative correlating RNA binding proteins (using expression in cRPKM) and their dynamics to right.

In a similar way, we could also correlate the average PSI values of exons becoming more skipped or more included at each transition with RNA binding protein gene expression at the single cell level (for human and mouse only), within each stage separately (to avoid time-dependent correlation). This identified multiple significant correlations >0.8 or <-0.8, with some well-known splicing factors, including SNRPB, SRSF2 and SF3A3 (Figure R3.3.b; Supplementary Table 3.3 a and b). Showing that

the more exon skipping at genome activation is negatively correlated with expression levels of the three RBPs.

In mouse, the top candidates at genome activation were Sf3b1, Rbm28, Srrm1 (positively correlated) and Bcas2, Naa38 and Lsmd1 (negatively correlated).

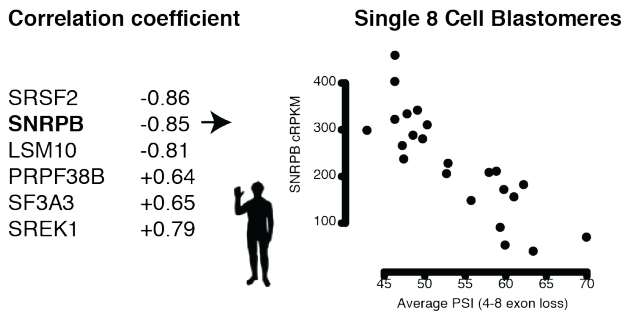


Figure R.3.3.b. Single cell correlations at genome activation stage (human). LEFT: Top three correlations (+ and -) of RBP expression and level of skipping at 4 to 8 cell stage, using single cells. RIGHT: Correlation of SNRPB at the 8-cell stage vs skipping level of alt exons (4 to 8 cell stage).

3.4 RBP knockdown/overexpression correlations

Using publically available RNA-Seq data from 88 human and 45 mouse knock- down or overexpression datasets in human and mouse RBPs in different cell or tissue types (Supp. Table 1- RBP_knock_downs [Hsa and Mmu]), we could calculate for each RBP, the average change in PSI (or delta [dPSI]) between knockdown and control for all events with sufficient read coverage, and compare these with the dPSI of all AS events changing at any given transition (Supp. Fig. Item 3Re.1 and Supp. Table 4 [Hsa and Mmu]). Given there are at least around 200 known RBPs/associated factors (Supp. Table 3.4), we have only checked around a quarter (or less) of the potential regulators. The most significant factor at any transition was SNRPB at ZGA in human (Figure R.3.4.a; $p=3.6 \times 10^{-41}$, binomial test of quadrants Q2 and Q3 vs. Q1 and Q4).

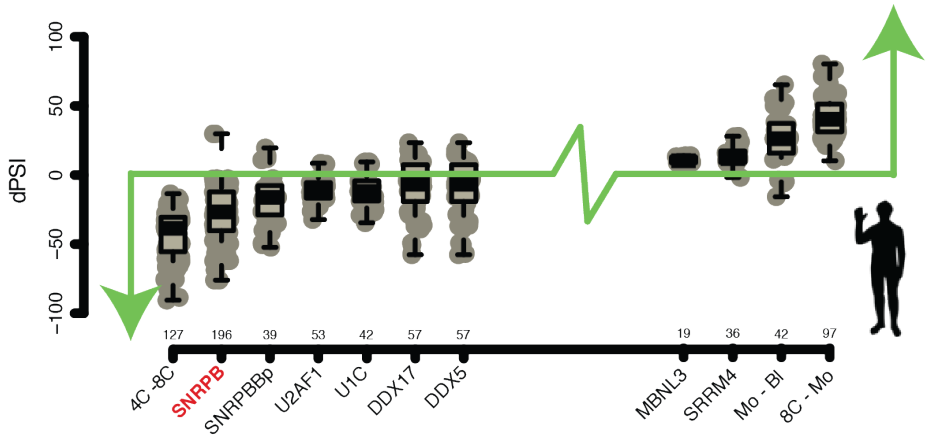


Figure R.3.4.a. RBP knock-down correlation highlights SNRPB. Exon skipping events skipped at 4-8C stage in human and the change in PSI (dPSI) between control and specific RBP knockdowns. Numbers of exons with coverage are listed.

Given the strong association between SNRPB and exon skipping changes at ZGA, we decided to study the potential role of this factor in more detail. SNRPB is part of the Sm heptameric ring, which is required for the biogenesis of the U1, U2, U4/U6, and U5 snRNP molecules on the pre-mRNA, which means it comes into close contact with the upstream poly-pyrimidine tract, the area with poly-U enrichment in human (Results 3.1. –Kmer enrichment).

Unexpectedly in human, SNRPB is lowly expressed before ZGA, and then increase sharply at the 8-Cell stage (Figure R.3.4.a. BOTTOM). This is the exact opposite of what could have been expected, as SNRPB knockdown showed a skipping response, but here we see skipping at the point where SNRPB goes from low to high expression. Remarkably, a similar expression profile was found for Snrpb in mouse and SNRPB in cow (Figure R.3.4.b). In addition, three other genes encoding Sm ring proteins, SNRPA1, SNRPC and SNRPD2, had similar activation patterns at ZGA in the three species (Supp. Fig 3.4).

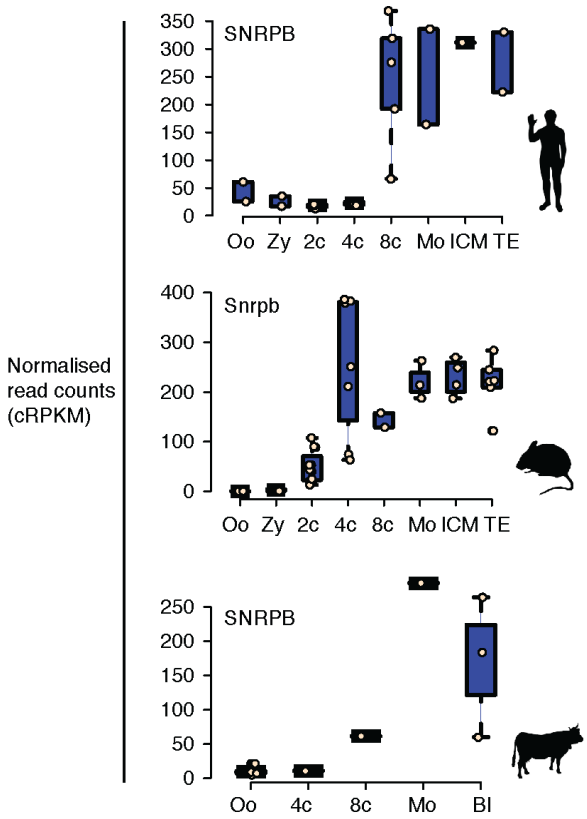


Figure 3.4.b: Expression of Snrpb in human, mouse and cow across developmental stages. Oo:Oocyte, Zy:Zygote, Mo:Morula, ICM: Inner cell mass, TE: Trophectoderm, Bl: Blastocyst. Each point is a separate merged replicate.

3.5 Knockout phenotypes

Finally, we also checked which RNA binding proteins have relevant knockout phenotypes in mouse using the MGI database (Mouse Genome Informatics Webserver; The Jackson laboratory). This found 8 RBPs that were embryonic lethal prior to implantation, including Sf3b1, Sf3a1, Srsf3, Snrnp200, Hnrnpl, Prpf19, Rnpc3 and Rbm19 (Supp. Table E1). No information was available for Snrpb/d2. Further, checking splicing related protein, that are not RNA binding per se, we also find Xab2 is pre-implantation lethal (fails at morula compaction) (Yonemasu et al. 2005) and Plrg1, 1.5days post fertilisation (Kleinridders et al. 2009).

3.6 Conclusion

In conclusion, through multiple RBP motif/correlation/search techniques, we have found several potential regulators, many of them were found to be core spliceosomal

factors (Figure R.6). The strongest correlative evidence for regulation of the skipped ZGA exons was from the RBP-knock down correlations, which highlighted SNPRB. However, the gene expression in the early embryo follows the opposite pattern to what we might expect, given expression increases in the embryo. However, looking at the gene expression alone, we cannot directly link these potential regulators to the splicing changes. Considering all these data together, we hypothesized the following scenario. Genes that are actively transcribed at ZGA, when the levels of these specific Sm (Snrpb/d2 or others) are still low, results in the production of transcripts with skipping exons. As development proceeds, the burst of expression of these SNRP genes will elevate protein levels (with some delay), which will in turn, leads to inclusion of those exons in the following stages, thus resulting in a peak-like pattern of exon skipping at ZGA stages.

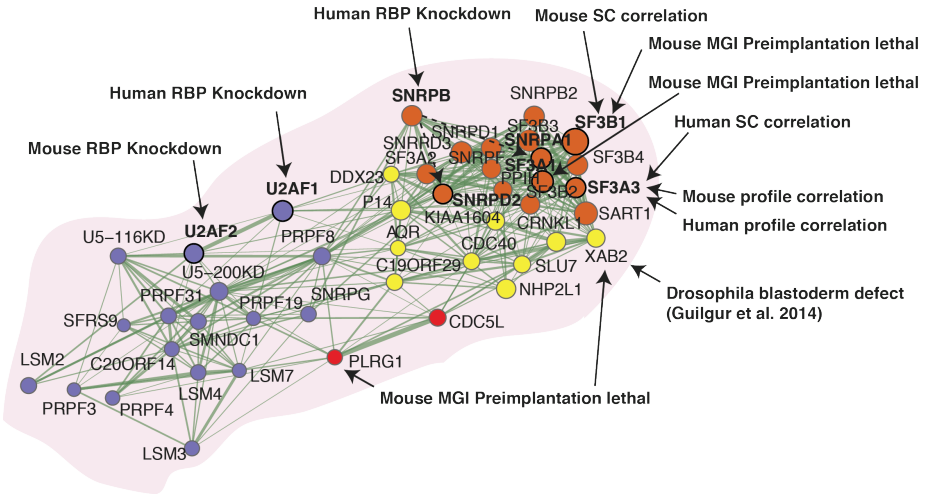


Figure R.3.6. Network of core splicing components adapted from (Papasaikas et al. 2015; Figure 4). Showing part of a network with core splicing proteins that have a positive functional correlation as predicted by knockdown of each given factor. In bold are the genes that have arisen from the various analyses carried out in mouse and human, either from the RBP knockdowns, SC-single cell correlation analysis, RBP expression/cluster analysis or MGI preimplantation knockdown lethal (mouse-only). Dotted lines represent inferred potential regulators, due to expression changes similar to Snrpb. Colour of nodes represent different sub-networks from (Papasaikas et al. 2015).

Chapter 4. SNRPs and DNA repair

Thus far, we have determined that a set of exons are skipped around ZGA (many leading to disruption in the three species). These are potentially regulated by Snrpb/d2 (as predicted from human SNRPB knockdown only) and are in genes enriched in DNA repair pathways (all three species). It has also been shown that DNA damage response, as measured by gamma-H2AX phosphorylation is also inactivated from the 1cell to 2-cell stage, becoming active around the 4-8cell stage in mouse (Yukawa et al. 2007), in line with our predicted exon-skipping phenotype (disruption in DNA damage genes).

To fully investigate the connection between Snrpb, exon skipping and DNA damage. We: **4.1**) Estimated the number of events potentially regulated by Snrpb in mouse (by orthology inference to human events); **4.2**) Determined mRNA and protein abundance of Snrpb and Snrpd2, from oocyte to blastocyst (mouse); **4.3**) Overexpressed Snrpb and Snrpd2 at the zygote stage in mouse, to assess changes to exon skipping; **4.4**) Co-injected Snrpb/d2 with etoposide treatment, to evaluate the affect on the DNA repair pathway; **4.5**) Analysed the expressed levels of DNA repair genes across development, and explored a potential splicing related DNA damage response in the embryo.

In the following section, Barbara Pernaute (Irimia Lab) designed and conducted the experimental work in mouse (with the help of Marta Miret Cuesta). Bioinformatic analyses were completed by myself (sections 4.1, 4.3, 4.4).

4.1 Snrpb/d1 may regulate exon skipping in mouse

Given the difficulty of working with human/bovine embryos (not established in the lab and ethical considerations), we wanted to test our hypothesis that Snrpb may regulate a subset of exons around genome activation in mouse. For this we took the 1433 events that were skipped at genome activation in human and also skipped after SNRPB knockdown (dPSI 15). 736/1433 had an orthologous mouse events in vast-tools. Of these, 67 were found alternative in mouse (any pairwise transition: oocyte to blastocyst) and a further 101 events had a standard deviation >=5 (median of each stage-oocyte to blastocyst). Of these, 91/168 (54.2%) were in peak-down-like clusters (Supp. Table 4.1). Of the 67 AS events in mouse, 48 (72%) were in peak-down Mfuzz

clusters around genome activation (Figure R.4.1). This gave us an indication that Snrpb may indeed affect a subset of the same exons as in human, leading to skipping around genome activation in mouse.

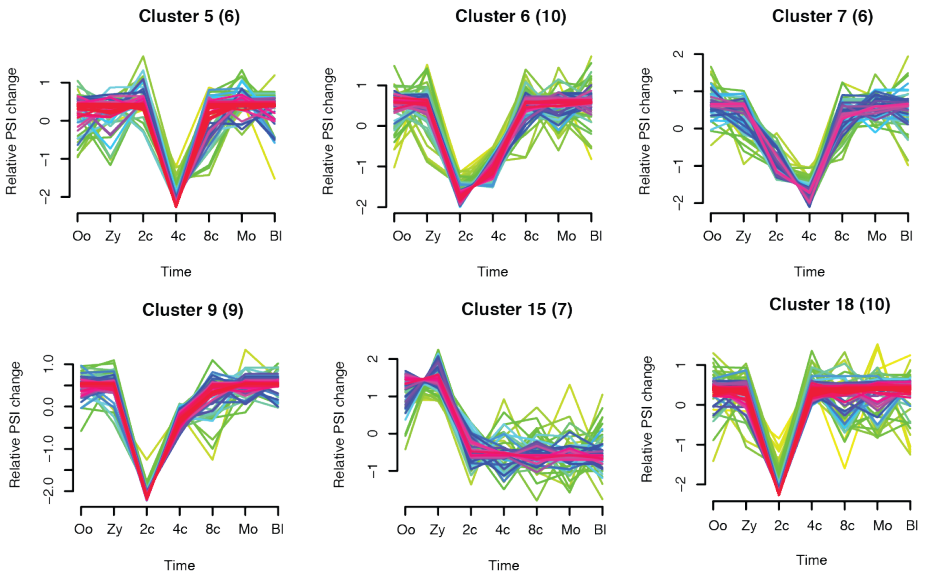


Figure R.4.1. SNRPB predicted events in mouse. The top 6 mouse clusters with SNRPB orthologous events. In brackets are the numbers of events in each cluster. For each plot the x-axis shows oocyte:Oo, zygote:Zy, 2c-8c:2-cell to 8-cell, Mo:Morula, Bl:Blastocyst; y-axis shows relative change in PSI (percent spliced in).

4.2 Snrpb/d2 expression and protein levels

Our RNA-Seq data show that Snrpb mRNA expression greatly increases at genome activation, which we could confirm through RT-PCR (Supp. Figure 4.2.a). However protein levels may not correspond to levels of mRNA. In addition, the oocyte can be loading with maternally supplied proteins, so it is important to check protein levels in the embryo. For this, we performed a western blot of Snrpb, which shows there are lower protein levels from 1- to Morula stage (Figure 4.1), further using immunostaining of Snrpb/d2, we could show that levels of the protein increase from the 4 to 8-cell stage in mouse. The immunostaining fits with our hypothesis. However, further experiments may be needed to accurately measure relative protein levels.

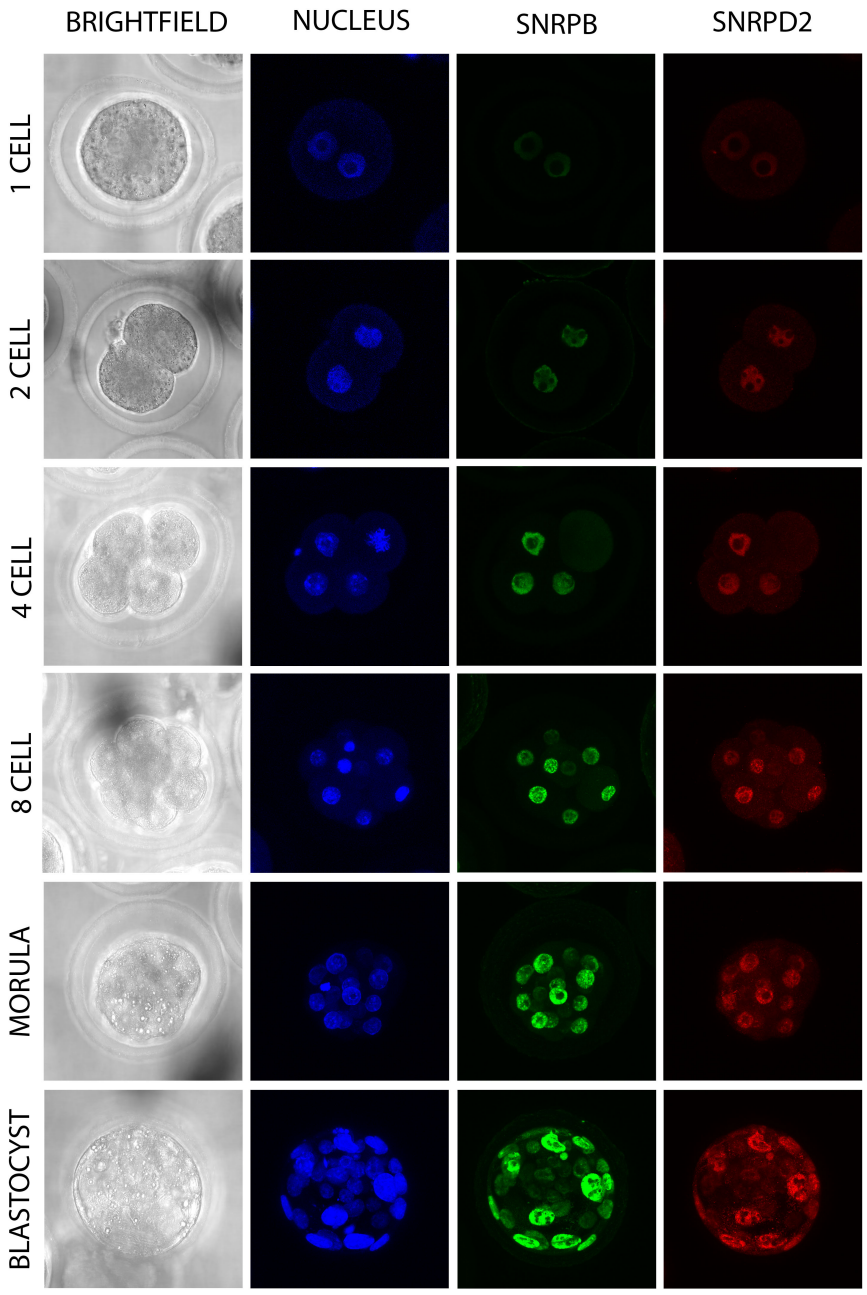


Figure 4.1. Immunostaining of Snrpb/d2 in mouse pre-implantation embryos. Showing increased protein levels by the 4 cell stage.

4.3 Microinjection of SNRPB can prevent exon skipping.

With Snrpb levels lower during the genome activation stages, we next wanted to test whether Snprb is responsible for the exon-skipping pattern in mouse. Through microinjection of Snrpb mRNA into the zygote, we could increase expression at an earlier stage in the embryo (Supp. Figure 4.3).

Next, we then performed an RNA-Seq experiment, from control-injected versus Snrpb, Snrpd2 and Snrpb/d2-injected zygotes (see methods p.166). This experiment was particularly hard, as we needed to capture the embryos before and directly after genome activation (at the point of greatest skipping). Plotting the results in a PCA (Supp. figure R.4.3), you can see that our injected and control 2-cell samples are not fully activated and cluster closer with zygote-stage embryos (from published data). However, we were able to measure some changes in splicing between the control versus the micro-injected. Of 362 events that change from zygote to 2-cell stage in mouse as predicted from public RNA-Seq, 95 events had a dPSI >=15 in our control Zygote and "2-cell" stages (Figure R.5.3 [left]). Of these 95 events we could compare the dPSI to that of our overexpression "2-cells" (Figure R.5.3 [right]). This clearly shows that many of the events skipped in both the public/our own data are not skipped in the overexpression, and further some skipped event have much higher inclusion after overexpression.

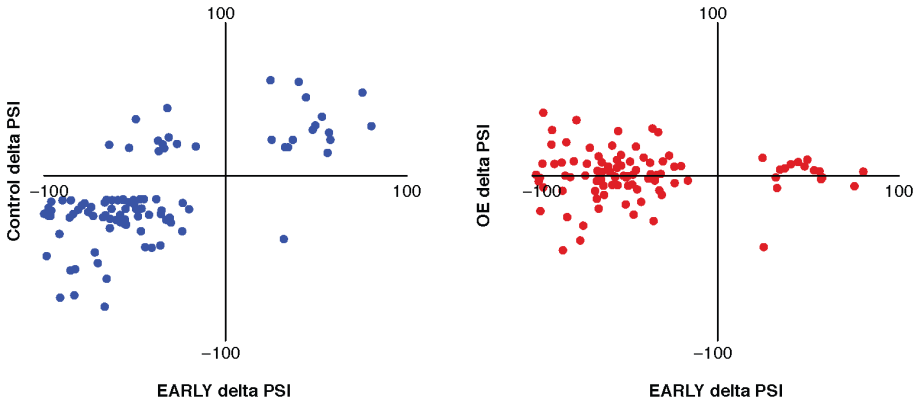


Figure R.4.3. Exon skipping reduction in Snrpbd2 microinjected embryos. LEFT: comparison of dPSI values for event changing in both public and our own control datasets from zygote to 2-cell. RIGHT: Comparison of dPSI values of control vs Snrpbd2 overexpression 2-cell embryos, and dPSI values of public.

4.4 DNA damage levels affected by SNRPB/D2 levels.

Finally, given we could in part prevent the exon skipping of these exons in early development by combined overexpression of of Snrpb and Snrpd2, we wanted to know what effect this has on the DNA damage repair activities of the embryo. To explore this, we used a combination of Snrpb/d2 injections and etoposide treatment to measure changes in response to DNA damage by gH2AX and phospho-p53 (Ser15) immunostaining. For this 1-Cell embryos were microinjected with Snrpb and Snrpd2 mRNA or mCherry mRNA as control, and treated at 2-Cell stage with 10uM Etoposide for 1.5h or left untreated (See methods p.166). A preliminary experiment showed that Snrpb and Snrpd2 overexpression did not induce a significant change on p-p53 intensity after etoposide treatment, however the levels of gH2AX were significantly higher after the treatment in Snrpb/d2 injected embryos (Figure R.4.4). This suggests that potentially Snrpb/d2 indeed does affect DNA repair in early development. Further, we know that potentially other Snrp(x) and U2af(x) genes may be involved, so Snrpb may not be the limiting factor affecting these particular exons, so this effect may be greater in the embryo, where multiple proteins in the complex could be absent.

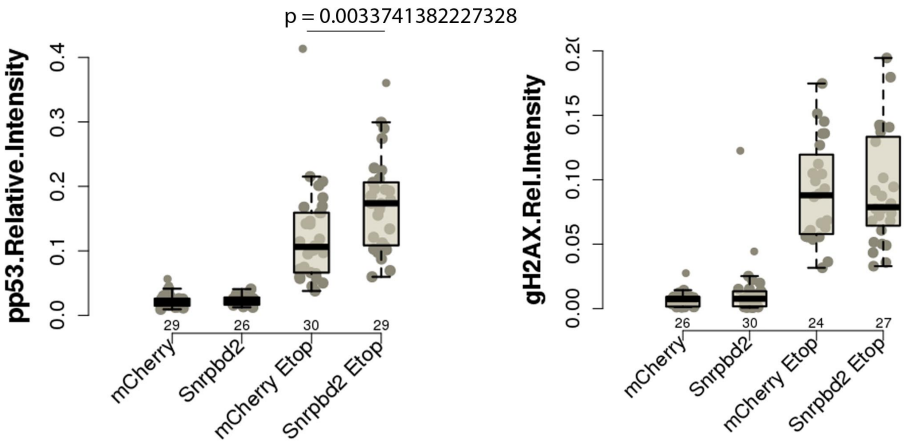


Figure R.4.4. Experiments showing etoposide treatment in 2-cell embryos after Snrpb/d2 overexpression. TOP: Relative intensity for pp53 and BOTTOM: Relative intensity for gamma-H2AX, with mCherry or Snrpb/d2 injected embryos, and treated with etoposide at the 2-Cell stage. Each dot represents relative intensity of a nucleus. This was completed doing multiple rounds of embryos, with the full result in Supp. Figure 4.4.

4.5 DNA repair gene expression

Next, given DNA repair genes were predicted to lead to disruption at ZGA, we wanted to know if these genes had any change in expression levels that may lead to dysfunction. In supplementary figure 4.5, I show the expression of gene (in cRPKM) in the three species and subdivided into the different DNA damage pathways. This shows that most genes in each pathway have expression across the time course, so are predicted to have adequate protein levels. Though a subset of DNA damage genes are not expressed (or lowly) before genome activation. Among the genes are important regulators of DNA damage repair, including: in human, RAD52, POLD4, XRCC3, POLN and XPA, in mouse, Polh, Atm, Mlh1, Kdm4d and Fancf. In addition, given NMD may not be particularly active, the disruption of DNA damage genes may not lead to lower mRNA levels, so we cannot assess their function by measuring mRNA levels.

4.6 Conclusion

In conclusion, we found that skipped exons may be the result of low levels of Snrpb/d2. Further, we could show that microinjection of Snprb/d2 could increase the DNA damage response as shown by phospho-p53 intensity. Further, ewe plan e to perform more experiments to better detect gamma-H2AX intensity, as we believe this marker too may be significantly different (not shown here). We cannot be certain that it is Snrpb/d2's splicing function leads to the changes in DNA damage response, yet it is striking that we find this noticeable change.

Chapter 5. Conservation of AS in core pathways

Next, we investigated the evolutionary conservation of AS patterns during pre-implantation development and extent of species-specific changes. This can potentially highlight specific exons with a critical function, which have been retained through evolutionary time and may revealed conserved function.

5.1 Conserved AS events

To obtain a global portrait of conserved AS events, we needed to determine the orthology of exons and calculate the number of comparable exons between any two species. To do this, we first used lift-over (see methods), which uses the synteny of gene structure to map each exon's coordinates to a second species. Next, we ensured the lifted coordinates had the correct splice site junctions (AG to GT), and then whether the lifted exon had a vast-tools IDs (see methods) and so is annotated for the second species in VASTDB (our database of exons). Finally, with the vast-tool IDs we could check if the event had sufficient expression in the second species to detect splicing (See methods; [> VLOW coverage in at least two stages]).

Using these levels of information, we could compare all events changing in any one species (Hsa=3554, Mmu=1934, Bta=3087), to the other two species as shown in Figure R.5.1.a. For example, of 3554 exon-skipping events in human, 2662 exons could be lifted to mouse (G [genomic]: exon lifts over to mouse with correct dinucleotide splice junctions). Of 2662 exons present in the genome of mouse, only 1329 had an alternative vast-tools identifiers in mouse (A: alternative, has vast-tools ID). Finally, of 1329 events comparable from human to mouse, 938 exons had sufficient read coverage in mouse to be measured for alternative splicing (Ex: expressed; see methods]). A total of 581 human, 727 cow and 530 mouse events were comparable to the other two species (Figure R.5.1.a-Venn). Of these, many were also found alternatively spliced in the other species (i.e. venn overlap). 93 were alternatively spliced in all three (during any preimplantation transition), 129 only in Human/Cow, 110 in Mouse/Cow and 79 in Human/Mouse only, which means that around 52%, 46% and 53% of comparable exon skipping events in human, mouse and cow respectively are also alternatively spliced in at least one of the other two species.

(Figure R.5.1.a: Top). Further, we found that events changing in two species, but not the third (based on our script's definition of alternative), were also changing when lowering the dPSI threshold to 15 (Figure R.5.1.a: middle). Using this approach we attained 259 putative conserved events between the three species (Table 5.1).

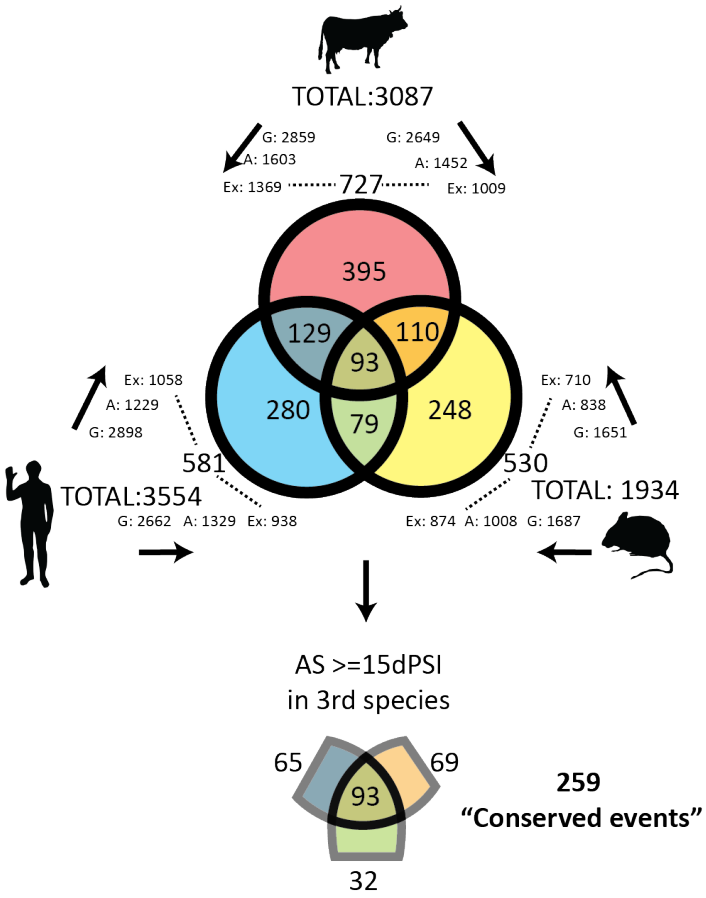


Figure R.5.1.a. Conservation of exon skipping events in human, mouse and cow. Upper: Overlap of comparable events between the three species. Showing the total number of exon skipping events, and compared to either of the other two species, G: Total number detectable at the genomic level, A: Total that have an annotated exon in vast-tools (and 'G'), Ex: Total number with expression for same event (and 'G','A'). Using the intersection of the two with expression for these events in the other two species the Venn shows the overlap of events that are also changing in the other two species. **Middle:** Number of events considered conserved between the three species, by lowering expected dPSI value to 15 in the 3rd species.

Conserved events have particular features. They are more likely to generate protein coding isoforms when included/skipped: human 69% (1.7 fold enrichment;

hypergeometric=1.59e^{-22}), mouse 70% (1.35 fold; 1.26e^{-10}), cow 66% (1.51 fold; 8.73e^{-15}) (Supp. Figure 5.1.a), compared with all exon-skipping events. They also are less likely to be peak-down clusters (Human: 2.25 fold less [2.7e^{-8}], Mouse: 2.62 fold less [1.09e^{-16}], Cow: 5.2 fold less [4.3e^{-31}]), and more likely to be shift-like clusters (Human: 1.61 fold more [1.19e^{-5}], Mouse: 1.2 fold [0.003], Cow: 2.06 fold [3.29e^{-8}]) (Supp. Figure 5.1.b), compared to all exon-skipping events.

To test if the conserved changes have a similar dynamic across time, we next compared the 365 events changing in human and mouse (includes those predicted through Cow AS [dPSI>=15], and those between Human/Mouse only). For these events we could also check the cluster type they belonged to, either peak up/down or shift up/down. Only 88/365 had these categories in both mouse and human. The number is low as many events didn't have coverage enough to determine transition type, or were classified as 'other' (unclear dynamic). This showed that the general patterns are often the same between the two species for the same exons (Figure 5.1.b.: Cluster type).

Next we calculated at which transition the first major change occurred (calculated at dPSI 25). Out of 365 exons, 227 were changing in a consecutive stage (e.g. 4-cell to 8-cell; in both species; the others may change between say 4-cell and Morula, were not considered). 145/227 were changing in the same direction at greatest change. Of these, we could show the correspondence between transitions changing in the two species (Figure R.5.1.b: Transition of greatest change).

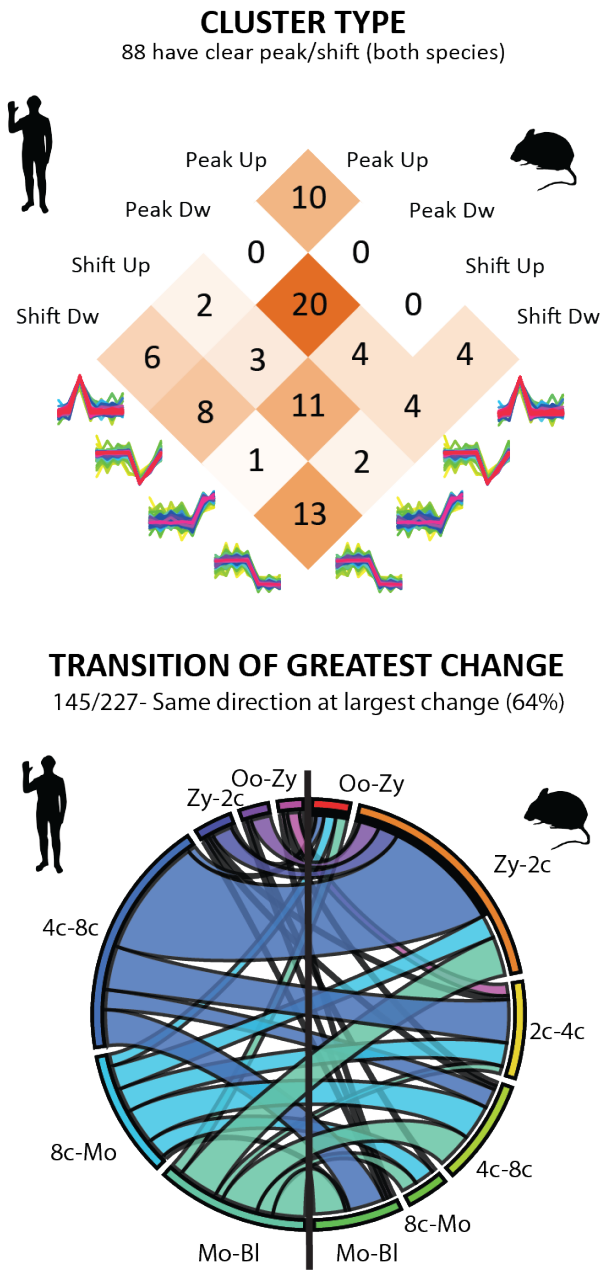


Figure R.5.1.b Comparison of human/mouse exon skipping events. Of events potentially conserved between human and mouse: **Top:** 88 events were in peak/shift cluster pairs (i.e. no other classifier), in each square shows the overlap between cluster type, either peak/shift up/down. **Bottom:** Overlap between the transition of greatest change for those (145) changing in the same direction in human/mouse.

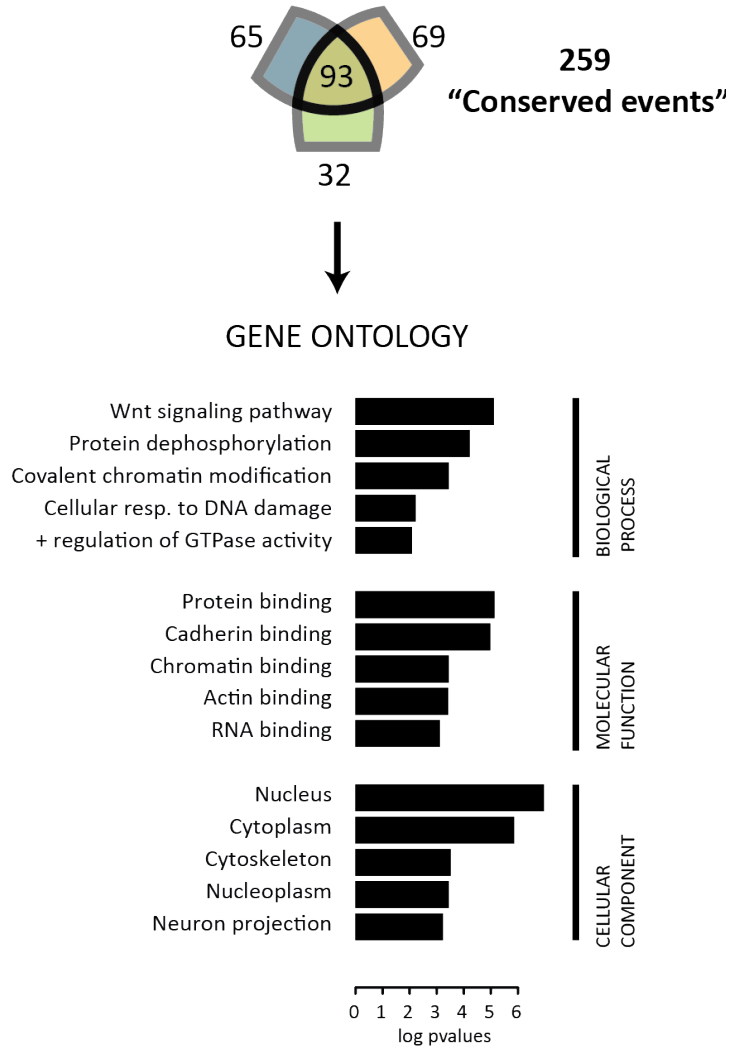


Figure R.5.1.c. Gene ontology of putatively conserved events in the three species. The 259 conserved events between the three species. Using the human list of ensembl gene IDs. Showing the top 5 terms for biological process, molecular function and cellular component. Log binomial p.values are calculated for each.

Finally, we explored what functional roles these exons. Using GO analysis, we find significant enrichment for genes involved in key signalling pathways (Hippo/Wnt pathways), transcription and chromatin modifiers, and genes related to morphogenesis (Figure R.5.1.c) (e.g. cell-cell adhesion). Many were in core genes of pathways (e.g. Yap, Nf2, and Tead4 in the hippo pathway). Full details of their potential affect in early development are discussed in the Supp. Discussion, on page 147, with an overview in Table R.5.3.

		Event (example)	Domain/Effect	Role
DNA damage	Whsc1			H3K36 mono- and di-methylation DNA replication and repair
				Exit of pluripotency
Epigenome	Bptf		PHD	H3K4me3 modifications
	Ezh2	MmuEX0017712		
	Top2b	MmuEX0048511	ORF disruption	TAD formation
	Dnmtb3b	MmuEX0015337		Denovo DNA methylation
Transcription	Tcerg1		Alt isoform	Transcriptional elongation
Apoptosis	Rpap4	HsaEX0055439	PIH1D1 int.	Regulation of apoptosis
Hippo path.	Tead4			
	Yap			
	Nf2			
Misc.	Titin		IG & PEVK	Unknown in embryo
	Rab6a			
	Dmn2			

Table R.5.3. Table of events expected to have an influence on important pathways involved during pre-implantation development.

In conclusion, many events appear to be changing in a similar way. Further, due to coverage issues we are likely underestimating the levels of conservation. Their enrichment in core pathways involved in embryo development suggest that at least some isoform changes have critical roles.

5.2 Species-specific events

Noticeably, many exon-skipping events were found to be species-specific (Figure R.5.1). This could help explain some of the species-specific differences in the early embryo. To explore these events, we studied two particular groups of species-specific exons, those that are not present in the other two species because the gene/exon is unique to the species (Mode 1- “unique”), and those exons present in the other two species (at the genomic level) and that have coverage to detect splicing, yet are found to change only in one species (Mode 2- “regulatory”).

For “Mode 1” exons, for example, 537/3554 human exons could not be detected at the genomic level in mouse and cow (Figure R.5.1.a). These exons tended to be enriched

for disruption upon inclusion (30%, 30%, 54% in human, mouse and cow) and in the 5'/3' UTR (53%, 42%, 27%). Almost none of these exons were predicted to cause disruption upon exclusion (Human: 0.01%, Mouse: 0.18%, Cow: 0%). They were also more likely to be peak up clusters (29%, 25%, 47%).

Comparing the gene ontology terms of genes with species-specific events (Table 5.2), highlights DNA-templated transcription in both human and mouse and cell division (Figure R.5.2), though cow did not show enrichment for the same terms (Supp. Fig. 5.2). Further, when comparing the genes that are affected, there is a large enrichment of ZNF (ZiNc Finger) proteins in human (43/537 compared to 1/259 conserved exons), in mouse similarly there was slight enrichment in Zfp (Zinc Finger Proteins) genes (12/219 unique vs 1/259 conserved). These exons most likely represent cryptic exons and likely do not contribute to the function of the gene, especially given almost none disrupt the open reading frame upon exclusion.

The other set of species-specific events are those that are detected in the other two species, and have coverage to detect splicing, but are only alternative in one species (Supp. Figure R.5.2.b; Type 2). Therefore, these events only change at the regulatory level. Gene ontology analysis of this set of exons revealed similar terms to those in the conserved set (Figure R.5.2.b; Human: DNA repair, G2/M transition of mitotic cycle, chromatin modification; Mouse: regulation of transcription, Wnt signalling), suggesting that the species-specific regulatory events may have a similar functional role (to known conserved events). This may mean that the regulation of these exons has changed between the species, although it is difficult to determine whether we may have missed an event changing.

Chapter 6. Lineage fates and pluripotency

One of the most important transitions in early development is the change from totipotency to pluripotency, and the appearance of cells restricted to a particular fate. The trophectoderm (TE) appears in the outer cells of the blastocyst and will develop into the placenta, while the inner cell mass (ICM) will further separate into the epiblast (EPI), that will form the embryo proper and the primitive endoderm (PrE), forming the amniotic tissues. How these cells separate in terms of gene expression has been well studied (Guoji Guo et al. 2010), yet we know very little about the regulation of splicing in these cells, with most our knowledge coming from ESCs (Han et al. 2013), that are derived from the ICM.

This section first dissects the embryo in terms of predicted fate to explore splicing differences (6.1), then explores how known embryonic stem cell (ESC) specific-exons are regulated during early development (6.2).

6.1 Lineage specific exons

To explore the splicing patterns in the different emerging lineages, we first determined the predicted state of each single-cell from human (Yan et al. 2013; Blakeley et al. 2015) and mouse (Deng et al. 2014; Boroviak et al. 2015) blastomeres, by clustering with known transcription factor markers for each lineage (Guoji Guo et al. 2010) (Supp. Figure 6.1.). This confirmed the original classifications given in each paper, and allowed us to separate Deng et al. 2014 (which had no lineage classifications). Further, we could cluster the individual cells by global gene expression values and by alternative exon inclusion (Figure R.6.1.a). Each experiment was run separately, as the data were not directly comparable, given different timings. This shows that gene expression separates the putative lineages reasonably well, but that alternative exon inclusion is less clear.

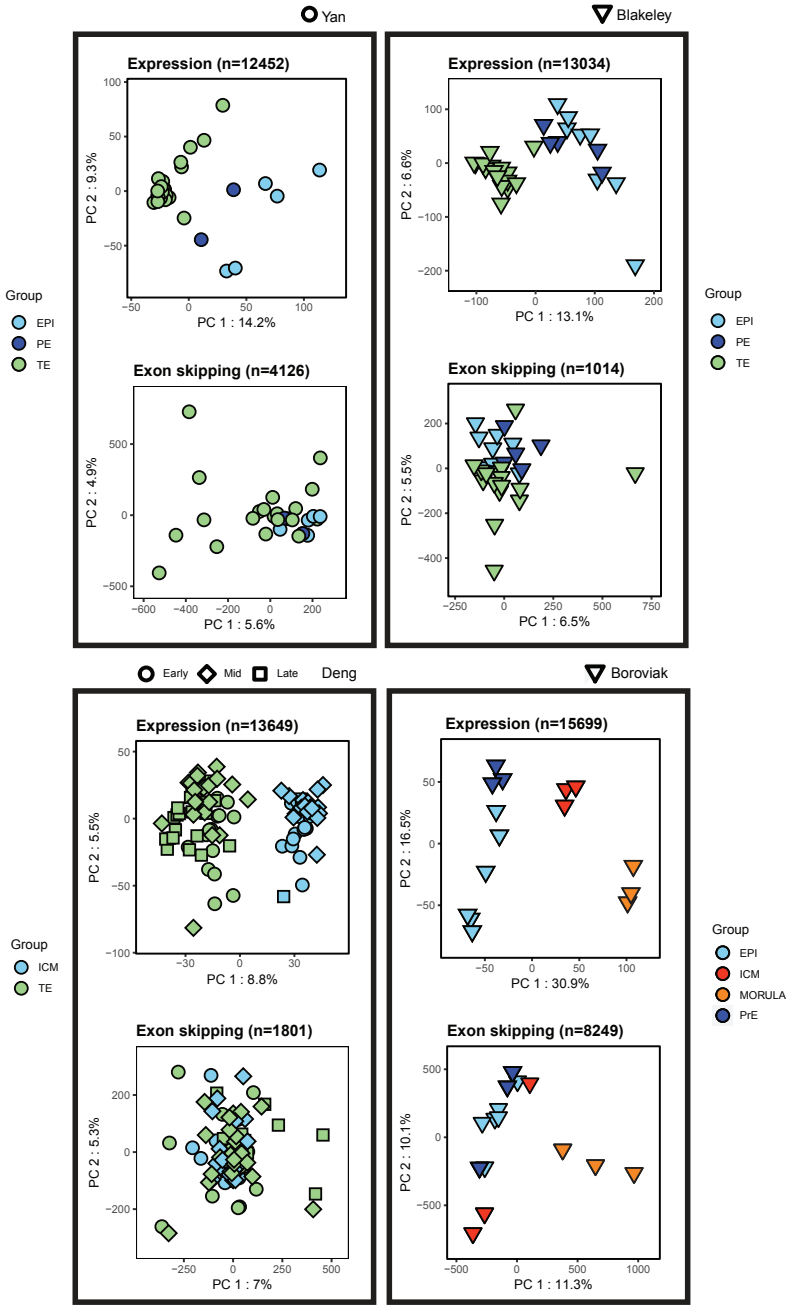


Figure R.6.1.a Principle component analysis of blastomeres in four experiments. Top-left, human blastomeres from (Yan et al 2013), Top-right, human blastomeres from (Blakeley et al. 2015), Bottom-left, mouse blastomeres from (Deng et al. 2014) and bottom-right blastomeres from (Boroviak et al. 2015). Gene expression PCAs were run using standard deviation 30 and alternative exon PCAs used coverage of 50% across blastomeres with imputation and standard deviation of 2.

Especially in mouse, there is indistinguishable difference in Deng et al. 2014 and inconsistency between replicates in Boroviak et al. 2015; this may suggest that in mouse at least there is not much separation of splicing identity by the blastocyst cells in different lineages (at this stage). Although, it could also be that the signal is weak in this data given the lower coverage in individual cells (Deng~0.96 billion base-pairs per sample, Boroviak=~33.7billion, Yan=~3.1billion and Blakeley=~20.4billion).

Next, we calculated the number of splicing changes, between the ICM and TE for mouse, and between EPI, PE and TE for human, using single cells as replicates (minimum two replicate per condition; Full details in methods: Lineage specific exons; Events: Supp. Data 14). For mouse, we found 109 events different between ICM and TE (Deng). In human, we found, 76 and 74 events different between EPI/TE in Blakeley and Yan respectively, plus 74 and 18 events between TE/PE, as well as 61 / 43 between PE/EPI. These numbers are quite low and are consistent with the PCA clustering (Figure 6.1.a).

For human, as we have the same lineages sequenced from two experiments, we could look at the overlap of events that change in a similar way between (Yan et al. 2013) and (Blakeley et al. 2015). In addition, we used a relaxed definition of differentially spliced events (dPSI of 25 and no range overlap filter), to explore further the potential overlap between these experiments (given the first approach may be too strict). In this way, we found that the strict approach could only find 2,1 and 0 in EPI/TE, EPI/PE and PE/TE, whereas the more relaxed approach found 38, 21 and 14 events in 38, 18, 12 genes respectively, that change in the same direction between the lineage pairs (Figure R.6.1.b). Of these genes, 10/68 are involved in the regulation of transcription (TBP, WBSCR22, HELLS, NCOR1, RBM39, ZNF195, ZNF232, POLR3G, INO80C and PMF1), Two are epigenetic regulators (DNMT3B and HELLS), of which, HELLS is known to be up-regulated in the ICM and DNMT3B in the trophectoderm in cow (Hosseini et al. 2015), where they have been implicated the restructuring of the epigenome in the two lineages. Having alternatively spliced differences between the two lineages may also be important for their development, but will require in depth study of these isoforms to discover potential function in the lineages. However, we do not see much change between lineages even when using relaxed thresholds for differentially spliced events, so likely spliced isoforms do not play a major role in the differentiation of lineages at this stage.

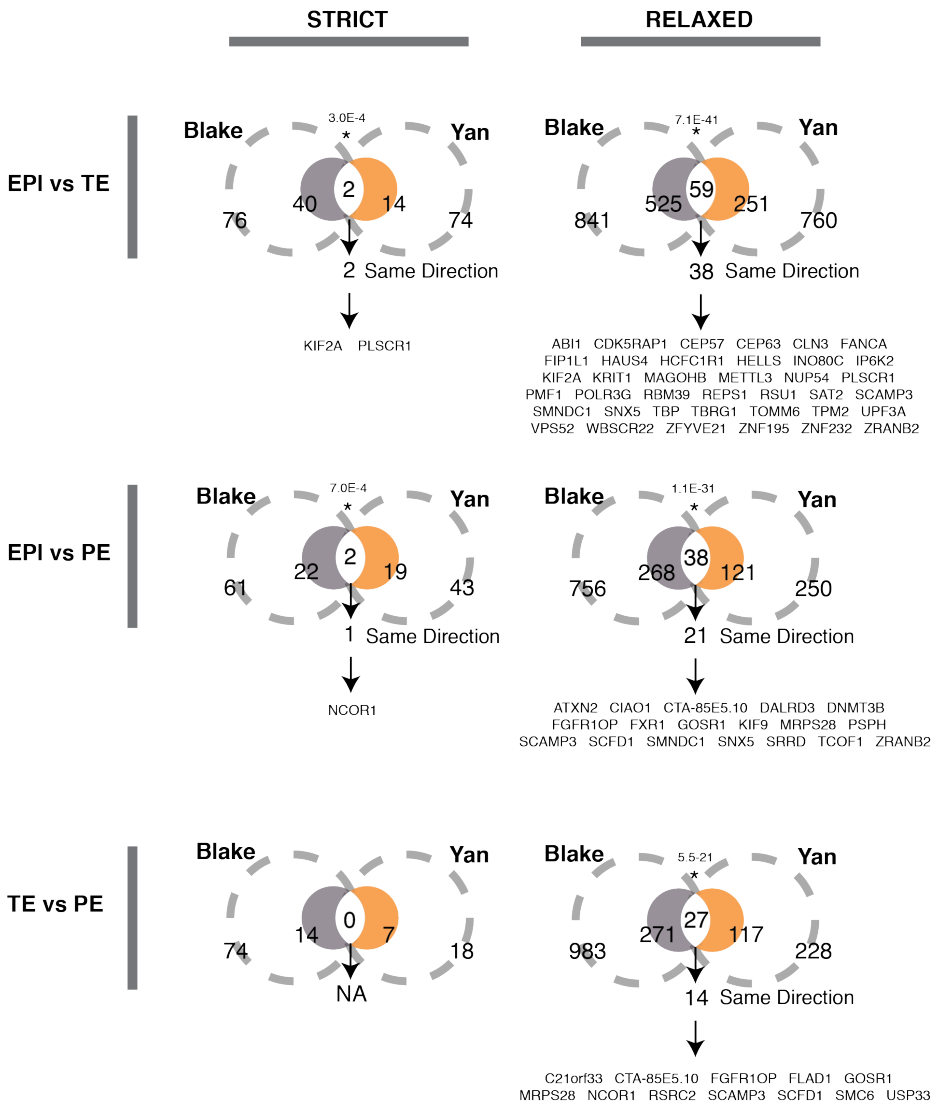


Figure R.6.1.b. Comparison of alternatively spliced events (Alternative exons, Intron retention and 3/5' splice site), between pairwise comparisons of human blastomeres from (Blakeley et al. 2015) (Blake) and (Yan et al. 2013) Yan, representing epiblast (EPI), Trophectoderm (TE) and primitive endoderm (PE). Using both a strict (left: single blastomeres; n >= 2 in each condition; range overlap of 10, same as Figure R.1.3.a: human) and a relaxed (right: single blastomeres; n >= 2 in each condition; dPSI 25 only, no range limit) approach. Larger circles represent the total number of AS events different between the lineages; smaller coloured circles show numbers of events that had coverage in the other experiment; the overlap is the number of events shared between the two experiments. The numbers below each overlap show the number changing in the same direction in both experiments, with gene names of these below. A hypergeometric p.value above shows the significance of each overlap. The total number of events that could be compared between experiments was 21499 10547 and 13017, for EPI_TE, EPI_PE and TE_PE respectively.

6.2 ESC specific exons

Given we could not detect many splicing changes between the lineages, we were curious to explore the inclusion of known ESC-specific exons. (Han et al. 2013) could show that specific exons were included in ESCs, compared to all differentiated tissues (or the reverse: skipped in ESC/included in differentiated tissues), suggesting isoform identity may be important to the proper functioning of pluripotent cells. Pre-implantation embryos transition from totipotency to pluripotency, which are associated with epigenetic changes (G Guo et al. 2017) and characteristic gene expression pattern (F. Lu and Zhang 2015), so we expected to see changes in ESC-specific exon inclusion.

To investigate this, we first defined ESC-differential exons using a large set of ESC samples, and differentiated cell and tissue types from VastDB (Tapial et al. 2017) (Samples and INCLUSION tables: Supplementary data 8), using expanded definitions for ESC-differential exons following (Irimia et al. 2014)(see Methods; comparing all ESC samples versus all differentiated tissues, but not including other pluripotent stages).

This identified 897 and 576, human and mouse ESC-differential exons with either increased (human: 374, mouse: 227) or decreased (human: 522, mouse: 355) PSI ($\Delta PSI_{ESC\text{-}DIFF} >= 25$; Supp. Table 6.2- Tab1/2).

For each ESC-specific exon ($\Delta PSI_{ESC\text{-}DIFF)}$, we could compare its ΔPSI, to the ΔPSI of each preimplantation stage versus differentiated tissues (e.g. $\Delta PSI_{2Cell\text{-}DIFF(EXCEPT\ PLURIPOTENT[EX])}$, $\Delta PSI_{Morula\text{-}DIFF(EX)}$), between iPSCs ($\Delta PSI_{IPSC\text{-}DIFF(EX)}$) and primordial germ cells (PGCs; $\Delta PSI_{PSC\text{-}DIFF(EX)}$) or between other differentiated tissues (e.g. $\Delta PSI_{NEURAL\text{-}DIFF(EXCEPT\ NEURAL)}$, $\Delta PSI_{\ MUSCLE\text{-}DIFF(EXCEPT\ MUSCLE)}$). Examples for 2-cell/8cell (ZGA stage), iPSC and Neural are shown in Figure R.6.2.a (scatter plots) for human and mouse, where it is noticeable that both toti/pluripotent stages have many exons that are included/skipped in the same way (warm colours), whereas for neural most points are not changing in the same way (dark blue). This was quantified for all the early stages (Figure R.6.2.a; barplots) and split into proportions of events changing compared to ESCs. This shows that ESC-specific exons are mostly skipped/included in a similar way, and ~40% of pre-implantation (and pluripotent) stages had an even stronger difference (brown) in inclusion/skipping compared to ESCs. In contrast, differentiated samples showed the opposite pattern (Figure

R.6.2.a). A similar pattern was obtained within the different lineages of the blastocyst, suggesting that the inner cell mass does not have a greater pluripotent AS signature compared to trophectoderm (Figure R.6.2.a). These ESC-differential exons included most previously experimentally validated examples (Han et al. 2013), involved in pluripotency and regulation of differentiation, including FOXP1-ES (Gabut et al. 2011) and NUMB (Revil et al. 2010) (Supp. Table 6.2- Tab3/4). Finally, these results are consistent with the very low expression of Mbnl1 and Mbnl2 genes during early embryogenesis in the three mammalian species (Figure R.6.2.a – MBNL expression). This is consistent with which its known role as the master regulator of ESC-differential alternative splicing (Han et al. 2013).

These results suggest that there is a common set of alternative exons that are differentially spliced between toti/pluripotent and differentiated cells. Unfortunately, we didn't analyse data from oocyte maturation stages and later post-implantation stages, which would help show when cells acquire/lose this specific "pan-pluripotent" splicing program.

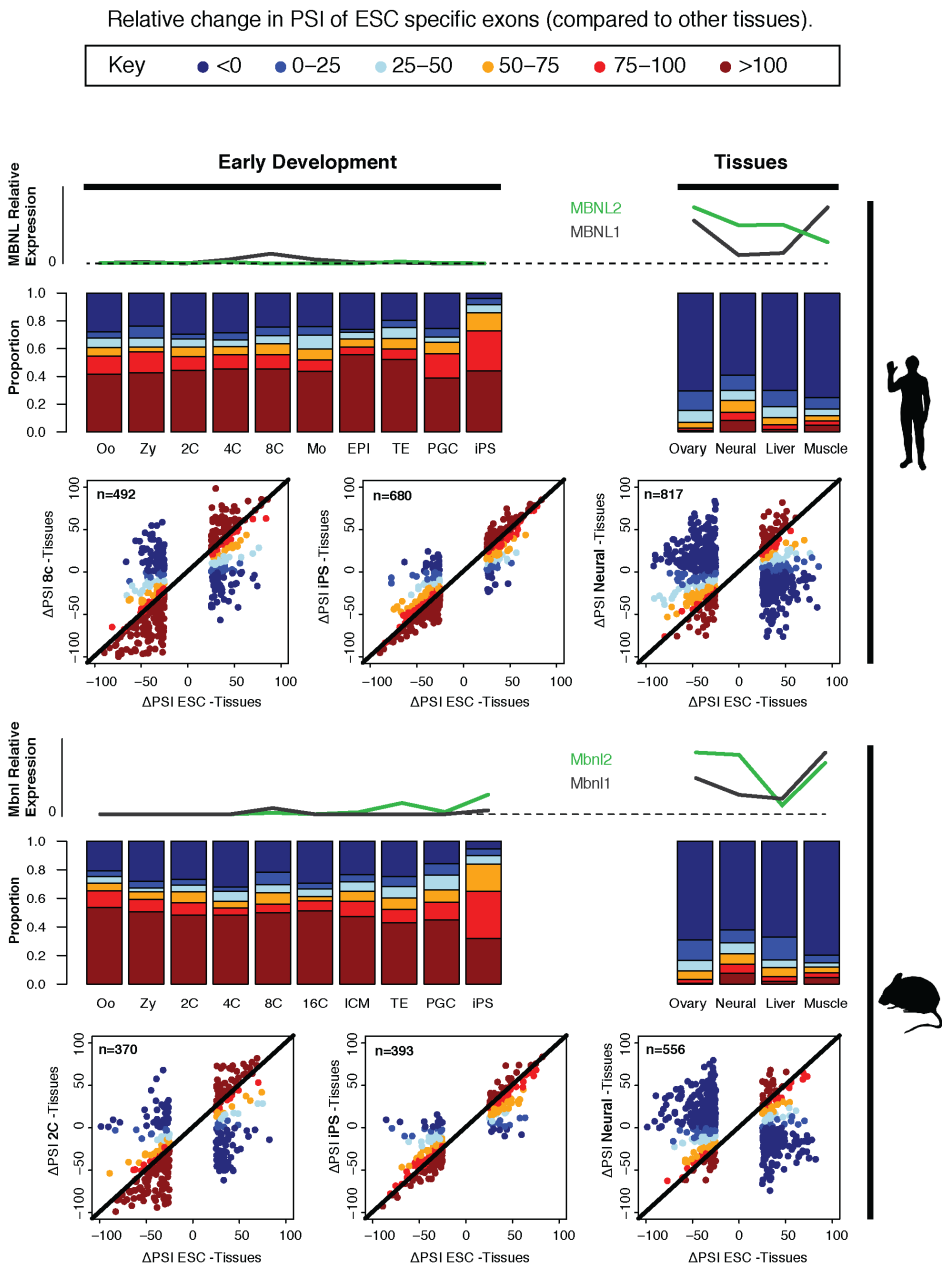


Figure R.6.2.a Mbnl-regulated, ESC-specific exons in early development and tissues. For human (top) and mouse (bottom). Upper: Showing the relative expression of MBNL1 and 2 for the stages of oocyte to blastocyst/pluripotent stages and to right the differentiated tissues. Middle: The proportion of ESC-specific exons with coverage in each stage and their behaviour, either changing in a greater difference to ESCs (dark red), of the same direction (red tones), not changing (lighter blues) or changing in opposite direction (dark blue). Lower: For the stage of genome activation (2-Cell or 8-Cell), iPSC cells and neural cells the delta PSIs of each to tissues versus ESCs to tissues.

6.3 Pan-pluripotent alternative splicing

To extend our discovery of "pan-pluripotent" exons, we calculated the number of exons specific to the whole set of early stages from oocyte to male and female primordial germ cells (PGCs), induced pluripotent stem cells (iPSCs) and ESCs. For alternative exons with sufficient read coverage across these samples, we required a consistently higher or lower PSI ($|\Delta PSI|>15$) between each of these pluripotent samples and the average in differentiated cells (see Methods). We identified 450 human and 344 mouse alternative exons with increased and decreased inclusion in toti/pluripotent cells compared to differentiated tissues (Figure R.6.3; 6.2- Tab5/6). "Intracellular signal transduction" was the most enriched GO term in the two species with "immune system process" also present in both species.

We found just 14 genes specific in both human and mouse of 126/99 (11-14%) common in the two species (Figure R.6.3), which although has a significant hypergeometric p.value, is quite low. Of these genes, many could have specific-roles in pluripotent cells, including SORBS1, which is involved in cell adhesion and growth, with roles in insulin-stimulated signalling and glucose uptake. NUMA1 was shown to be a novel component of SWI/SNF (or BAF complex) in ES cells (Ho et al. 2009), a complex which is essential for the formation of totipotent and pluripotent cells through its role in chromatin remodelling (Ho et al. 2009).

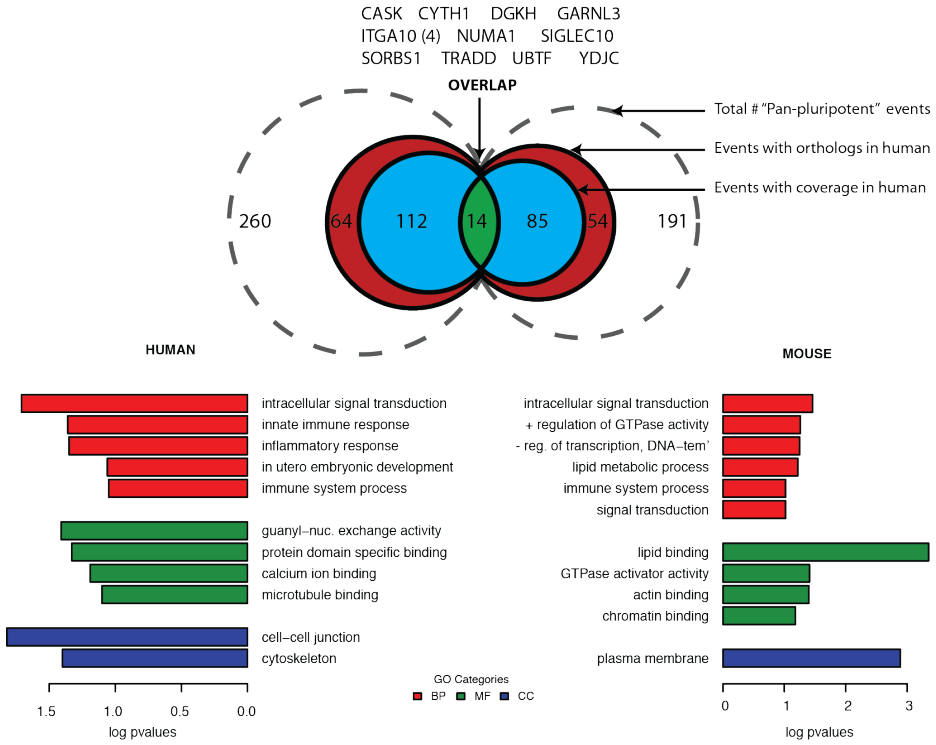


Figure 6.3. Pan pluripotent exons. TOP: Venn showing overlap of pan-pluripotent exons that have ortholog in the other species and coverage. Comparing human and mouse. BOTTOM: Gene ontology (barplots) to show the top enriched terms for genes with pan-pluripotent exons (total: 260 and 191)

In summary, pluripotent splicing characteristics are not restricted to the ICM/ESCs, but are present from the oocyte to blastocyst stage as well as in PGCs, trophectoderm and iPSCs, so are "pan-pluripotent". This splicing identity appears negatively correlated with expression levels of MBNL, a known regulator of ESC-specific exons (Han et al. 2013). Finally, despite some overlap in pan-pluripotent events, there appears to be a great deal of species-specific variation.

DISCUSSION

Overview

Alternative splicing of a single gene can result in multiple protein isoforms, or lead to changes in total protein level. This can be adjusted in different developmental stages or between specific cell types or tissues, allowing great flexibility in the evolution of gene function. In the early embryo, a time where massive epigenetic, transcriptomic and proteomic changes occur, the role of splicing and its ability to diversity the transcriptome and function has not been studied at depth.

During this thesis, we have detailed the dynamic changes in the embryo at the level of splicing, finding unparalleled levels of mRNA diversity at the stage of zygotic genome activation (ZGA) in human, mouse and cow. This diversity was mostly caused by hundreds of temporary exon-skipping events, many of which were predicted to lead to premature stop codons in the mRNA. We could also show that these transcripts are not fully removed by NMD, yet appear less translated by comparison to "in-frame" isoforms using TRiP-Seq.

Interestingly, PTC-containing events were enriched for DNA damage functions, suggesting DNA repair may be affected in the embryo. To understand these exon-skipping events better, we used multiple bioinformatic strategies to find potential regulators. This suggested the role of core spliceosomal proteins, including Snrpb/d2. Experiments performed in the lab (Barbara Pernaute) could show that Snrpb/d2 protein levels were lower before the stage of genome activation, in concert with known lower DNA damage response marker phosphorylated phospho-p53 and gamma-H2AX. Microinjection of Snrpb/d2 mRNA into the oocyte could reduce the level of exon-skipping and also increase the levels of phospho-p53, suggesting that DNA damage repair response may have improved due to the prevention of disruptive exon-skipping changes in DNA damage repair genes.

Further, we could also show that a fraction of events are conserved, preserved for ~90 million years (between human, mouse and cow) that have potentially vital roles in early embryogenesis, that need to be explored further.

Finally, we explored the dynamics of embryonic stem cell (ESC) specific exons during pre-implantation development, showing that in fact these exons are not only in ESCs, but are found in all early developmental stages, and are regulated by the

absence of muscle-blind (MBNL) RNA-binding proteins (RBPs), as previously shown in the literature.

Here, I discuss some of the main questions that arose during this PhD thesis and their wider implications.

Detection of splicing in preimplantation embryos

Measuring splicing dynamics in early development was a complex task, owing to the starting material being mainly low-coverage single-cell RNA-Seq, from multiple independent laboratories. For this reason, the input was heterogeneous, and contained various levels/types of bias/variability, that we needed to consider.

We can split the potential biases/variables into two categories (Figure D.1), 1: **Biological**, those that contribute variability owing to natural phenomena. 2: **Technical**: those that lead to variation that are not biologically informative (e.g. mRNA read length; Sequencer type- Illumina, etc.).

It is important to point out these differences/biases that were inherent to each dataset, as these have a great influence on what we can interpret.

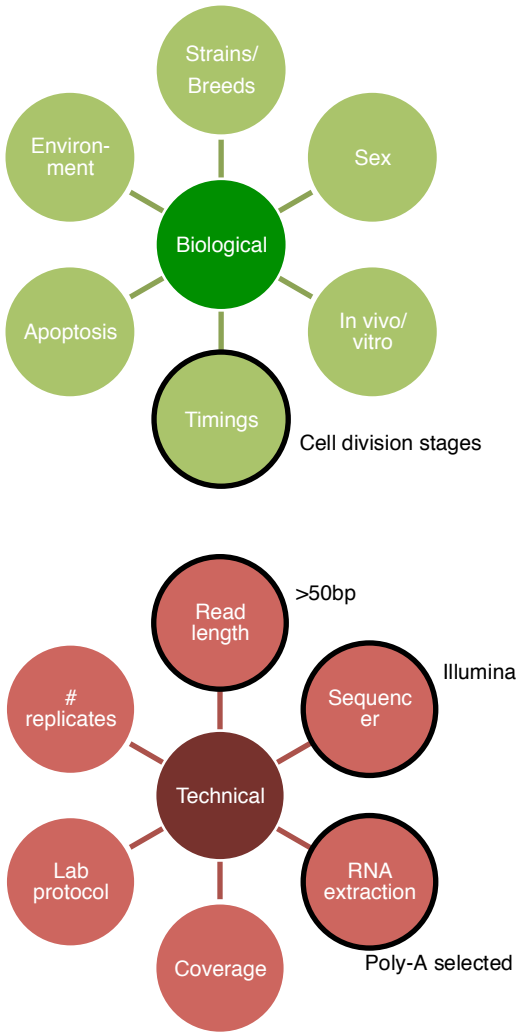


Figure D.1 Biological and technical variables. Can differ between samples and between species. Black circles show the factors that were informative of the initial RNA selection.

We initially restricted the data analysed to more obvious technical and biological differences. Highlighted in Figure D.1 (black circles). **1. Read length (>50nt).** This was a restriction imposed by vast-tools (Irimia et al. 2014). **2. Specific time points**. We only compared samples with comparable cells of each developmental stage (at specific hours post fertilisation). For this reason, we could not use human cells from (Petropoulos et al. 2016), which were not sampled at specific times post-fertilisation (and therefore exhibited a range of timings). **3. Illumina sequencers**. We discarded

SOLiD and other non-Illumina sequencing data, as comparison across different platforms can lead to biases (A. Roberts et al. 2011), in addition, most embryo data in the literature were performed on Illumina machines **4. Poly-A selection**. There are two main selection strategies to recover RNA, both resolve to remove the highly abundant ribosomal RNAs whilst selecting for either coding-only (Poly(A) selection) or non-coding & coding RNAs (ribosomal RNA depletion; Ribo-). Poly-A selection uses beads of poly-dT to cause hybridisation to the 3' tail of coding mRNA; whereas rRNA depletion entails probing the RNAs for ribosomal sequences then removing them from the sample, leaving the coding and non-coding RNAs. Each has its own bias, making it difficult to compare between the types (P. Cui et al. 2010). Given the abundance of Poly-A selected datasets, and poly-A's better coverage of exon-exon junctions (S. Zhao et al. 2018), we chose to restrict our analysis to this type. Cow data from (Graf et al. 2014) used oligo-dT and random primers, therefore the cow analysis is somewhat separate to human/mouse (oligo-dT only) and this explains the high number of intron retention events detected at genome activation, where incompletely spliced transcripts are detected.

Next, once we started to analyse the data, we ran into more potential biases/variations in the data, most of which required us to control which data we were analysing for the final comparison between stages.

1. Specific timings. Despite using data from timed experiments, the exact time of stages sampled in each paper differed. For the most part, embryos were sampled at specific hours post fertilisation (hpf) or post hCG (human chorionic gonadotropin; a hormone treatment to promote egg maturation), but differences existed in the timing used in each source paper. In mouse, (Biase, Cao, and Zhong 2014) could not be used, as timings were different from (Deng et al. 2014) (Figure D.1.). Biase *et al.* 2014 sampled zygotes at 26h post-hCG, 2-cell at 48hpf and 4-cell at 52hpf, whereas Deng *et al.* 2014 sampled zygotes at 31-32 post-hCG, 2-cell from 31-48hpf and 4-cell at 54-56hpf. These timing differences increased the apparent variation in each stage, making it more difficult to distinguish changes between stages. Therefore, in this case, (Biase, Cao, and Zhong 2014) data was removed from the global analysis, in order to fairly compare transitions.

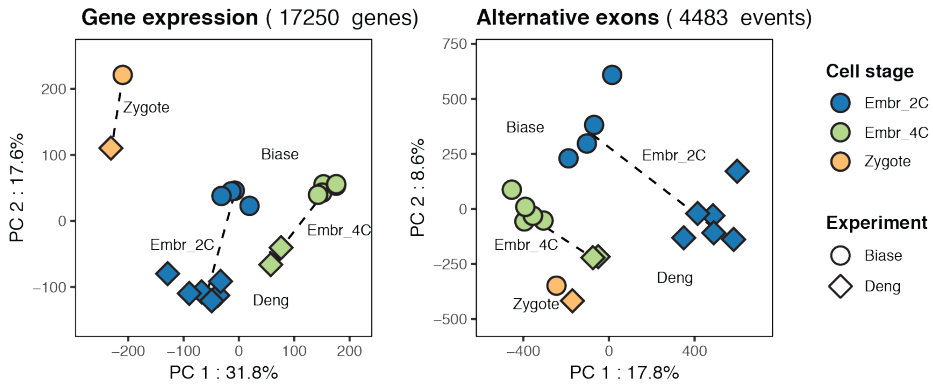


Figure D.2. PCA plots showing the difference between (Biase, Cao, and Zhong 2014) and (Deng et al. 2014) Zygote, 2-cell and 4-cell stages (mouse). Principle components 1 and 2, with percent variation explained on each axis.

2. Three prime bias. Different Poly(A) selection protocols can lead to 3' bias. This is not a major problem when comparing within a single experiment, but when comparing multiple RNA-Seq dataset (where 3' bias can differ), reads can be biased. Where more reads map to the 3' poly-A tail end, meaning exon-exon junctions will have higher coverage compared to the 5' end. A full analysis of 3' bias is found in Supp. Figure D.1. For this reason, several RNA-Seq datasets had to be removed from the study to ensure there was no bias (Supp. Table 1 [Tab 1 and 2]). For example, the SuperSeq method (Fan et al. 2015) had high levels of 3' bias, making it incomparable to the other datasets. To make fair comparisons, only samples where less than <40% of reads were in first fifth of the transcript from the 3' end were considered.

3. Methodological bias (number of replicates). The variable coverage in samples across the time course led to bias in the calculation of differentially spliced events. For example, in cow, we could only attain one replicate for 4, 8 and 16-cell embryos (after merging); plus zygote/ 2-cell were not available. This meant that we ended up using two rules to call an event different, for cow one replicate and in human/mouse two replicates were required (Figure R.1.3.a). These differences led to substantial biases in the total numbers of alternatively spliced events detected for each transition and species. Therefore the numbers and accuracy in cow is questionable, and likely not fully comparable to the other two species. It is also highly likely to affect the numbers between stages, as more replicates increase the potential noise/variability in each sample.

4. Anomalous cells. Some cells did not cluster well with the same cells of their respective stage (Supp. Figure 1.1.i; listed in Supp. table 1[red highlighted]). These cells were removed from the analysis, as we could not determine the reason why they did not cluster with cells of the same embryo, though we presume these were apoptotic/dying cells. These anomalies were not consistent across multiple embryos, therefore it seemed reasonable to exclude these individual cells from the analysis.

There were also other biases, most of which we could not take into account (see a full list in Figure D.1), which may have had an influence on our results.

1. Laboratory/Protocols. Exact reagents and equipment varied between different RNA-Seq datasets, and often they were badly documented within each publication. This could potentially have a big effect on the sequencing results, given the method of cell manipulation, pipetting, reagents and concentrations varied. For instance, the proteases used to remove the zona pellucida or separate blastomeres into single cells can differ. This could have profound affects on the cells.

2. Bulk vs Single cell. Cow RNA-Seq used in this thesis was pooled-cells/bulk sequencing, ~10 embryos per sample (Graf et al. 2014), whereas for human (Yan et al. 2013) and mouse (Deng et al. 2014), mostly came from single cells. Considering we merged cells to gain coverage, we do not think these differences would have a great bearing over out results, but is worth taking into account.

3. Coverage. Differences in the numbers of reads per sample could affect the level of noise in a sample. It was partially for this reason that high-throughput technologies, such as Drop-Seq (Macosko et al. 2015) and Smart-seq (Picelli et al. 2014) could not be used (which have <10million reads per single cell). We did however run the 1529 human pre-implantation cells (Petropoulos et al. 2016), which showed much greater variation in PSI (data not shown). This is likely because the samples were taken from a continuous time course (not staged at 24 hour, 48 hours etc) and due to the need to merge >20-40 cells to have single replicate with >150million reads (necessary to measure splicing changes). For this reason, we couldn't incorporate this dataset into the analysis.

4. Read length. Potentially comparing different read-lengths can affect measurements of splicing, as short reads are more likely to be mapped multiply and therefore discarded from the calculation (Chhangawala et al. 2015). Our data ranged from 50nt

single reads to greater than 100nt paired-end. Read length is likely only to affect coverage, leading to loss of detection of some events, but not in false positives.

In summary, due to the existence of bias in different datasets, we could not compare all the available data. Despite this, we were able to find enough comparable samples to have decent coverage across the time course for the three species (Figure R.1). Ultimately, we had to rely on the multiple validations (in mouse, performed by Barbara Pernaute; Supp. Fig 1.3.iii), which showed the majority (89% of those tested) of changes predicted were real. Therefore, we are reasonably confident with the inference of PSI values for each exon, yet due to coverage issues, we likely fail to capture all the alternative splicing events present in the embryo.

Transcriptomic complexity at genome activation stage

A surprising initial result was the level of isoform "complexity" that we found at the corresponding genome activation stage in each species (Figure 1.3). Using large quantities of RNA-Seq from a wide range of tissues taken from VastDB (Tapial et al. 2017), and using a simple measure (1 or 2 isoforms per exon or "event"), not a single tissue or developmental stage had such a high complexity.
One criticism of this result is that only 66 human and 70 mouse-differentiated tissues were analysed (Supp. Table 1), so we cannot fully discount that there may be a tissue/cell-type/stage with more isoform diversity. Though, this represented a very broad range of cell/tissue types and cell lines. In addition, given our data is mainly low-coverage single-cells, levels of isoform diversity may be underestimated compared to the differentiated tissues (that are also composed of multiple cell types, and mostly have higher coverage [see Supp. Table 1; read numbers per sample]). Therefore potentially the signal we detect could be weaker than the reality. Despite technicalities, this suggested that zygotic genome activation stage is highly diverse in terms of isoform.

Analysis of frame preservation could tell us that many of these newly synthesised isoforms were predicted to be NMD targets or lead to disrupted protein isoforms, suggesting that this diversity may not necessarily translate into protein diversity. Our

analysis of TRiP-Seq also confirmed that these disrupted isoforms may not be translated (Figure R.2.3.). In addition, given many genes are not pre-implantation embryonic lethal (Artus and Cohen-Tannoudji 2008), likely due to loading of maternal proteins/mRNAs, it has been suggested that the initial burst of gene expression and spliced isoforms are not required at this time, or that they could represent a pioneer round of translation (Maquat, Tarn, and Isken 2010). This could mean that they will not be required for protein synthesis. This could fit with existing proteomics data, showing that even though at the RNA level most genes change at ZGA, most protein level changes happen at the blastocyst stage (in mouse; Gao et al. 2017). It could also be that given NMD potentially is inactivate at this time, and given genome activation may overload the machinery, what we detect at this stage may be biological "noise", that does not have a function, but is tolerated by the cell because of maternal proteins are in control.

However, we should not expect all isoform changes to be translated into protein or have a role. It could be that a small numbers of events are required, as we have suggested for the potential role of events in DNA damage repair genes and their regulation by Snrpb/d2. It is also important to note, that we do not detect general exon-skipping but it is specific exons that are affected, so it seems likely that the changes are not random, but a coordinated response in the three species. This pattern of temporary disruptive isoforms was also found in zebrafish (Aanes et al. 2013), suggesting this phase of temporary disruption may be a conserved feature beyond mammals. However, we found little enrichment in any of the specific subsets of DNA damage categories (Homologous recombination, mismatch repair, etc.) and little overall conservation in the exact exons changing.

In summary, we found a distinctive exon-skipping phenotype at ZGA in three species, which contributes to a massive change in isoform diversity. The enrichment for disruption and less translation suggest that the majority of changes could be non-functional. However, disruption or truncation in specific genes could lead to lower protein levels and to function.

Are core splicing factors responsible for exon-skipping at zygotic genome activation?

We next looked at potential regulators, to determine if a specific RNA-binding protein may be responsible for the patterns in splicing detected. Despite finding no clear enrichment of a motif/s around alternative exons, our comparisons to RNA-binding protein knockdown/control was more informative. SNRPB and other core splicing components (e.g. U2af1 in mouse) had a similar exon-skipping phenotype to that during genome activation. This could explain in part why we didn't detect any major motif enrichment profiles for these events, as Snrpb/d2 does not directly bind RNA, but is central to the biogenesis of the Sm ring used into multiple splicing complexes (U1, U2, U4, U5). Though it does not explain why specific exons are skipped.

The knock-out for SNRPB came from a paper studying glioblastoma in human (Correa et al. 2016), which found mostly exon skipping after knockdown. Gene ontology of events changing after knockdown were highly enriched for DNA damage related terms (Figure D.3), as we found for events becoming skipped at genome activation in our data. Further, (Saltzman, Pan, and Blencowe 2011) has shown that a knockdown of SNRPB (siRNA in HeLa cells), results in 20x more exon skipping than exon inclusion, many of which were due to skipping of constitutive exons and a high frequency of premature termination codon (PTC)-containing alternative isoforms, in line with our results. In contrast, Saltzman et al. 2011 found that changes to PTC-containing isoforms, had lower gene expression, and the reverse had higher expression, suggesting that NMD may be reponsible for the change.

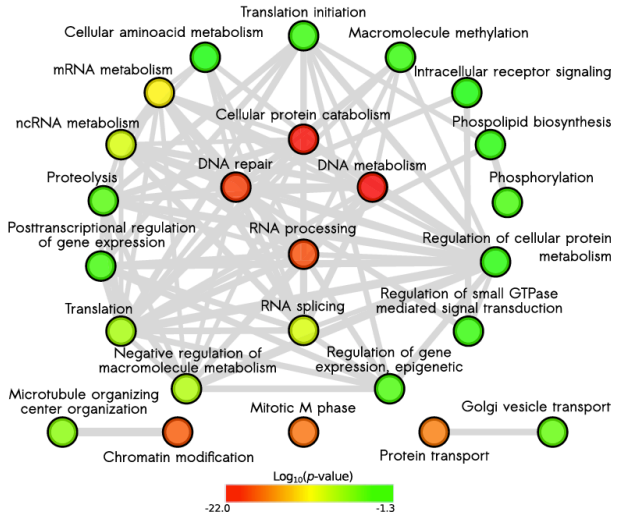


Figure D.3. Gene ontology of genes with splicing events changing after SNRPB knockdown. Node colour represents significance. (Correa et al. 2016).

The genes in the complex in which Snrpb is associated are known for their mutations that are linked to cancer. In myelodysplastic syndromes (MDS; U2AF1) (Okeyo-Owuor et al. 2015), and in hematopoietic malignancies (SF3B1, U2AF1 and SRSF2) (Makishima et al. 2012). Given SNRBP knockdown in human affects DNA damage repair genes (through splicing), it could be that mutations in related core factors may lead to similar DNA repair deficiencies, which would make sense if they lead to genomic instability. Further, the core splicing components were found enriched in a genomic wide siRNA (small interfering RNA; ~20,000 gene) screen of RAD51 foci after DNA damage (IR) (Herr et al. 2015). In another study, is was shown that U2 snRNP factors (including SNRPA1, SF3A3, SF3B3) are required for genomic stability, by maintaining levels of DNA damage proteins and preventing R-loop induced DNA damage (Tanikawa 2017).

Is DNA repair truly weaker at genome activation?

Previous studies have found pre-implantation embryos to be both DNA-damage insensitive or hypersensitive. (Yukawa et al. 2007) found that gamma-H2AX phosphorylation is much lower in mouse from the 2-cell to 8-cell stage. (X. Wang et al. 2017) showed that Rad51 (another marker for damage) and gamma H2AX foci

were reduced in macaque oocytes and early embryos, yet were present in mouse. (Gawecka et al. 2013) could show that sperm chromatin fragmentation leads to stalling of the zygotic cell cycle, suggesting that the embryo can respond to damage at this time. Finally, (Derijck et al. 2006) could also show that the zygote appears to respond to double-stranded breaks (as shown by gamma H2AX staining), yet did not look at later stages. Therefore we cannot be sure that DNA repair is more/less active at this time, and it may be that we cannot detect the components.

There are several reasons why DNA repair may not be detected at this stage,

1. Transcriptional activation. The massive levels of transcriptional activation could potentially overload the system, in that the embryo could be compromised whilst activating the genome, and therefore cannot adequately activate the DNA damage response.

2. DNA-damage overload. Given the early embryo inherits a potentially fragmented genome from the paternal sperm, the early embryo may be inundated with damage to repair. This is turn could either mean DNA damage foci are difficult to detect at this time, or they are diluted.

Given that DNA damage itself can lead to exon skipping in several cell contexts (Shkreta and Chabot 2015), we also tested whether the exon skipping in early development may in part represent a direct splicing response to DNA damage.

Using comparisons with transcriptomes of cells exposed to DNA damaging agents (Supp. Table 1: [section: DNA damage]), we could make similar comparisons to our RBP knockdowns (Supp. Figure 4.6 [folder of plots]). If we look at genome activation stage in human (Figure D.4) the top hit is etoposide treatment (6e-33), followed by Methyl methanesulfonate (MMS) from two different cell lines (and experiments), where interestingly they have a similar profile to SNRPB, with massive overlap of skipped exons at 4-8 cell stage and skipped exons with treatment of DNA damage agent. Mouse comparisons were confounded by low coverage, but did not show much significant difference at genome activation stage, although etoposide treatment did show a similar exon skipping phenotype, though barely enriched (7e-2; Supp. Figure 4.6 [on hard disc]).

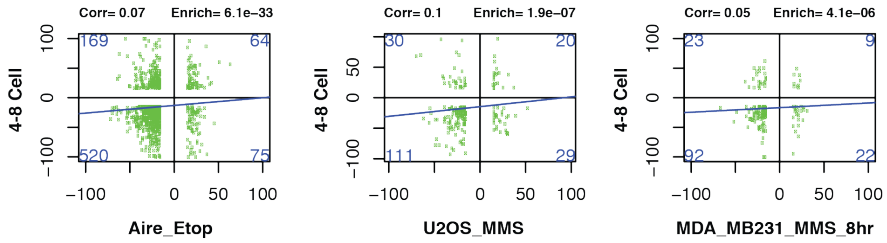


Figure D.4 Top 3 human DNA damage to 4-8 Cell transition comparisons of alternative exons. Showing the 4 to 8-cell transition dPSI values (y axes) and dPSI values for different DNA damage treatments compared to control (x axes). Correlation coefficients are top left and binomial p.values are top right.

This could suggest that the embryo is responding to DNA damage at ZGA. Equally, it could suggest that DNA damage leads to a cascade where a core protein (such as Snrpb) is deactivated, leading to skipping of specific exons.

However, given we find injection of Snrpb/d2 increases phospho-p53 foci intensity sensitivity/responsivity at this time. We believe that there are DNA damage deficiencies at this time.

Why would DNA damage be weak at genome activation?

It seems odd that DNA repair would be inefficient at this crucial time in the organism's life cycle, when any error would be carried forward to all the developing cells. There are several explanations that could explain this dichotomy, and may explain why DNA damage checkpoints may be inactive.

Chromatin-remodelling constraint. The need to completely overhaul the chromatin marks (demethylation, H2K4me3, etc), and protamine to histone replacement could require specific DNA repair genes to be removed from the chromatin, to order to allow remodelling proteins to assemble.

Transposon control. Potentially linked to early chromatin changes; we know that transposons are particularly active at this time, and some require the use of the repair machinery to transpose (Izsvák, Wang, and Ivics 2009), therefore turning off the machinery may actually increase genomic stability, and not the reverse, at this time.

Cell-cycle progression. In C.elegans, the embryo needs to divide quickly, in order to correctly time the expression of genes (Holway et al. 2006). Growth phases in the cell

cycle allow time for genes to be expressed in C.elegans. Therefore DNA damage checkpoints could increase the length of growth phase, and erroneously turn on genes earlier than required. Timing of gene expression could also play a part in mammals, where delaying the cell cycle could lead to defects in expression, though this has not been tested.

Specific telomerase activity. During the first few cell divisions it is known that the telomeres are not actively lengthened by the conventional telomerase (Figure I.3.g.) (Kalmbach et al. 2014; L. Liu et al. 2007). Telomerase is a reverse transcriptase that uses an RNA subunit as a template to produce DNA, lengthening the chromosomes. The activity of telomerase, has been linked to the Sm ring (which contains SNRPB) and LSm (like Sm; an orthologous heptameric ring) complexes in yeast (W. Tang et al. 2012). In that, the Sm complex is known to be involved in 2,2,7-trimethyl guanosine capping of telomerase mRNA, and the binding of LSm complex to the 3' end of telomerase mRNA prevents degradation and activates telomerase activity. In human, it is known that telomerase co-purifies with Sm/LSm protein complexes (Fu and Collins 2007). This may suggest that the changes in SNRPB levels while affecting DNA damage, could also affect telomerase activity. It could be that levels of SNPRB are reduced to allow a different telomerase activity at this time, which could also affect SNRPB's potential DNA damage repairing activities.

Summary

Overall, we have tried to show that splicing has an important role in the early embryo. We have detected specific patterns of splicing change in the three species, some of which show conserved dynamics. We believe at least the temporary changes at genome activation may lead to the changes in DNA damage responsonsivity that we detect.

CONCLUSIONS

Chapter 1.

- Thousands of genes are alternatively spliced during mammalian pre-implantation development, especially during genome activation.

- Exon skipping changes were largely temporary (or peak-like), and led to a high transcriptional complexity.

Chapter 2.

- Changes at genome activation had high levels of open-reading-frame (ORF) disruption, mainly through exon skipping of constitutive exons.

- This did not lead to detectable levels of non-sense mediated decay (NMD), yet potentially disrupted isoforms are less translated, by comparison to ribosome-bound isoforms.

- Disruptive exon skipping showed enrichment in DNA damage genes, suggesting that repair may be disrupted at genome activation.

Chapter 3/4. (Experiments by Barbara Pernaute)

- RNA-binding-protein knockdowns revealed the potential role of core spliceosomal factors responsible for exon skipping at ZGA, including Snrpb/d2

- Snrpb mRNA and protein levels were reduced around genome activation in mouse, at the same time as our exon-skipping events.

- Microinjection of Snrpb/d2 could partially reduce the levels of exon skipping. and could show an increased DNA damage response.

Chapter 5.

- Apart from temporary exon skipping events, scores of alternatively spliced events are conserved, often with similar dynamics in the three species.

- Conserved exons are mostly protein coding, less likely to be peak-like and are enriched in the Wnt signalling, chromatin and cadherin binding pathways.

Chapter 6.

- "Pan pluripotent" exons are characteristic of all stages of pre-implantation development, and are likely regulated by the absence of muscleblind proteins.

SUPPLEMENTARY DISCUSSION

The role of specific splicing events in key developmental pathways

Finally, to explore the potential affect of splicing in core pathways, I used the list of conserved events, in combination with manual annotations of events in VASTDB (Tapial et al. 2017), to investigate the potential affect on specific exon-skipping changes.

Epigenetic-related genes

Bptf (bromodomain and PHD domain transcription factor) has three conserved exons whose inclusion decreases in the three species after ZGA. One of these exons overlaps the PHD domain that is known to directly bind the N-terminal tail of histone H3 and its methylation state (Sanchez and Zhou 2011). H3K4me3 patterns are highly unique during mammalian early embryo development (Dahl et al. 2016). Starting from very broad domains across the genome in mouse oocytes, H3K4me3 is removed after ZGA, becoming restricted to its usual domain size, and overlapping mainly transcription start sites in subsequent stages and into adulthood. These dynamics seem crucial for correct genome activation (Dahl et al. 2016). Therefore, it is possible that concerted AS changes impacting chromatin modifiers, associated with H3K4me3, may be a response (or even a cause) to some of the epigenetic changes that occur during pre-implantation development.

Ezh2 is the catalytic subunit of the PRC2 complex, which can regulate the mono-, di- or tri- methylation state of histone 3 lysine 27 (H3K27), a mark that leads to the repression of gene expression in its targets (W. Mu et al. 2018). It is also known to be embryonic lethal, and inhibition in the embryo leads to a loss of pluripotency marker (Oct4, Sox2, Nanog) expression (X. J. Huang et al. 2014). Exon 14 (MmuEX0017712) in Ezh2 is known to be alternatively spliced between tissues and its up-regulation has been linked to cell cycle progression (W. Mu et al. 2018), where the authors could show that during sperm meiosis, the skipping isoform is prevalent where new H3K27me2 is formed, and in mitotic cells the inclusion isoform was

prevalent (that contains the active SET-domain), with active catalytic activity that increased the amount of tri-methylation (H3K27me3). The authors do not show the changes during early developmental stages; in our data we find that in the three species, all exhibit some skipping around genome activation (4C in Cow, 4C in mouse and 8C in human), yet human and cow show high inclusion (80-100 PSI) at oocyte and blastocyst (or peak-like), whereas mouse has low inclusion at blastocyst stage (~20PSI). The appearance of the skipping isoform coincides stage with lowest H3K27me3 marks around genome activation in cow and human, which is known to have low expression of some core PRC genes (Bogliotti and Ross 2012) or cytoplasmic localisation (Ross et al. 2008), and its inclusion after genome activation therefore could allow the active tri-methylating component to be re-established. In mouse, the loss of H3K27 mark is at the morula stage (Bogliotti and Ross 2012), and so does not quite fit our prediction. Overall, these patterns could further support the evidence that lack of some core components of PRC2 at this time, leads to loss of H3K27me2, in our case by exon skipping.

Top2b is a DNA topoisomerase, which can alter the topological state of the genome, involved in chromatin condensation and relaxation, during DNA transcription and replication. It is considered a core protein involved in stabilising/formation of TADs (p.25).
Interestingly the gene has lower expression at ZGA in the three species, compared to rest of the time-course. All three also have a conserved event (HsaEX0066493, MmuEX0048511, BtaEX0037409), which is a 92nt exon in the middle of the gene which when skipped is predicted to lead to disrupt the ORF. The pattern of inclusion is ~80PSI before genome activation and ~100PSI after, in all species (human, mouse and cow). Given TADs are known to be firmly established after genome activation, the removal of the truncated/NMD transcript from the embryo could facilitate the function of the active TOP2B protein.

Transcription/Splicing

Tcerg1 (or transcription elongation regulator 1) is a regulator of transcription elongation and pre-mRNA splicing (Montes et al. 2012) and localises to the periphery

of nuclear speckles (Sánchez-Hernández et al. 2016), where it is known to bind regulatory elements in the exons of specific genes (inc. Bcl-x) to promote inclusion, potentially through modulation of the elongation rate of RNApolII (Montes et al. 2012).

A well-conserved 'shift-down' 63-nt exon was found, which shows a conserved switch in PSI, one stage after ZGA in all three mammalian species. This exon is found between the first two WW domains, which are known to mediate protein-protein interactions involved in pre-mRNA splicing, specifically with the U2snRNP and splicing factors SF1 and U2AF (K.-T. Lin, Lu, and Tarn 2004).

It is unknown how this event may change the protein's function. However, Bcl-x (which contains a known target exon) is not expressed during early development, so we could not detect if splicing in this gene changed.

DNA damage

Wolf-Hirschhorn syndrome candidate 1 (WHSC1) is a histone-lysine N-methyltransferase (also known as NSD2), known to be critical for H3K36 mono- and di- methylation (Kuo et al. 2011). Methylation of this mark has been associated with regulation of transcription, DNA replication and DNA repair (Wagner and Carpenter 2012). In particular, it was shown loss of expression in U2OS cells (bone osteo-carcinoma cell line) led to loss of expression of DNA repair proteins and decreased recruitment of proteins to sites of double stranded DNA damage (Shah et al. 2016). Finally, Whsc1 has been linked to the exit from pluripotency (Unpublished, Graf lab). It is ubiquitously expressed during early development, and interestingly a 94bp event in the three species shows high inclusion at the early stages of pre-implantation development, and is predicted to lead to disruption when included, this could suggest that Whsc1 is in part responsible for the DNA damage phenotype that we detect, in that the active form of Whsc1 is required for a normal response to double-stranded DNA breaks.

Ercc1 has a 72bp exon (MmuEX17238) and (HsaEX0022996), that shows some skipping at genome activation in both species and skipping specifically only in a few trophectoderm cells but never in the inner cell mass. This event is already annotated (Sun et al. 2009), with the skipping isoform shown to be present in ovarian cancer

cells and leads to the decrease in protein levels, decreased excision repair function, leading to a reduced resistance to cisplatin (a chemotherapy drug). This may suggest that some single cells have reduced excision repair at either genome activation or a subpopulation of trophectoderm.

Apoptosis

RPAP3 contains an exon (HsaEX0055439) that is required for the protein's interaction with PIH1D1, which protects the cell against apoptosis. Knockdown of the isoform containing the exon down-regulates PIH1D1 protein level, without affecting mRNA level, suggesting the inclusion isoform stabilizes the R2AP3 protein complex (involved in cell survival) of which they both belong, whereas the skipping isoform has a dominant negative effect on the R2TP complex. Interestingly, in human the exon is skipped in several of the single cells at the 8-Morula stage, and may suggest these cell are leading to apoptosis. It does not change in mouse.

Hippo pathway signalling

The hippo pathway is important for cell proliferation, apoptosis and control of organism size, and strikingly in the three species we have three core elements alternatively spliced in a similar fashion in the three species, Yap (Yorkie in Drosophila [Dre]), Nf2 (or Merlin in Dre) and Tead4 (Saclloped in Dre). Experimental evidence for how these exons operate has already been published in both human and mouse (see references in next sections).

Yap (yes-associated protein 1) is a transcriptional co-activator that associates with TEAD family genes (including Tead 4) to activate the hippo pathway. Eight alternatively spliced isoforms have been described (Gaffney et al. 2012). We find a conserved event in the three species (MmuEX0052491, HsaEX0071941, BtaEX0001274) changing the inclusion of a 48nt (Figure S.D.2 [orange]) long exon in the gene. (Finch-Edmondson et al. 2016) could show that inclusion of this exon into the leucine zipper (Figure S.D.2 [blue]), reduced the activity of Yap (As measured by luciferase). Therefore, through alternative splicing the activity of this important gene may be changing in the early embryo.

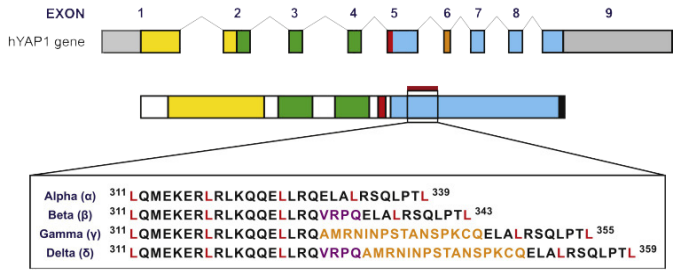


Figure S.D.2. Exon-intron structure of Yap1. Showing four isoforms Alpha to Delta, where exon 6 (orange) in this description is the conserved exon in the three species, contributing 16 amino acids to the final protein (beginning AMRN amino acid sequence). Colours denote protein features: Tead binding region (yellow), first WW domain (green), second WW domain (exon 4-green), SH3-binding region (red), trans-activation domain (blue). Exon 6 is within the leucine zipper of the trans-activation domain. UTRs are shown in grey. Figure taken from (Finch-Edmondson et al. 2016)

Miscellaneous

Titin is the largest protein in humans (Opitz et al. 2003) and the third most abundant at 0.5kg of the average adult (Labeit, Kolmerer, and Linke 1997), it is known for its role as a molecular spring involved in muscle elasticity (K. Wang, Ramirez-Mitchell, and Palter 1984). It is expressed in muscle-related tissues only (Figure R.5.8). It is greater than a micron in length and recorded to be between 3000-3700kDa depending on isoform (Bang et al. 2001).

Unusually, we found that Titin is also lowly expressed in early pre-implantation embryos. We detect inclusion of exons not found in muscle and many unannotated exons, making embryonic-Titan potentially the longest protein in any tissue or stage (Exon inclusion details in Supp. Table S.D.1-hard disk only). However, the expression is extremely low and may suggest that the role of Titin in early development may be at an earlier stage, such as oocyte maturation. The only non-muscle related function that has been published (to our knowledge) is the role of D-Titin in mitosis (Fabian et al. 2007), and predicted to be used in chromosome movement. This could suggest that a mammalian embryonic-Titin may have a role in chromosome movement. Given Titin is not expressed in all cells, it could be that this novel isoform has a putative function in meiosis (not mitosis) or even fertilisation

where specific mechanical movements are required and are specific to this stage in development only.

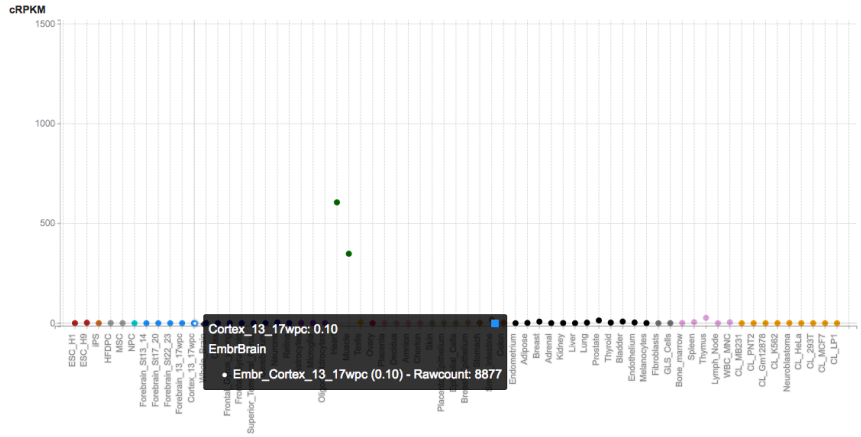


Figure R.5.8. Titin expression across human tissues and cell-lines. Taken from vastDB.

Rab6a is a Rab GTPase that contains a mutually exclusive exon (A=HsaEX0051725, A'=HsaEX0051726), which are nearly identical except for three amino acids and are alternatively spliced during pre-implantation development (Supp. Data 5.2). Isoform A and A' have the same GTP-binding properties and are localized to the golgi. A' however, cannot induce redistribution of golgi proteins into the endoplasmic reticulum, suggesting it cannot stimulate golgi-to-endoplasmic retrograde transport. In line with this, A' cannot interact with Rabkinesin-6, a golgi-associated Rab6A effector. Specifically the position 87 (either T or A) appears to be responsible for the difference between isoforms (Echard et al. 2000).

In the embryo, Rab6a knock-down in mouse oocytes is associated with 'severe disorganisation of the endoplasmic reticulum', changes to intracellular calcium levels and absence of 'corticle granule free domains' (Ma et al. 2016), which are involved in corticle granule translocations that prevents polyspermy (Cheeseman et al. 2016)(see p.5). In addition, is has been shown that this protein is also critical for meiosis and its knock-down affects actin-cap formation and failure of other cytoskeleton structures that lead to spindle defects and chromosome misalignment phenotypes (Ma et al. 2016). Therefore there are multiple testable phenotypes that could be used to explore the function of the two isoforms in the embryo.

Dynamin 2 is a GTP-binding protein that associates with microtubules and known to be involved in endocytosis, cell motility and cytokinesis in mitosis (Altschuler et al. 1998). All isoforms have been shown to localize to clathrin-coated pits at the plasma membrane (H. Cao, Garcia, and McNiven 1998). A mutually exclusive 139bp exon in the "Dynamin_M" domain, was found to change in early development (e.g. Mouse: MmuEX0015305; MmuEX0015299). Inclusion of MmuEX0015299 leads to localisation at the golgi (Y.-W. Liu et al. 2008). Dynamin 2 has been well studied in the early embryo, inhibition in 2 to 8-cell mouse embryos stalled the cell cycle and prevented compaction (Q C Wang et al. 2015) and it has also been shown to have a role in regulating the polar–body extrusion during meiosis (Qiao Chu Wang et al. 2014). In sperm, DNM1 and 2 are specifically localised to the acrosome heads and suggested to be involved in sperm exocytosis (W. Zhou et al. 2017).

In the three species there is a preferential inclusion of MmuEX0015305 (and equivalents in human/cow) at around genome activation stage, and in human only, skipping at the blastocyst stage. This conserved shift is curious, and could relate to the many exocytosis mechanisms at play in the early embryo, or even in the release of cortical granules, with a change in preferential isoform required for the change in cell cycle dynamics in the early embryo. It is unclear to which functions these isoforms have using this data.

METHODS

Data assembly

Publically available RNA-Seq samples from oocyte, 2-cell, 4-cell, 8-cell, 16-cell/morula, blastocyst, were downloaded from the "sequence read archive" (SRA). All data up to 09/9/2017 were considered. All samples used are found in Supp. Table 1 (Tab 1,2,3: Human [Hsa], Mouse [Mmu], Cow [Bta]). This comprises four, six and three experiments with 234, 345 and 123 single cells/samples in human, mouse and cow respectively.

Measuring gene expression

To measure gene expression (GE) of each gene in each cell/pool, we used *vast-tools* [align module](Irimia et al. 2014), which provides a normalized count measure for each gene (cRPKM, reads per kilobase of gene per million mapped reads, corrected by mappability; see (Labbé et al. 2012)) and corrected gene counts. The output expression tables from vast-tools we call a "cRPKM" tables and a "COUNTS" tables.

Calculation of exon/intron inclusion

To calculate an exon's **PSI** (percent-spliced-in) value (From 0% to 100% inclusion) for alternative exons or alternative donor/acceptor sites and **PIR** (percent intron retention; 0-100%) value for introns; we used vast-tools align (Irimia et al. 2014). Vast-tools relies on a database of known exon-exon and exon-intron junctions: VASTDB (vast-tools database), of which we used the version from 2017. Vast-tools maps RNA-Seq reads to all junctions, and takes into account complex events, where multiple alternative exons can be in tandem (Supp. Figure 1.1.ii). For a sample/event to have a PSI/PIR value, we set a minimum coverage filter of "VLOW" (in vast-tools). For alternative exons and intron retention, "VLOW" means that we must have at least 10 raw reads (before mappability correction- see previous section) must be mapped to the exclusion junction, or at least 10 and at least 5 reads mapped to the two inclusion junctions, else the value is "NA". For alternative 3/5', we required at least 15 reads on any junction ("VLOW" in vast-tools), else the value is "NA". Further, for

intron retention we filtered events to ensure the binomial p-value for imbalance between inclusion junctions was >0.05 (not variable across junctions), to ensure a single intron was being measured per event. A simplified calculation of PSI and PIR is shown in Figure R.1.1.b. The output of vast-tools align is a table of PSI/PIR values, which we call an "INCLUSION" table.

De-novo exon-skipping events

To ensure all exons were present in the *vast-tools* database, we conducted a de novo search for alternative cassette exons, using our early developmental RNA-Seq (Supplementary Table 1, Early development data only). Vast-tools libraries were not explicitly based from RNA-Seq from all these early stages, therefore we expected to find some novel exons. To make an additional vast-tools library, all RNA-Seq samples were mapped to their respective genomes (Builds: mm9, hg19, bostau6) using tophat2 (Trapnell et al. 2012) to gain mapped reads (in BAM format), then run through cufflinks (Trapnell et al. 2012) to attain gene transfer format (GTF) files for each sample. These GTFs were then merged using Cuffmerge (Trapnell et al. 2011) for each species and processed using SUPPA (Alamancos et al. 2014), to find alternative exon-exon junctions. Custom scripts (available upon request), were used to build a new vast-tools library, with novel exon skipping events added. Only exons found within the gene body (i.e. were not the first/last exons), and had at least one annotated ensemble exon as a donor/acceptor were added to the final list of novel exons.

Using this approach, we found 508 novel exons in Human, 267 in Mouse and 180 in Cow, not previously found in newest Ensembl assembly (hg38, mm10, bostau8). These are listed in exon information tables (Supplementary data 1; Column AN: Novel_to_hg19). These exons may have been missed previously due to their embryo specificity and show that even well studied genomes may lack some genetic features.

Calculation of alternative splicing

Given this thesis mainly relied on low coverage single cells with few replicates, and different numbers of replicate per stage, conventional statistics could not be used. Therefore custom scripts (available upon request) were used to calculate differential splicing between stages, following specific rules (see Figure R.1.3.a). In general, the

rules are quite strict. If the PSI range was large, we required a larger mean PSI difference between stages (Figure R.1.3.a). We also required PSI difference (dPSI) of at least 20 and a range difference of -10 (can overlap by 10 PSI values), and coverage in a minimum of two replicates (for human and mouse). The range difference removed many events where, for example, one sample had 100% inclusion and another 0%. However, this could either be real variance (single cell level) or noise, but we decided to filter these events, as they are likely to represent noise. Using this range filter also means if a stage is composed of cells from various slightly different timings (an issue of combining multiple RNA-Seq experiments), an event could be considered too variable in range to be counted as alternatively spliced between stages. Some of these issued are discussed on page 132.

Cell identity and initial clustering

Gene expression values were clustered using hierarchical clustering (Supp. Figure 1.1.i;), performed in R. Cell/samples clusters were used for the merging of data (next section). Clustering also revealed individual cells that did not fit into its expected stage, and may represent dying, damaged or mislabelled cells. These were removed from further study (Supplementary Table 1 for details). For principle components analysis see p.157.

Merging of samples

Given single cells often have low coverage across exon-exon junctions, we used the clustering to create pools representative of a particular stage (using vast-tools 'merge' module), with the aim to recover as many samples (pools of single cells/samples) with >150 million reads from each stage (not always possible), these pools are listed in Supplementary Table 1 (Merged group column). High coverage was needed to confidently assign levels of inclusion or exclusion (of exons/introns). In all cases, cells from the same embryo were kept in the same pool, and where two embryos need to be merged this was based on the hierarchical clustering results.

Principle components analysis

Principal Component Analysis (PCA) was conducted in R and custom scripts (SCRIPTS: PCA_from_inclusion.pl), which use the function 'princomp' to calculate components from an input counts or a vast-tools inclusion table. For alternative exons, all AS events with sufficient read coverage (VLOW or higher) in >80% of samples and standard deviation of 5 were considered. Gene expression measurements took all genes as input, with standard deviation greater than 5 across samples.

Differentially expressed genes

To call a gene 'differentially expressed' between each pair-wise transition we used fold change values between the mean cRPKM values for each stage (mean of the merged groups). We filtered off lowly expressed genes, those with a mean cRPKM of less than 2 in both stages. Then for genes with mean expression above 10 cRPKM (in one stage), a fold change of >=2 was required to be called differential, and for genes with mean expression lower than 10cRPKM, a value had to be below 1 cRPKM and above or equal to 5 in the other stage to be considered differential. This simplistic measure was used due to the difficulty to measure gene expression changes with samples from multiple independent experiments. Yet we acknowledge that this will not be as sensitive as a conventional differential expression program (such as DESeq2).

Transcriptional complexity

Short read RNA-Seq does not allow the absolute quantifications of gene isoform types, yet using the exon-centric percent inclusion data we could create numerous measures of isoform complexity, based on the number of exons that exhibit one or two forms. Using mean PSI values for each early development stage (e.g. 2-cell or Morula) and mean values for differentiated tissues. Data for tissues are taken from Supplementary table S1 from (Tapial et al. 2017), using a subset shown in Supplementary data 7. We could assign a score of 1 if an exon is found to be present in two inclusion forms (for example, a PSI between 20 and 80, meaning an exclusion and inclusion isoform exist) or 0 if the exon is found to have one main isoform (PSI <20 or >80).

A set of custom scripts (SCRIPT/Chapter1/Transcriptional_complexity/:) were used to plot the fraction of events ('P-mode') that show two forms over the total number

events checked with coverage (VLOW,LOW, OK, SOK; can be set in script) across a specified fraction of all samples (e.g. 0.5 : will only count events with coverage across 50% of input samples). We could also change the type of events surveyed (Exon skipping, Intron retention, Alt 5/3' SS), and range of PSI to check (e.g. 5-95: events considered two isoforms if between PSI 5 and 95). In addition, scripts can be run in 'T' mode, which calculates the total number of events with two potential isoforms.

Mfuzz clusters of exon skipping

Given median PSI values for each stage from oocyte to blastocyst (Human/Mouse) and Gv-oocyte to Blastocyst (Cow), we could clusters the events using Mfuzz (Kumar and E Futschik 2007) for events with zero NA values (coverage across the time course). Default settings were used and numbers of clusters were determined by using the using the Dmin function. This function calculates the minimum distance between cluster centroids, to show if splitting your data into more/less clusters will better represent the diversity of your sample. Here, I chose the number of clusters where an additional cluster would not noticeably reduce the minimum distance (as suggested in (Futschik and Carlisle 2005)). Next, I imputed the missing values in events with two or less "NA"s (one or less for cow) using a custom script (SCRIPT/Chapter1: Time_course_PSI_impute.4). This script imputed values based on the assumption of no change, i.e. if NA values were at either end of the time course then the imputed value would be the closest known value, and for missing values between known values the average was taken for the missing point/s. These events could then be assigned to an Mfuzz cluster using the 'Mfuzz: membership' function. Membership values for each event (with coverage), are found in Supplementary data 1 (Column: Cluster). In addition, Mfuzz clusters were made for mean gene expression values, in the same way that has been described for alternative exons (plots and gene to cluster assignments in Supplementary data 3.GeneExpression/Mfuzz).

To classify the profiles (expression and exon skipping Mfuzz clusters) into Peak/Shift or Up/Down, I used custom scripts (SCRIPT/Chapter5: FINAL_TABLE_to_Mfuzz_Characteristics.8.pl/ FINAL_EXPR_to_cluster_Characteristics.8.pl). These split the values into upper, middle and lower values for each stage (given (Max-Mix)/3) and scored dependent on

the pattern of change across the time course. If a profile had 2+ 'middle' values it was given the status: OTHER_M or if first or last stages changed between upper/lower thirds its was given the name: OTHER_E (event changing at end, cannot assign peak/shift). If a profile started and ended within the same third it was a putative peak, else a putative shift. For putative peaks, if the number of stages in the top thirds was greater/equal to two times the number in the low third the profile was peak-down (P_Dw), and the opposite for peak-up (P_Up). For putative shifts, if the biggest change in PSI divided by the range in PSI was less than 0.4, we considered the profile to exhibit gradual change: OTHER_G. For the remainder, if the first or last values were in the mid third they were categorized at OTHER_MID. Finally, depending on the direction from upper to lower (or reverse direction), we could assign shift up (S_Up) or shift down (S_Dw). To calculate proportions of alternative exons in each category, only events with coverage in at least 6 time points were considered (5 for cow).

Enrichment of differentially expressed genes with splicing events

Using the number of differentially expressed genes at specific transition (see Methods: Differentially expressed genes) divided by the total number of genes tested gives a proportion of genes that are differential. Using this we could compare the actual numbers of differentially expressed genes with splicing events compared to what is expected by chance. Using these proportions in 'prop.test' (R; (Newcombe 1998)), p values (two sided, with Yates correction) could be assigned for enriched and depleted lists, given observed events in differentially expressed genes (Up and Down-regulated). For genome activation in the three species, I further separated the events in those with greater inclusion, or greater skipping, as well as separating by type: exon skipping or intron retention. A full table of expected/observed proportion and pvalues are found in Supplementary table 1.5.

Open reading frame (ORF) analysis

The expected impact of the ORF was calculated as in (Irimia et al. 2014).
Comparisons with Mfuzz clusters (Figure R.2.1.b), focused only on those changing at genome activation (Meth Table. 1), plotting only those events that were labelled as protein isoform, disruptive upon inclusion or disruption upon exclusion. PSI values in

differentiated tissues was calculated from Supp. Table 1, for the same three types of ORF prediction together.

Methods Table 1. Clusters considered changing at genome activation. Up/Down in stage of ZGA or stage after.

Hsa	Mmu	Cow
1	1	1
2	2	2
3	3	3
7	4	4
8	5	7
9	6	8
10	7	9
11	8	10
12	9	11
13	14	13
14	15	14
15	16	15
16	18	16
19		17
20		19
21		20
22		21
23		22
24		
25		
27		
28		

ORF to DEG comparison

A custom script (ORF_to_expr.pl) was used to compare differentially expressed genes for ZGA stage in the three species, split into protein isoforms, disruption upon inclusion/exclusion. Prediction of numbers of events within up and down regulated genes was calculated by taking the number of UP/DOWN DEGs divided by the intersect of genes with coverage for the differential gene expression analysis and the genes with events with coverage for the splicing analysis.

TRiP-Seq analysis

RNA-Seq was downloaded from (Blair et al. 2017). We used vast-tools to obtain PSI values, then measured changes in PSI (dPSI) between the cytosolic and high-polysome fraction (using vast-tools diff). For each category of alternative exons changing we could plot the dPSI as a box-whisker.

Gene Ontology Analysis

GO term significance was calculated using 'TotalGO' (available upon request), a novel script made to calculate binomial p.values given a target and background list of events/genes. Splicing events were converted to ENSEMBL gene IDs, and run in the same way as gene expression target lists (except splicing events in the same gene would only be counted once). The background (splicing lists only) was determined by listing the events that had adequate coverage in the comparison being made (see Calculation of alternative splicing). For example, when running gene ontology analysis for events changing between 2-cell and 4-cell, the background included all events with sufficient coverage in both stages (two replicates for human/mouse, 1 for cow). TotalGO has the advantage of being a one-line command program, with multiple options for filtering and plotting, and can accept gene/vast-tools ID lists with multiple categories/time-course data. It can print both a histogram of the top terms or if a category list, a heatmap of terms.

Motif enrichment

For known motifs, the CISBP database of predicted RBP binding sites (Ray et al. 2013) and all kmers (length 2 and 3) were used within the program 'matt' (Gohr and Irimia 2019) to search over a sliding window the frequency hits across the intron regions around the alternative exons and with the exon itself. The backgrounds in all cases were those alternative splicing events not in the target cluster/transition under study (Supplementary Data 4. RBP Maps, showing Clusters, Transitions and also enrichment of specific motifs: MBNL and ESRP).

For novel motifs, custom scripts were made to extract genome DNA sequences around the alternative exons (at each transition and for each Mfuzz cluster). 500 bp regions were extracted in the C1, up-stream, down-stream and C2 introns, with a 3bp overlap of the exonic spice junction. If a sequence was longer than the size of the intron, the sequence was shortened to the region representing the intron only. In

addition, exonic regions were also extracted, with 3bp overlapping the intron taken from either the 3' direction or 5', restricted to 50bp or shorter if the exon was less than 50bp. 'gffread' (John Hopkin Universitys, open source) was used to make the fasta files for each region (using an unmasked genome file). Meme (Bailey and Elkan 1994) was used to measure enrichment of novel <50nt sequences within the target set (Supplementary Table 3.2, Three species, transition and clusters). For the masked novel motif search 'pre-masked' genomes were downloaded from UCSC (Golden Path, bigZips). MEME (long novel motifs) and DREME (short motif) software was used to search for novel motifs in the sequences, full results in Supplementary Data 6.

Transposable element enrichment

To detect transposable elements around alternative exons, I used custom scripts (Supp. Scripts Chapter 3). Using the coordinates of all transposable elements for HMC, I used a script (EXON_info_to_TEcounts.repName or family .pl) to count the number of TEs that overlap a 500 nt region of the up/down stream intron around each event and within the exon itself, starting from Exon information tables (Supp. Data 1). This script can be adjusted to check differing length regions around the events of interest or restrict the search to specific families of transposable elements. TE_pivotplot.new.pl was used to make barplots of the proportion of TEs above and below the average (across all categories), with features in the EXON INFORMATION table (Supp. Data 1), such as the Mfuzz cluster ID, protein prediction or consecutive transition.

RBP knock down and DNA damage analysis

To compare RBP knockdown PSI change, we downloaded publically available RNA-Seq for as many RBPs (human and mouse), with compatible reads for both knockdown and control. The database was built by Sophie Bonnal (Valcarcel lab, CRG), who conducted a literate search to compile as many results as possible. With the merged early development stages, and RBPs (Supp. table. 1) I could merge their "INCLUSION tables" using a custom script (Supp. Scripts: MergeINC.pl), then create dPSI tables (Supp. Scripts: RBP_enrichment_setUP.pl), using settings of 10dPSI for mean difference and -10dPSI for range difference. Finally using custom script (Supp. Scripts: RBP_enrichment_PLOT) I could plot dPSIs of every sample versus each

other sample's dPSI (Supp. Figures X). Binomial p.values were given for significance of four quadrants in the plots and correlation coefficients.

For DNA damage treated cells (Supp. Table 1), I could use the same scripts to compare the events increasing in inclusion and skipping between control and treatment, with the changes at each transition in early development. To create the dPSI tables, I used a cut-off of minimum 5dPSI for range and 15dPSI for difference between means (number of samples >=1). Plots are in Supp. Figure (folder) 4.6 for all comparisons, not in thesis proper due to size constraints.

Conservation

To find the pairwise conservation between exons, we used lift-over (Hinrichs et al. 2006) to find orthologous syntenic regions in a second genome, we then used customs scripts (available upon request) to ensure the lifted dinucleotide splice junctions were (AG….GT) in the second species and matched these to Vast-tool IDs (e.g. MmuEX/ HsaEX in the second species where possible.
Several issues had to be solved before these keys could be used. One was that some exons could be lifted in one direction (say Human to Mouse), but would not be lifted in the reverse direction (Mouse to Human), in these cases the reciprocal entry was copied to the other conservation key. Second, a single exon could sometimes be lifted to multiple exons in the second species, this often could be biologically relevant, where an exon has undergone tandem duplication, creating two truly orthologous exons in the second species, whereas others appeared erroneous, where the two lifted exons had a different lengths, one matching that of the first species and the other not, for all these cases they were removed from the conservation key, to remove any speculative homologies, but these represented less than 1% in the comparisons.

Blastocyst Identity

To confirm lineage fate of blastocyst cells in each species, a more complex clustering was required to split cells into either trophectoderm (TE) or inner cell mass (ICM) and further into epiblast (EPI), primitive endoderm (PrE) or TE. This was performed first by clustering by gene expression of known transcription factor markers previously published to separate the cells (Yan et al. 2013; Guoji Guo et al. 2010).

This split the cells into the groups with strong expression of lineage markers and some cells that could not be definitively identified into one of the categories (Supp. Figure 6.1. A: Mouse, B: Human). In this way, we can compare equivalent cell types between species. The cells which could not be assigned may be due to technological biases (Bhargava et al. 2014), as mouse appear much more noisy (Supp. Figure 6.1. A), probably due to lower coverage than the human samples. Alternatively, these cells may be too early in development to show differences or could be due to apoptosis in particular cells.

Lineage specific exons

Principle component analysis was conducted as previously described, using single cells from the blastomeres of human and mouse. For alternative exons, a minimum of 50% of each event has to have coverage (VLOW), then imputation of missing values. Given the single cells have low coverage; we required lower % coverage to have sufficient events to cluster the cells with.

To find events differences between the lineages, we used single cells as replicates, and a strict (n=2+ replicates; with same rules as for human/mouse transitions) and relaxed approach (n=2+, but no range filter was used). Scripts: Vast-tools-diff_strict + relaxed.pl). For figure R.6.1.b, Hypergeometric p.values were calculated in R, using the number of events that had coverage in both experiments as a background (listed in figure R.6.1.b legend).

ESC-differential analysis

To measure ESC-specific exons in differentiated and our early-development time course, we first combined our inclusion table (vast-tools exon to PSI table) to that of multiple differentiated tissues (DATA 7. INCLUSION tables and configs). We could estimate the numbers of ESC and early-dev specific exons by using a custom script (get_tisAS_v5.pl; available upon request [written by Manuel Irimia]). This set removed novel exons predicted using the novel-exon discovery pipeline, as the differentiated set were not mapped against the novel junctions (these only comprise a few small fraction, so we do not expect them to change the result significantly). ESC specific exons could then be compared to all other tissue types by custom script (available upon request), to make scatter plots in figure R.6.2.a .

Supplementary Methods (Barbara Pernaute- BP)

Embryo collection (BP)

All embryos were obtained from B6CBA F1 crosses. Zygotes were collected from the oviduct of superovulated females 20h post HCG injection and cultured in KSOM media (Millipore) at 37^{o}C, 5%CO2 up to 2 cell, 4 cell, 8 cell, morula or blastocyst stage.

Snrpb and Snrpd2 overexpression and etoposide treatment (BP)

The pCS2+8NmCherry vector (Addgene) was used for in vitro transcription of mCherry mRNA. Snrpb and Snrpd2 cDNA were cloned into a modified pCS2+8NmCherry vector lacking mCherry tag. In vitro transcription was performed using the mMESSAGE mMACHINE® SP6 Transcription Kit (Ambion) according to manufacturers' instructions.
For all overexpression experiments 1 cell embryos were microinjected with 300ng/ul of mCherry mRNA (control) or 150ng/ul of Snrpb mRNA and 150ng/ul Snrpd2 mRNA. Injected one cell embryos were left in culture and either treated with 10uM etoposide for 1.5h or left untreated. Following the treatment embryos were fixed for immunostaining as described elsewhere.

Immunostaining (BP)

Embryos were fixed in 4%PFA/PBS for 10 min. Following fixation they were permeabilised in 0.5% Triton X-100 for 15min, blocked in 10%BSA/0,1% Triton X-100/PBS and incubated overnight in primary antibody: anti-Snrpb (Thermo Fisher), anti-Snrpd2 (Thermo Fisher), anti-phospho H2AX Ser 139 (Cell Signaling), anti-phospho p53 Ser15 (Cell Signaling). Hoechst was used for nuclear staining. Imaging was conducted in a Leica SP5 confocal and images processed with Fiji software. For quantification, relative intensity represents the mean fluorescent intensity of the nucleus relative to nucleus area, measured with Fiji. Each dot represents average

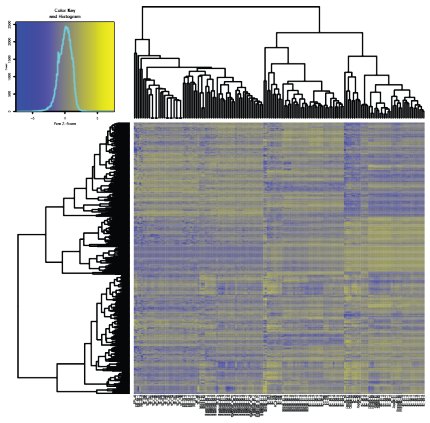
relative intensity of the 2 cells of the embryo. For etoposide treatment the plots represent quantification of embryos from 3 independent experiments. Statistical significance was tested by applying a Wilcoxon signed rank test.

RNA extraction (BP)

RT-PCR for alternative splicing event validations were performed using pools of embryos at different developmental stages obtained as described above. The Arcturus Pico Pure RNA extraction kit (Thermo Fisher) was used for RNA extraction and cDNA was transcribed with Superscript III Reverse Transcriptase (Thermo Fisher).
For RNA-sequencing experiments following Snrpb and Snrpd2 overexpression, one cell embryos were microinjected as described above and collected for RNA extraction at either 5h post injection (zygote stage) or 24h post injection (2 cell stage). For each condition 40 embryos coming from 3 independent experiments were pooled to extract RNA for sequencing. RNA was extracted using the Qiagen RNeasy Micro Kit. SMARTer Stranded RNA-Seq Kit was used for library preparation prior to Illumina sequencing.

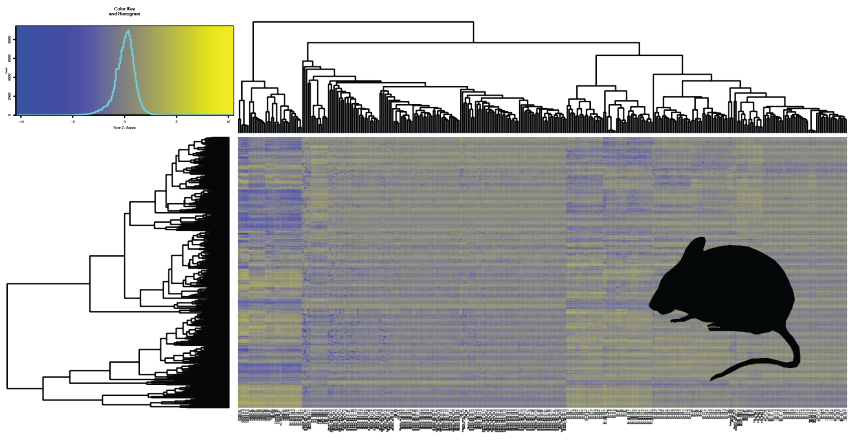
SUPPLEMENTARY FIGURES

Supplementary Figure 1.1.i: Early development heatmaps. Showing clustering of all samples, using the top 500 most variable genes as determined by edgeR. These trees were used to determine cell identity and locate anomalous cells, details of cells removed in Supplementary table 1 (Column J, red coloured text). Figures are too large to print (B5 or A4), therefore the full size figures are found on external disk (Section), overview below.



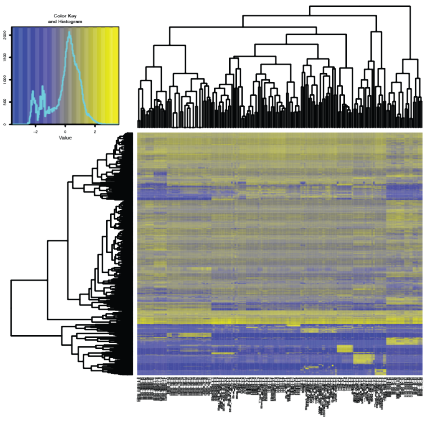
GENES

SAMPLES



GENES

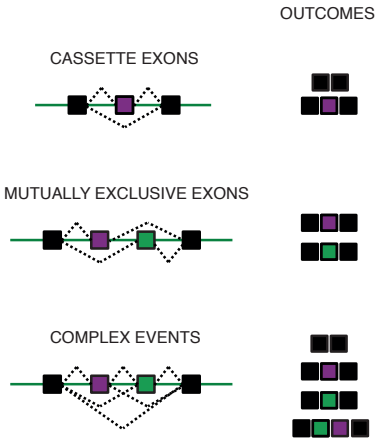
SAMPLES



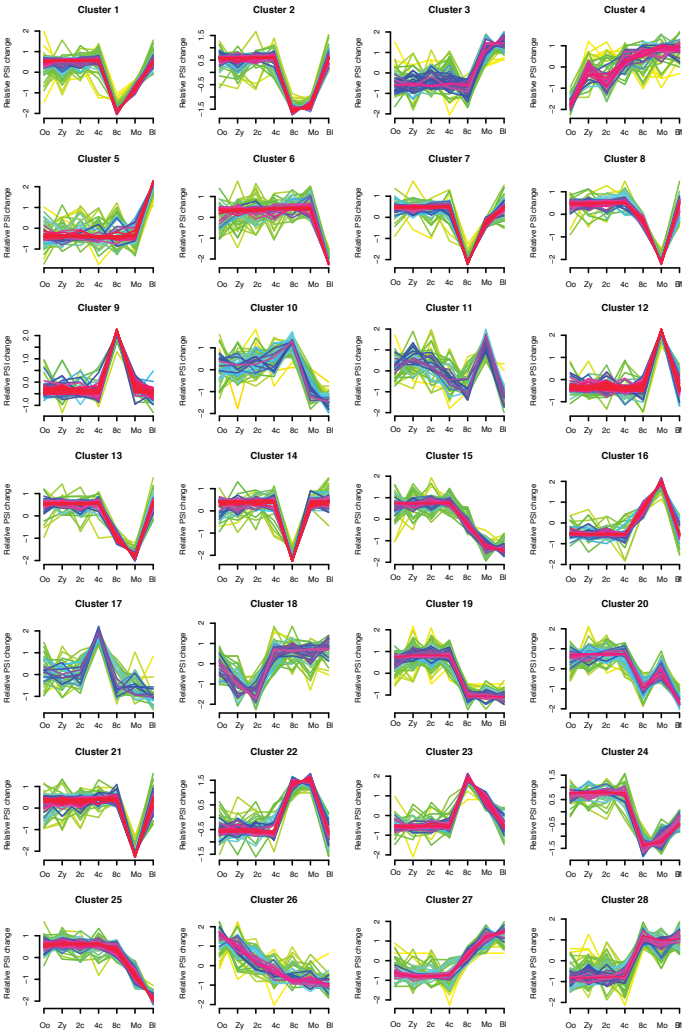
GENES

SAMPLES

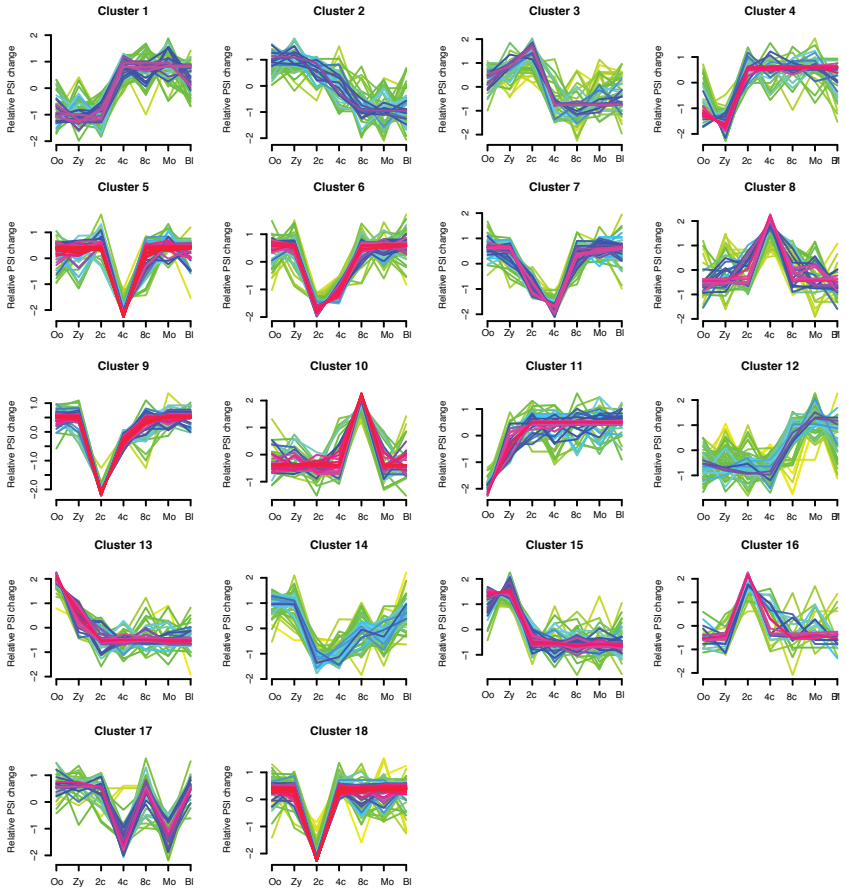
Supplementary Figure 1.1.ii. Complex splicing type. Top shows a simplified scenario with three exons and an alternative one in the middle, which can be result in two isoforms. Middle: shows an example where two alternative exons can be included into the final transcript but never together, which we call mutually exclusive, resulting in the inclusion of either exon. Lower: shows a more complex example four different outcomes can be produced from these two alternative exons, both can be skipped together, or included together, or either exon could be included exclusively. Given events can often be complex and involved greater than two tandem alternative exons, this makes it hard to accurately assign percent spliced in (PSI) values for each event.



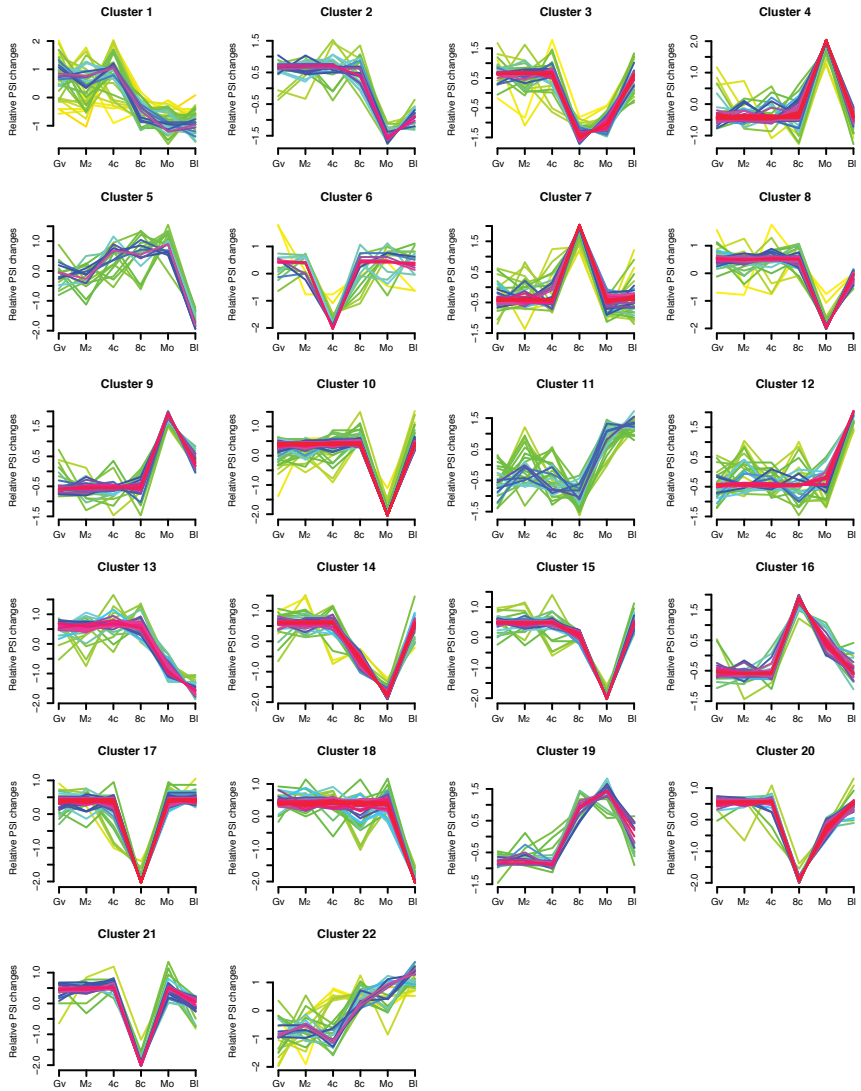
Supplementary Figure 1.2.i. (Part A- Human). Mfuzz clusters of exon skipping events, showing stages from oocyte (Oo), zygote (Zy), 2/4/8c (cleavage stage embryos), morula (Mo) and blastocyst (Bl). Y axis shows relative PSI change. Clusters made using Mfuzz (Kumar and E Futschik 2007)



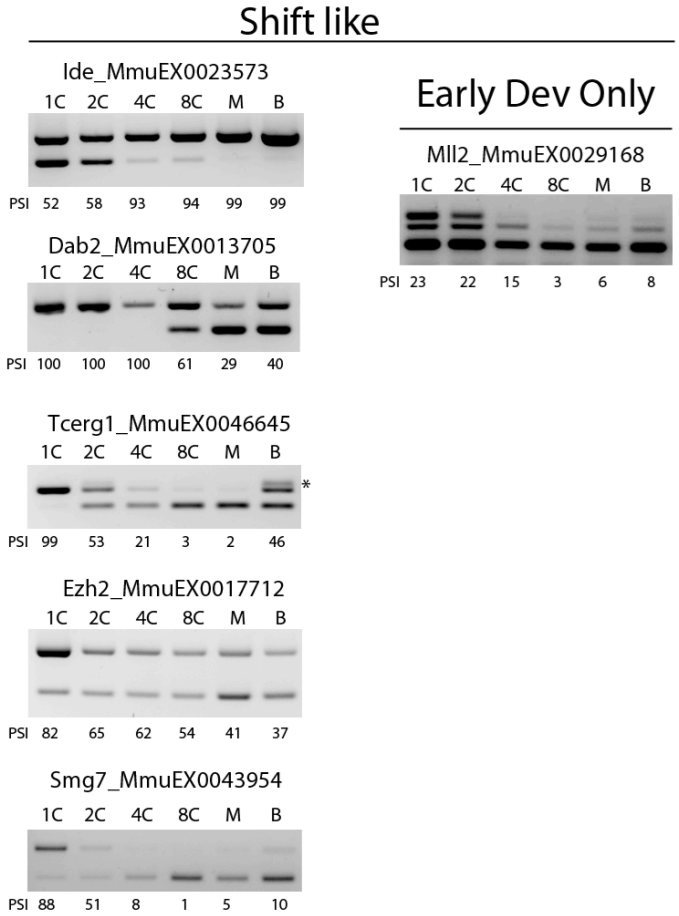
Supplementary Figure 1.2.i. (Part B- Mouse). Mfuzz clusters of exon skipping events, showing stages from oocyte (Oo), zygote (Zy), 2/4/8c (cleavage stage embryos), morula (Mo) and blastocyst (Bl). Y axis shows relative PSI change. Clusters made using Mfuzz (Kumar and E Futschik 2007)



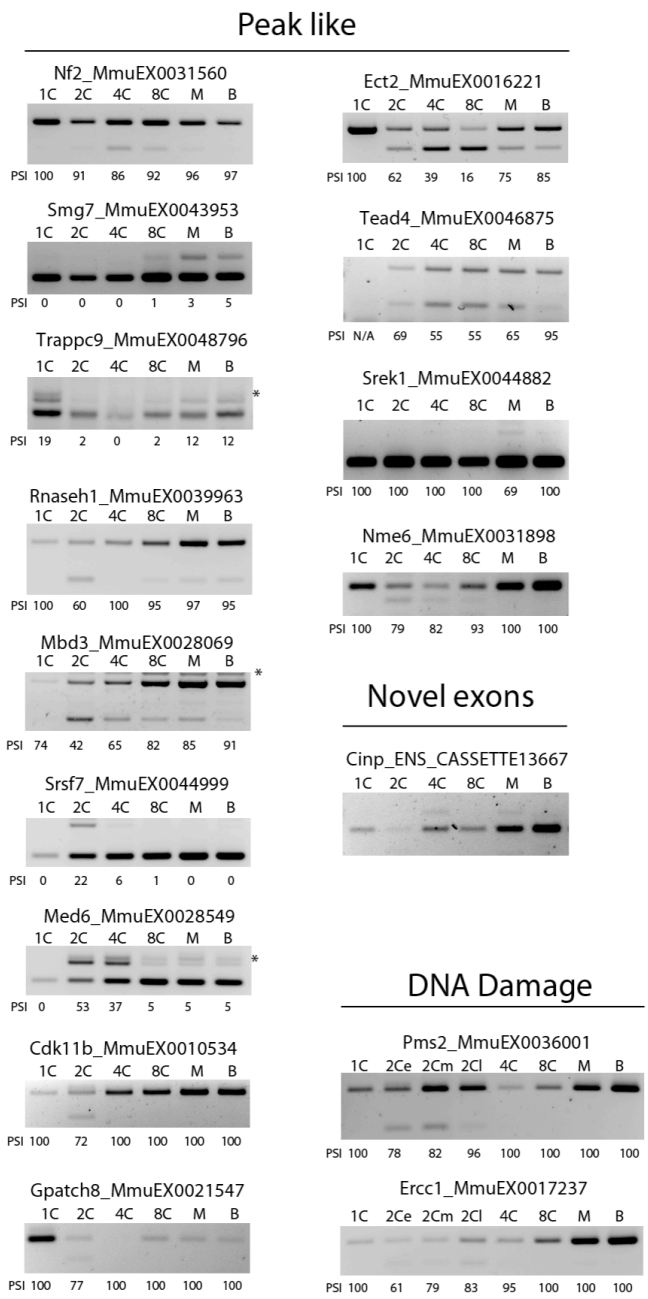
Supplementary Figure 1.2.i. (Part C- Cow). Mfuzz clusters of exon skipping events, showing stages from germinal vesicle oocyte (Gv), Metaphase II oocyte (M2), 4/8c (cleavage stage embryos), morula (Mo) and blastocyst (Bl). Y axis shows relative PSI change. Clusters made using Mfuzz (Kumar and E Futschik 2007)



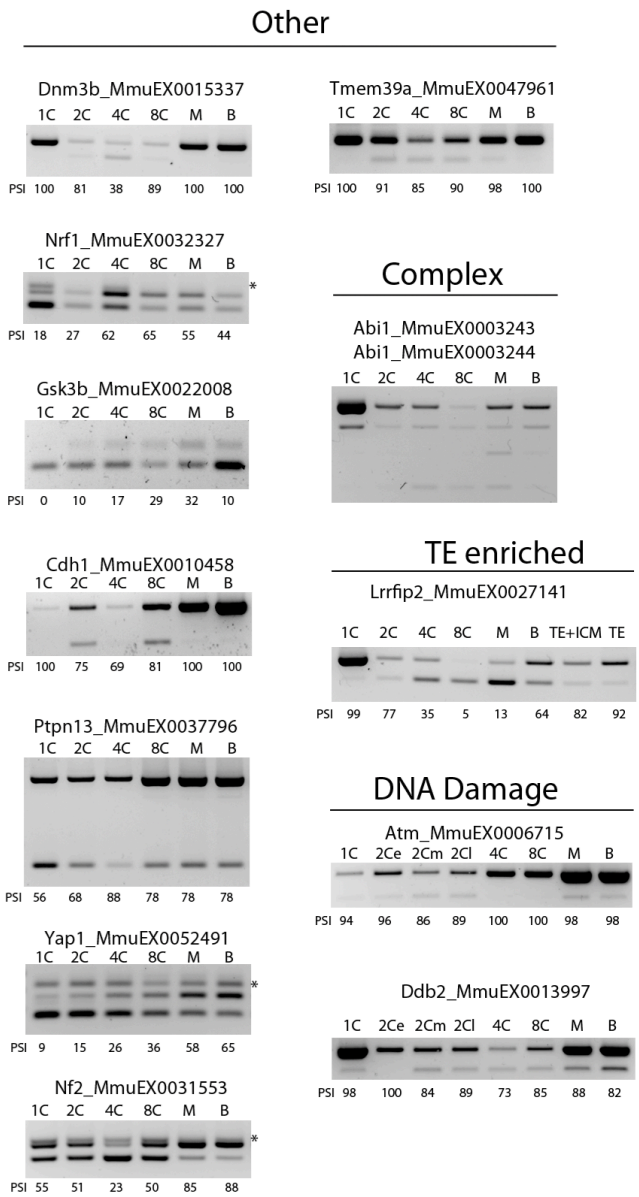
Supplementary Figure 1.2.ii (Part I): Validations of shift-like events in mouse. 1C=zygote, M=Morula, B=Blastocyst. Gene and vast-tools event names are shown above each validation. Upper band shows the inclusion isoform, longer PCR product, lower band is the short isoform. *Top band is PCR heteroduplex. Performed by Barbara Pernaute.



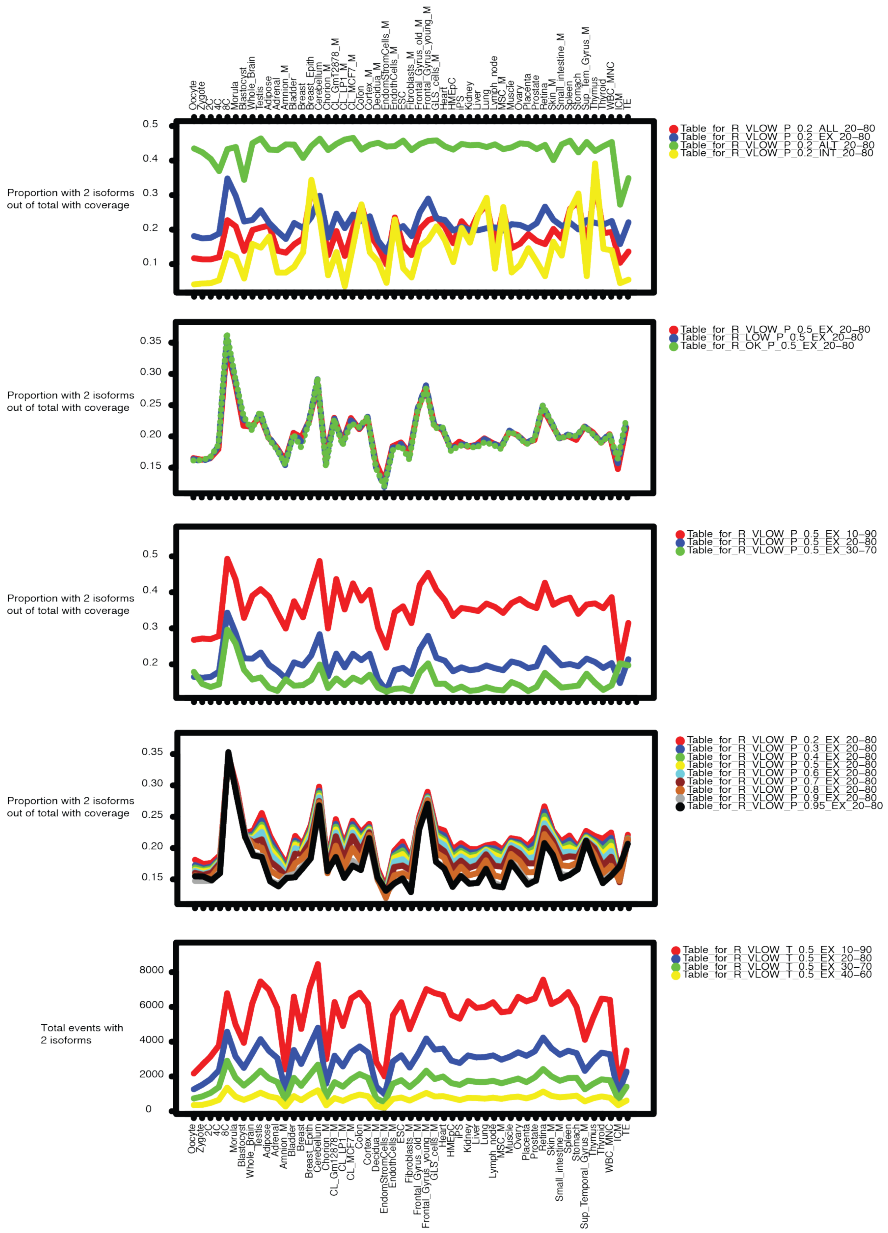
Supplementary Figure 1.2.ii (Part II): Validations of peak like events in mouse. 1C=zygote, M=Morula, B=Blastocyst. Gene and vast-tools event names are shown above each validation. Upper band shows the inclusion isoform, longer PCR product, lower band is the short isoform. *Top band is PCR heteroduplex. Performed by Barbara Pernaute.



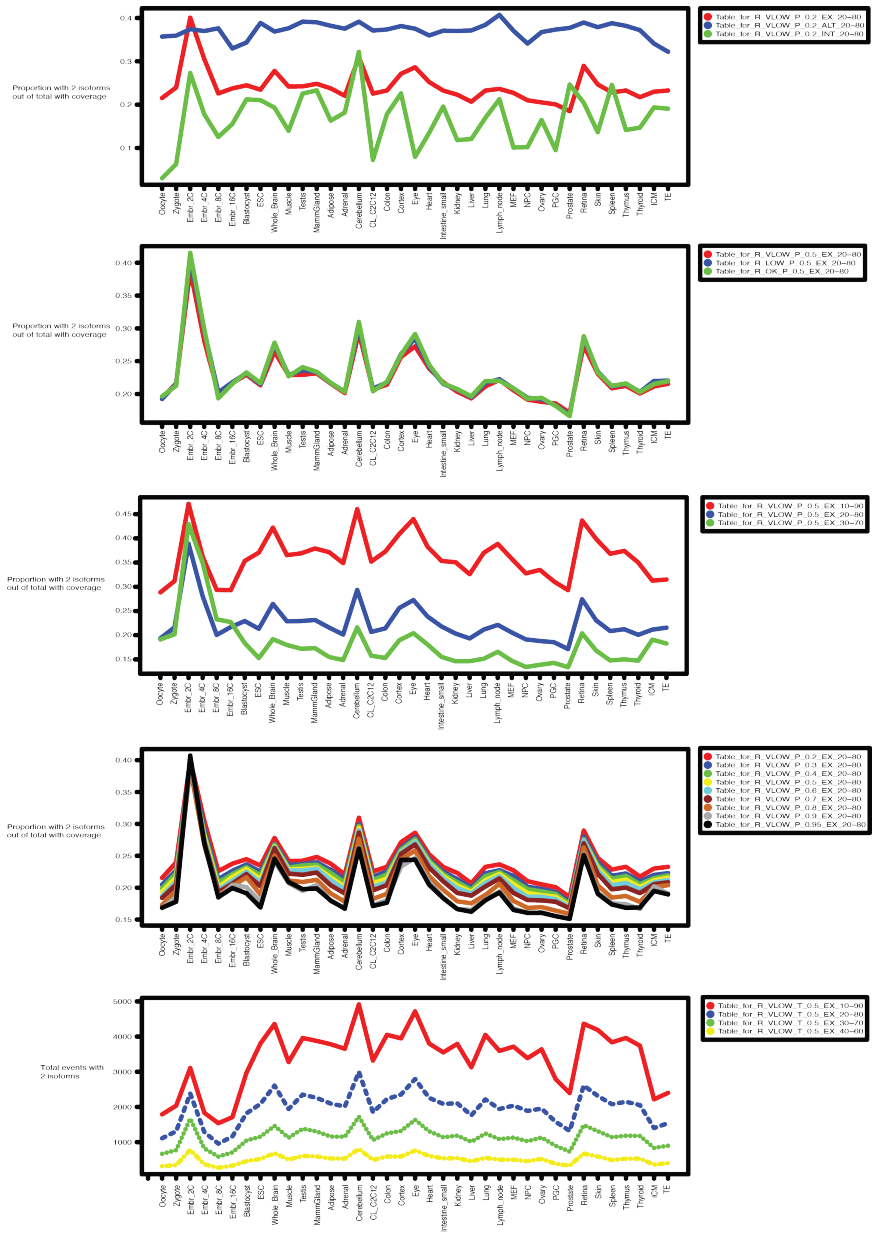
Supplementary Figure 1.2.ii (Part III): Validations of other-like (or complex), TE-enriched or DNA damage events (other-like). 1C=zygote, M=Morula, B=Blastocyst. Gene and vast-tools event names are shown above each validation. Upper band shows the inclusion isoform, longer PCR product, lower band is the short isoform. *Top band is PCR heteroduplex. Performed by Barbara Pernaute.



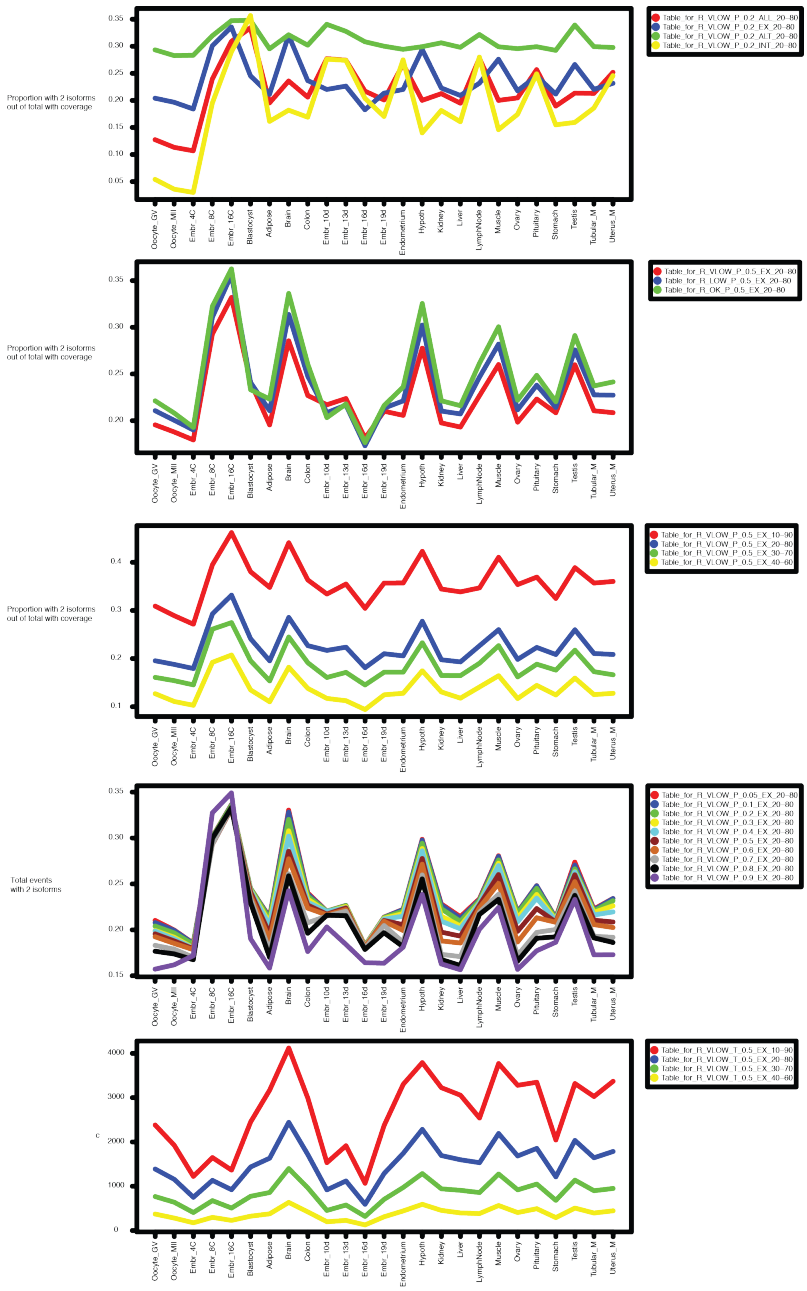
Supplementary Figure 1.3.i (Part A: Human- NEXT PAGE). Complexity measures across early development and differentiated tissues, with different settings to test the robustness that ZGA stage embryos have the most complex transcriptome. TOP: Complexity score (20-80 = 2 isoforms; coverage 80%) for alternative exons, alternative 3' 5' splice sites, intron retention and a combined score. UPPER MIDDLE: Complexity score (20-80 = 2 isoforms; coverage 50%; alt exons) for alternative exons changing the coverage required (VLOW, LOW, OK). MIDDLE: Complexity score (coverage 50%; alt exons) changing the range of PSI values considered to be two isoforms. LOWER MIDDLE: Complexity score (20-80 = 2 isoforms; alt exons) changing the percent with coverage expectation across all samples before making the score. BOTTOM: A different score showing the total number of events that support the existence of two isoforms, regardless of coverage.


Proportion with 2 isoforms out of total with coverage
Table_for_R_VLOW_P_0.0_ALL_20-80
Table_for_R_VLOW_P_0.0_EX_20-80
Table_for_R_VLOW_P_0.0_ALT_20-80
Table_for_R_VLOW_P_0.2_INT_20-80
Proportion with 2 isoforms out of total with coverage
Table_for_R_VLOW_P_0.5_EX_20-80
Table_for_R_VLOW_P_0.5_EX_20-80
Table_for_R_OK_P_0.5_EX_20-80
Proportion with 2 isoforms out of total with coverage
Table_for_R_VLOW_P_0.5_EX_10-90
Table_for_R_VLOW_P_0.5_EX_20-80
Table_for_R_VLOW_P_0.5_EX_30-70
Proportion with 2 isoforms out of total with coverage
Table_for_R_VLOW_P_0.0_EX_20-80
Table_for_R_VLOW_P_0.01_EX_20-80
Table_for_R_VLOW_P_0.04_EX_20-80
Table_for_R_VLOW_P_0.5_EX_20-80
Table_for_R_VLOW_P_0.6_EX_20-80
Table_for_R_VLOW_P_0.7_EX_20-80
Table_for_R_VLOW_P_0.8_EX_20-80
Table_for_R_VLOW_P_0.9_EX_20-80
Table_for_R_VLOW_P_0.95_EX_20-80
Total events with 2 isoforms
Table_for_R_VLOW_T_0.5_EX_10-90
Table_for_R_VLOW_T_0.5_EX_20-80
Table_for_R_VLOW_T_0.5_EX_30-70
Table_for_R_VLOW_T_0.5_EX_40-60

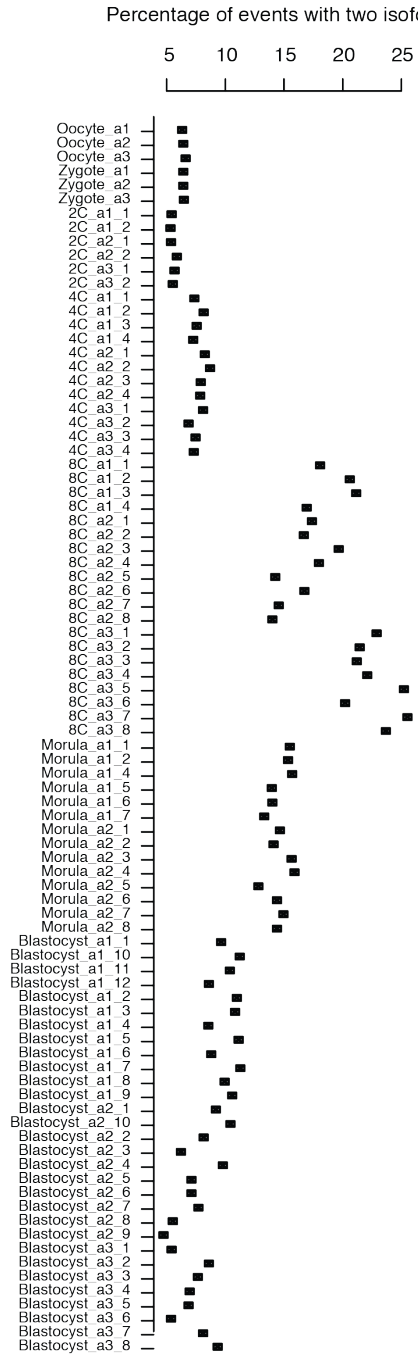
Supplementary Figure 1.3.i (Part B: Mouse). Complexity measures across early development and differentiated tissues, with different settings to test the robustness that ZGA stage embryos have the most complex transcriptome. Refer to Supplementary Figure 1.3.i (Part A) for scores.



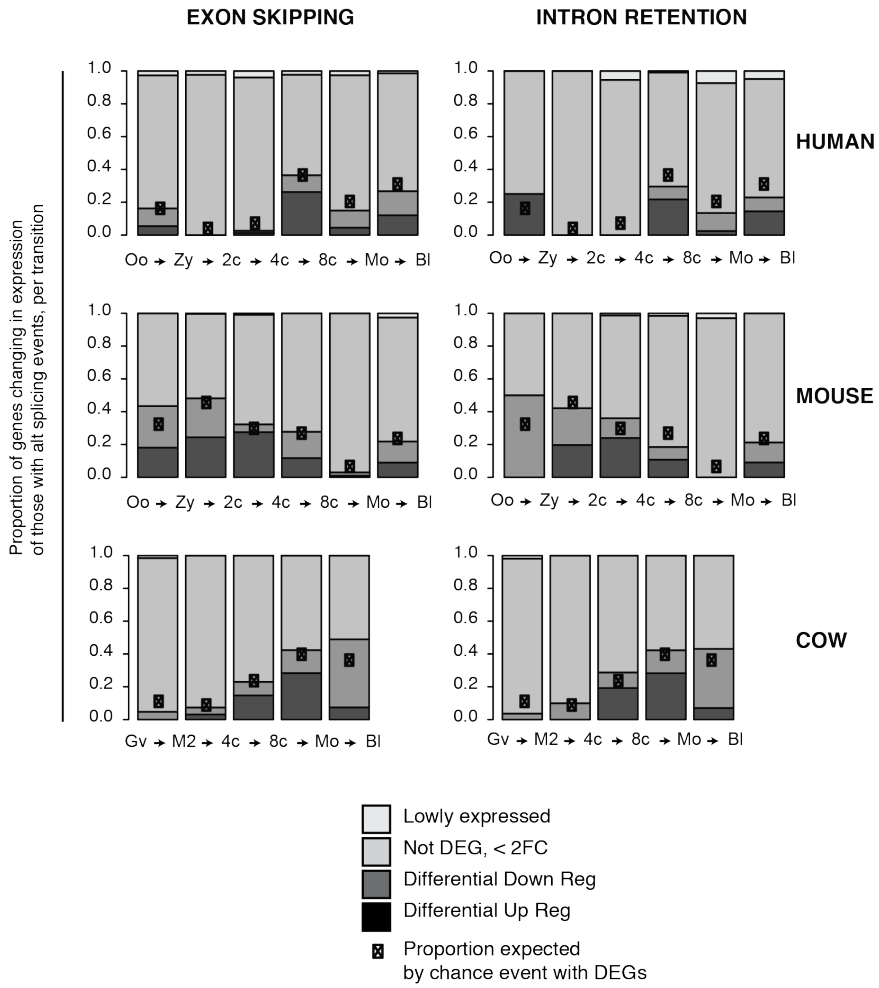
Supplementary Figure 1.3.i (Part C: Cow). Complexity measures across early development and differentiated tissues, with different settings to test the robustness that ZGA stage embryos have the most complex transcriptome. Refer to Supplementary Figure 1.3.i (Part A) for scores.



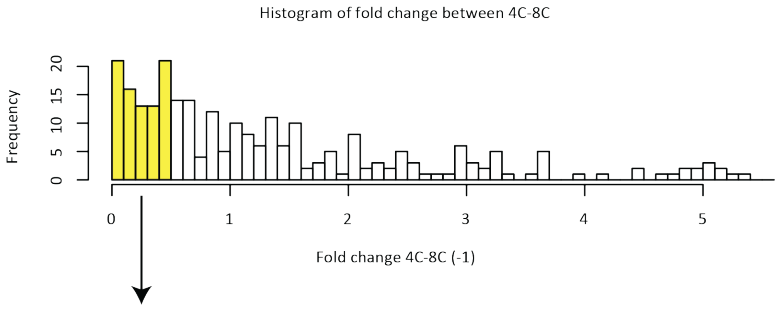
Supplementary Figure 1.3.ii. Transcriptomic complexity scores in human single-cells. Given 20-80 PSI=2 isoforms, <20/>80 = 1 isoform, and a minimum of 80% of the samples with coverage.



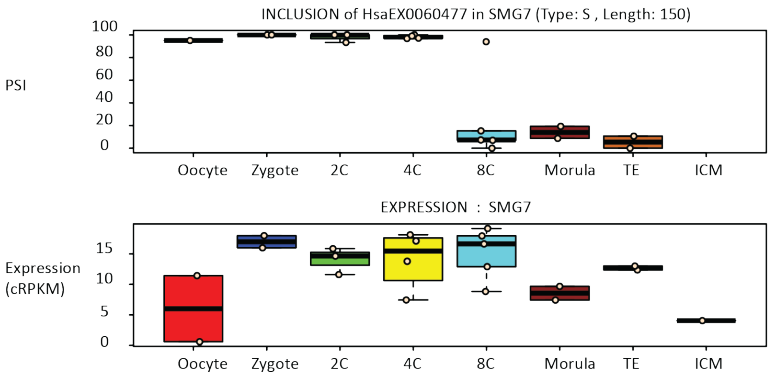
Supplementary Figure 1.4.i. Proportions of genes with events that change in expression (Up or Down), or are not differentially expressed (<2 fold change) or were not measured (often because of low expression). This is calculated for numbers of exon skipping events (left) and intron retention (right), for consecutive transition. The proportion expected by chance is shown.



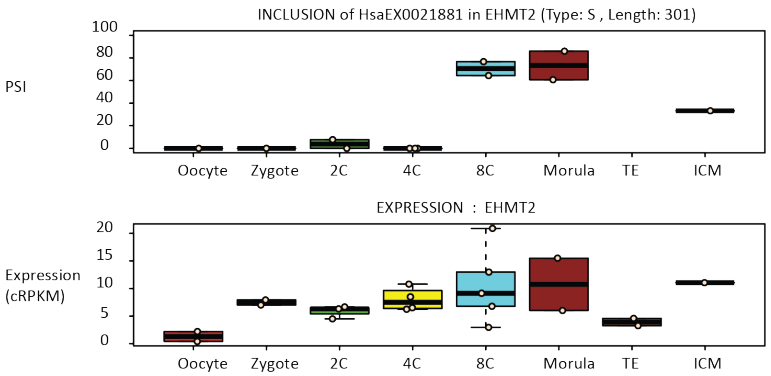
Supp. Figure 1.4.ii Exon skipping events in genes that show little to no change in expression. TOP: Shows the frequency of log fold changes minus 1 at genome activation (4C-8C) in human. BOTTOM: Two examples of exon skipping events, showing PSI changes and expression changes (cRPKM).



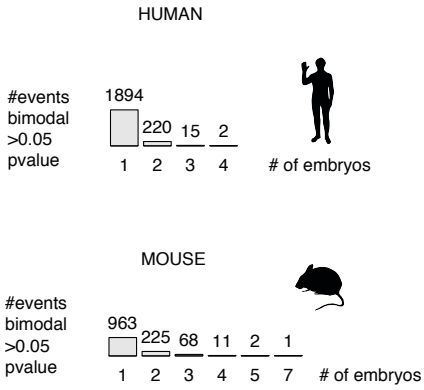
EXAMPLE 1: SMG7



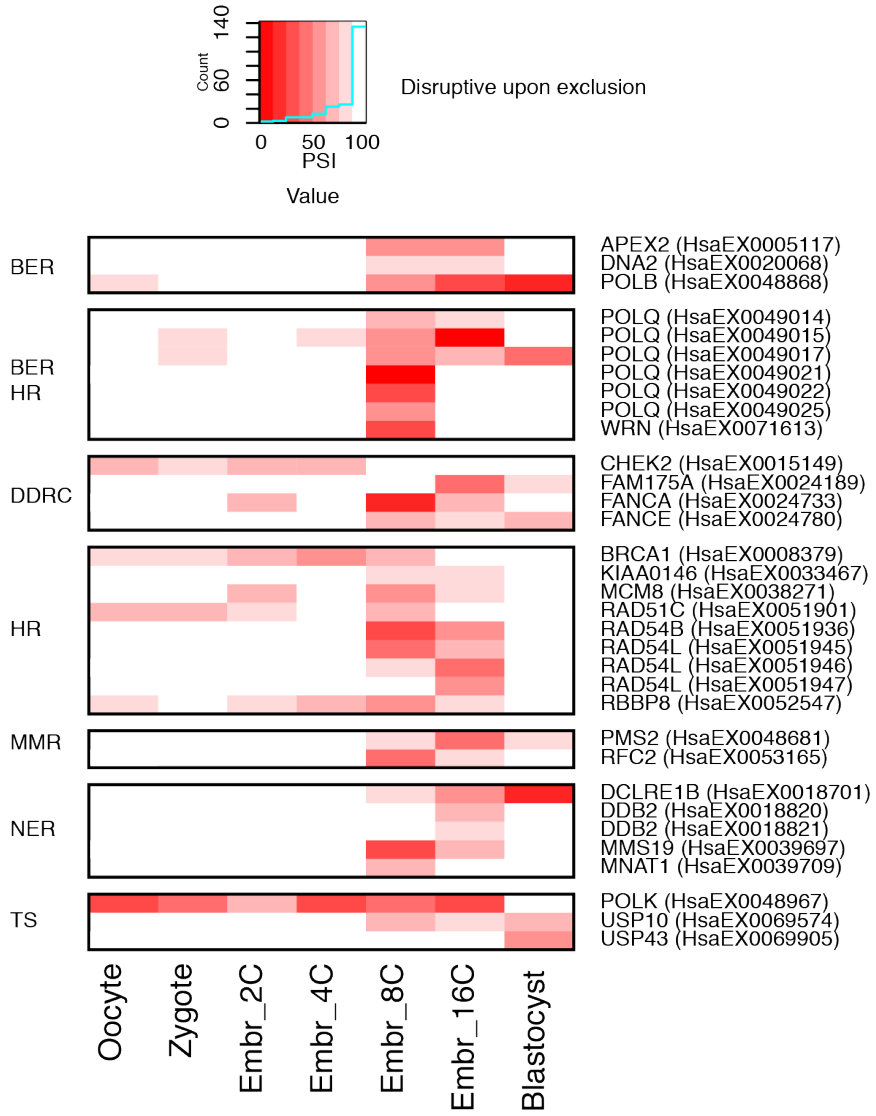
EXAMPLE 1: EHMT2



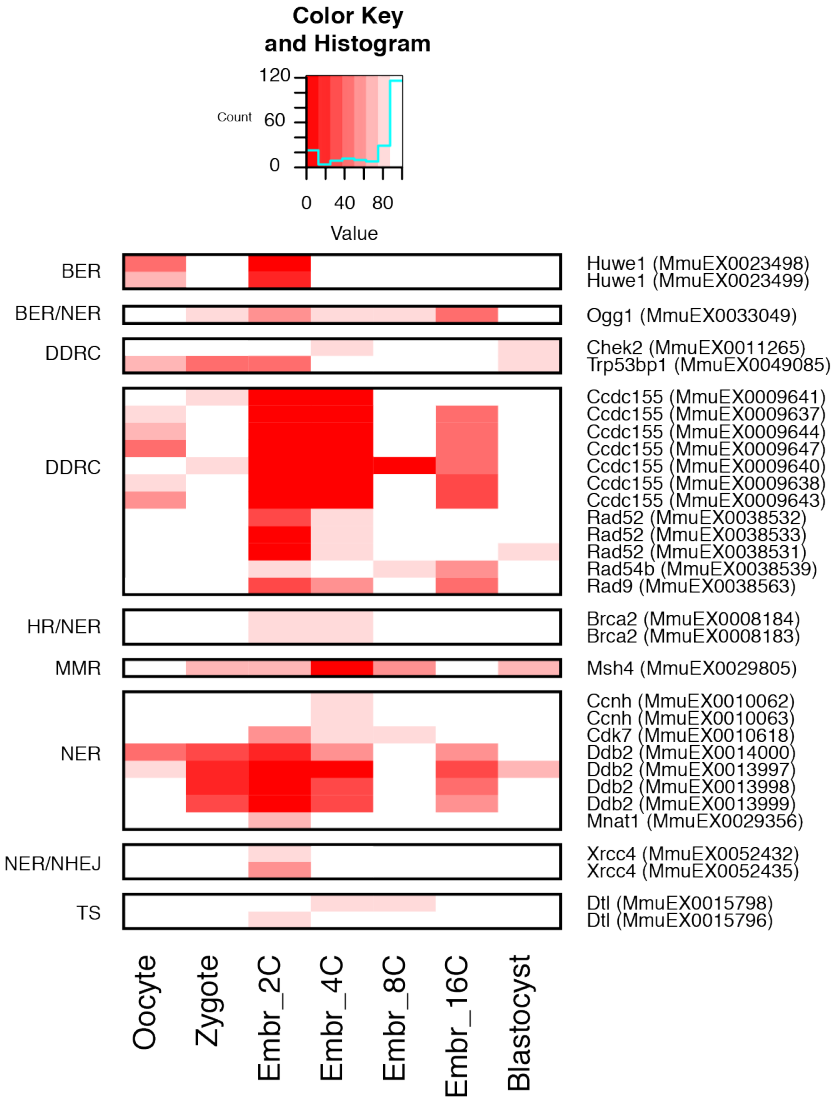
Supp. Figure 1.5 Bimodality in numbers. Barplots to show numbers of events that were significantly bimodality pvalue of <0.05, within any of 12 embryos in human and 18 in mouse. 2-cell stage embryos were merged, counting as a single embryo (as bimodality cannot be called with only two numbers). A minimum of 3 cells with coverage were needed an event in an embryo to be considered. This makes the number of bimodal/variable events biased depending on cell stage, as a 4-Cell embryo may be less likely to have coverage in 3 cells, that an 8-Cell embryo (if they have equal coverage).



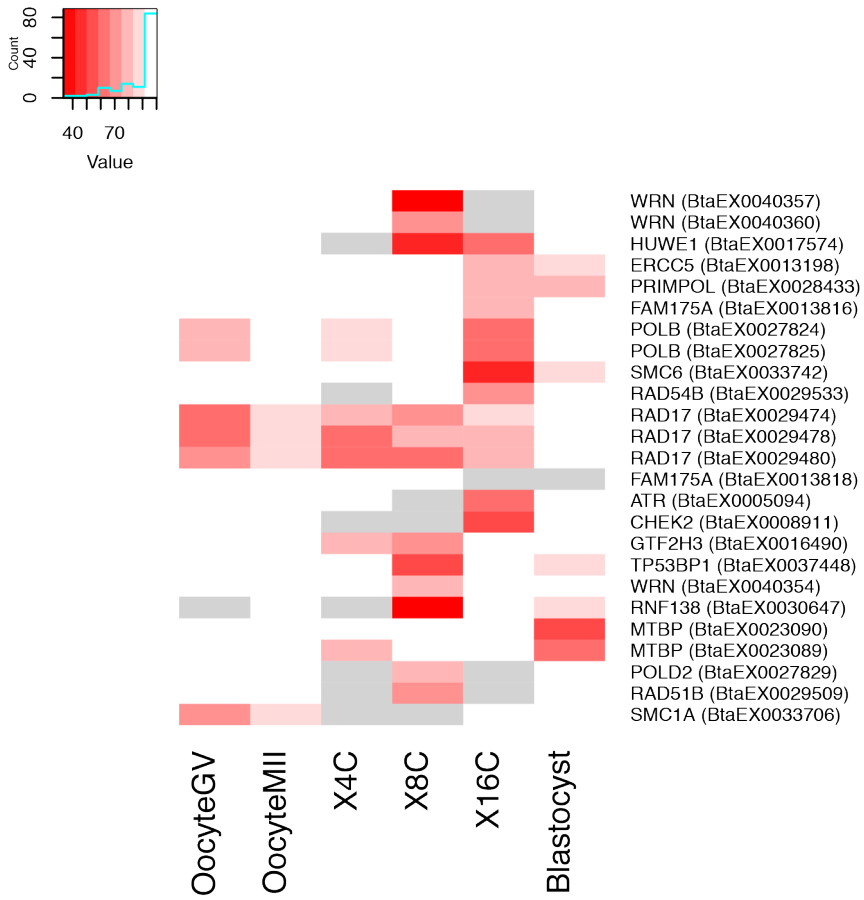
Supp. Figure 2.1 (Part A, Human). Percent spliced in values for events the different categories of DNA damage



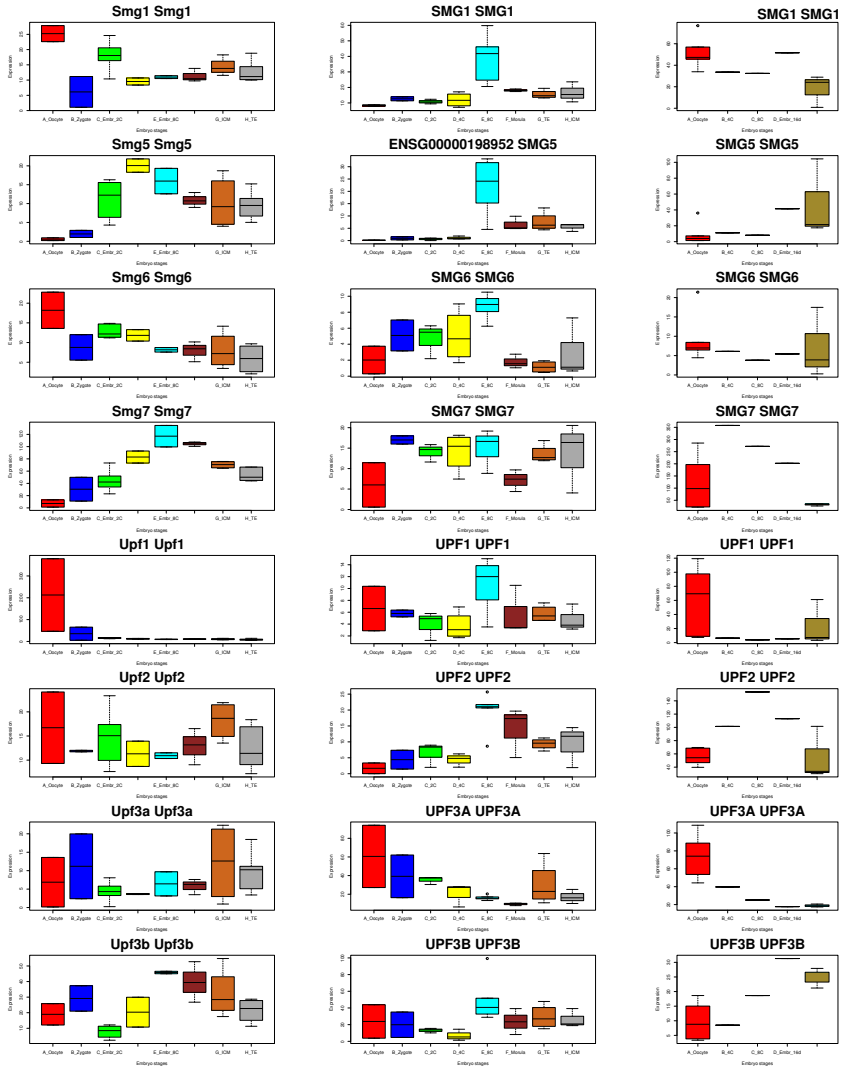
Supp. Figure 2.1 (Part B,Mmu). Percent spliced in values for events the different categories of DNA damage.



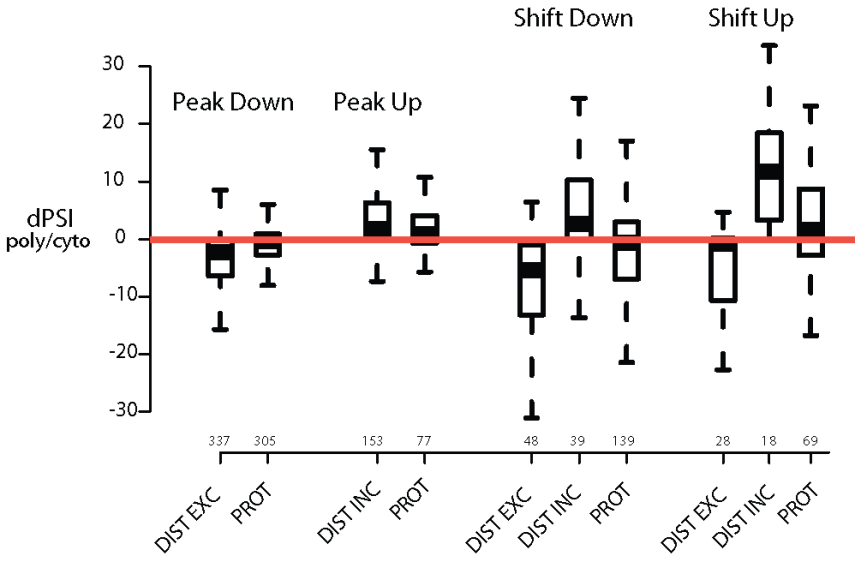
Supp. Figure 2.1 (Part C, Bta). Percent spliced in values for events the different categories of DNA damage.



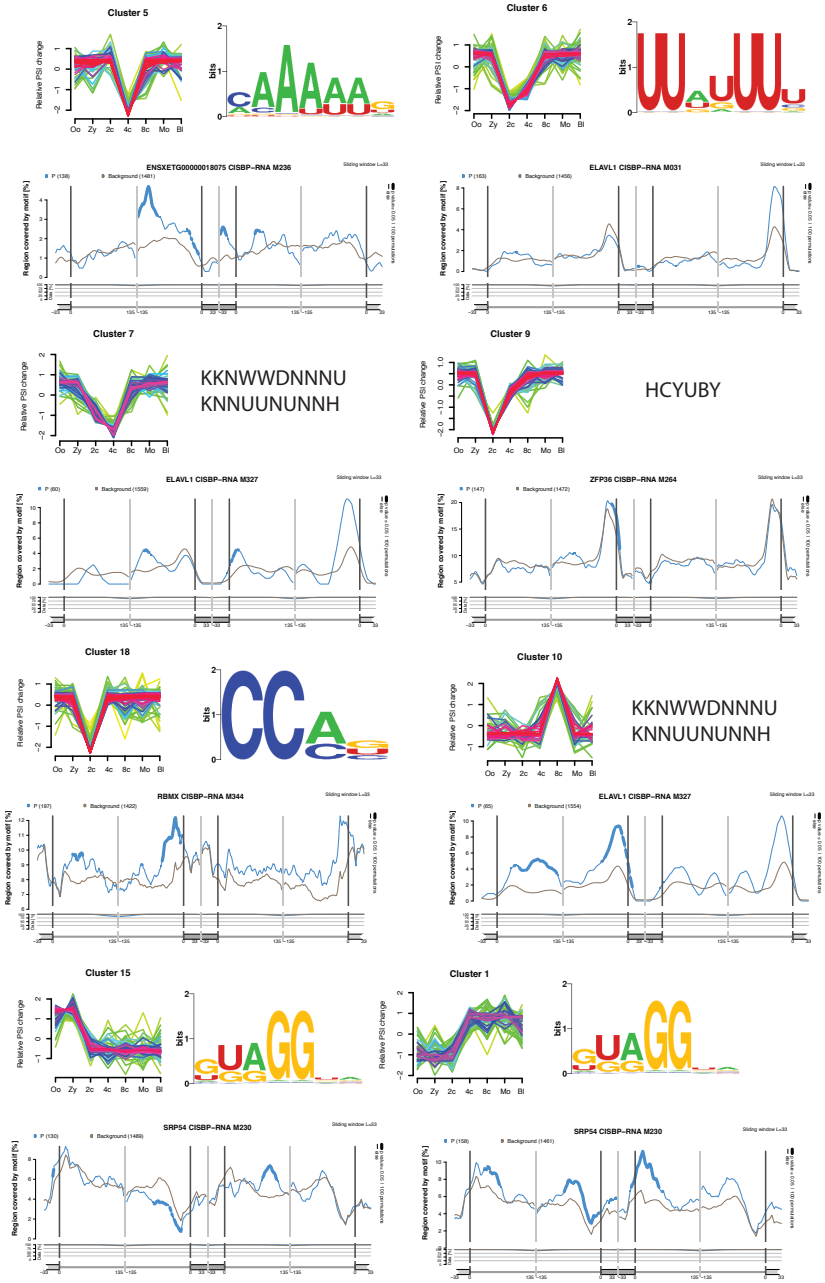
Supp. Figure 2.2.a. Expression of NMD-related genes in human (middle), mouse (left) and cow (right). The four main SMG and UPF genes are shown. Y-axis shows the cRPKM values for expression.



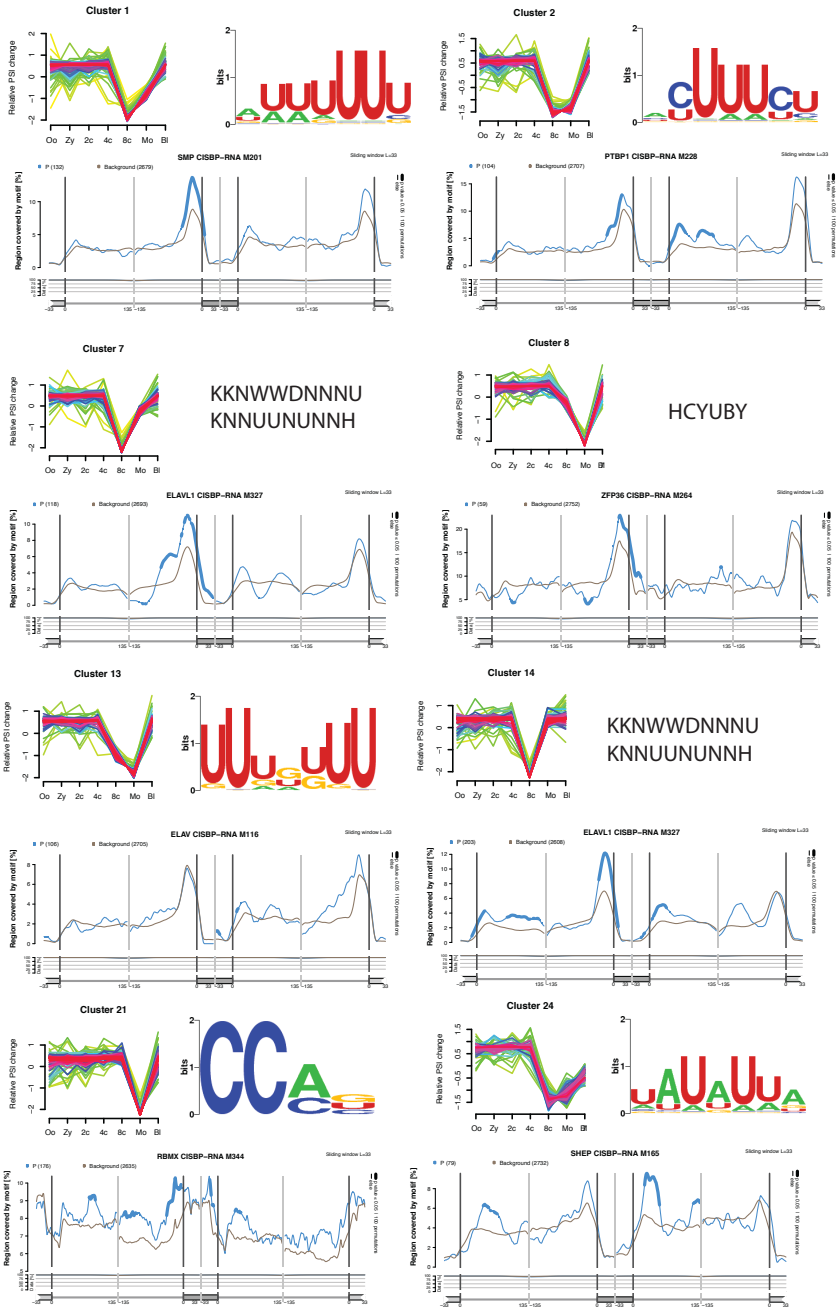
Supp. Figure 2.3. Disrupted isoforms are less bound by polysomes. For human 'Mfuzz' clusters changing at genome activation (4-8Cell stage), split into peak/shift and up/down, and further into those that are disruptive upon exclusion (DISR EXC), inclusion (DIST INC) or an in-frame protein isoform (PROT), (i.e. not changing tri-nucleotide order). We show the change in PSI between the polysome and cytosolic compartments in embryonic stem cells (data from (Blair et al. 2017)), with a positive delta PSI indicating the more polysome fraction.



Supp. Figure 3.1.i. Example mouse Mfuzz clusters, and their top RBP motif enrichment as determined by matt (Gohr and Irimia 2019).



Supp. Figure 3.1.ii.: Example human Mfuzz clusters, and their top RBP motif enrichment as determined by matt (Gohr and Irimia 2019).



Supp. Figure 3.1.iii. Kmer enrichment across alternatively spliced junctions, at the genome activation transition for the three species, with exons becoming more included (blue), skipped (red) or those that do not change in the transition (grey; ndiff). Thick lines show regions that have a p.value less than 0.05 compared to those that do not change in inclusion. From left to right (genomic region), you have the upstream C1 exon (35nt) and C1 upstream intron (135nt), upstream intron of the alternative exon (135nt), alternative exon (35 nucleotides from either side), downstream intron (135nt), C2 intron (135nt) and C2 exon (35nt).

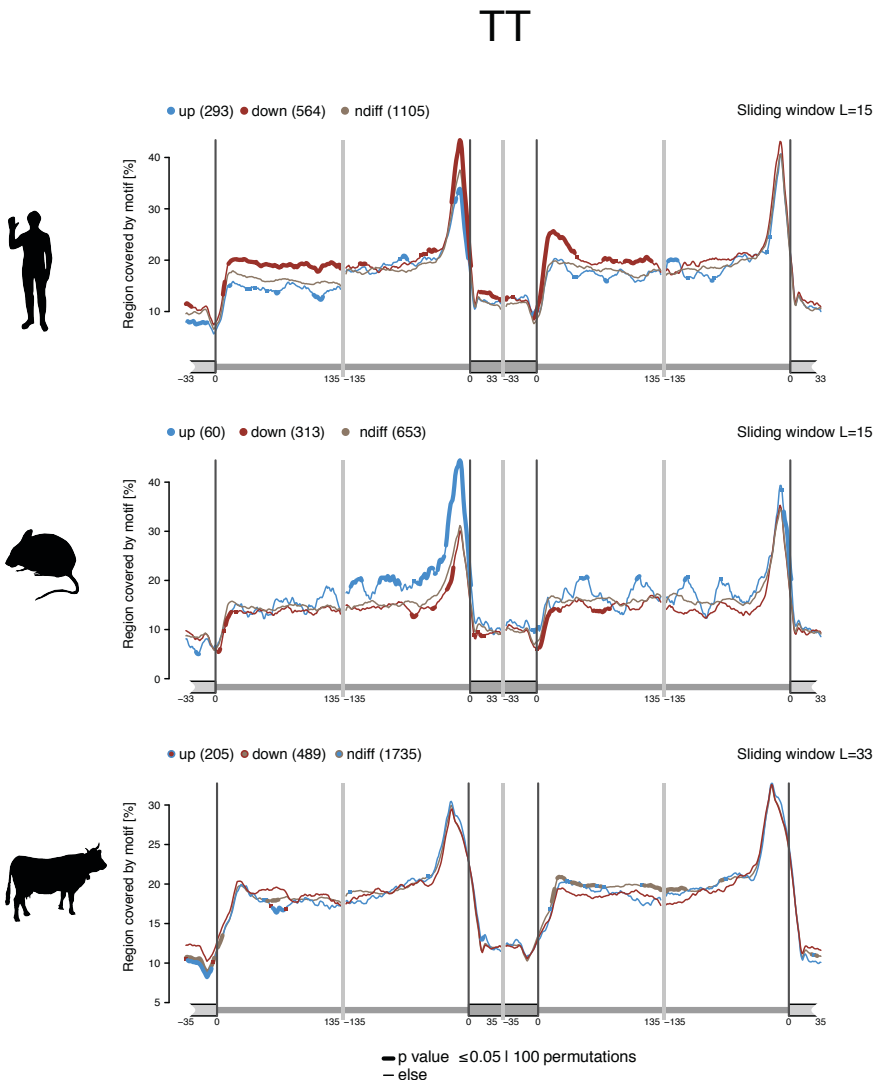


Figure S3.2.ii. The top three clusters in each species, for novel motifs as detected by MEME software for the region 500bp into the up or down stream intron (noted in figure for each cluster.

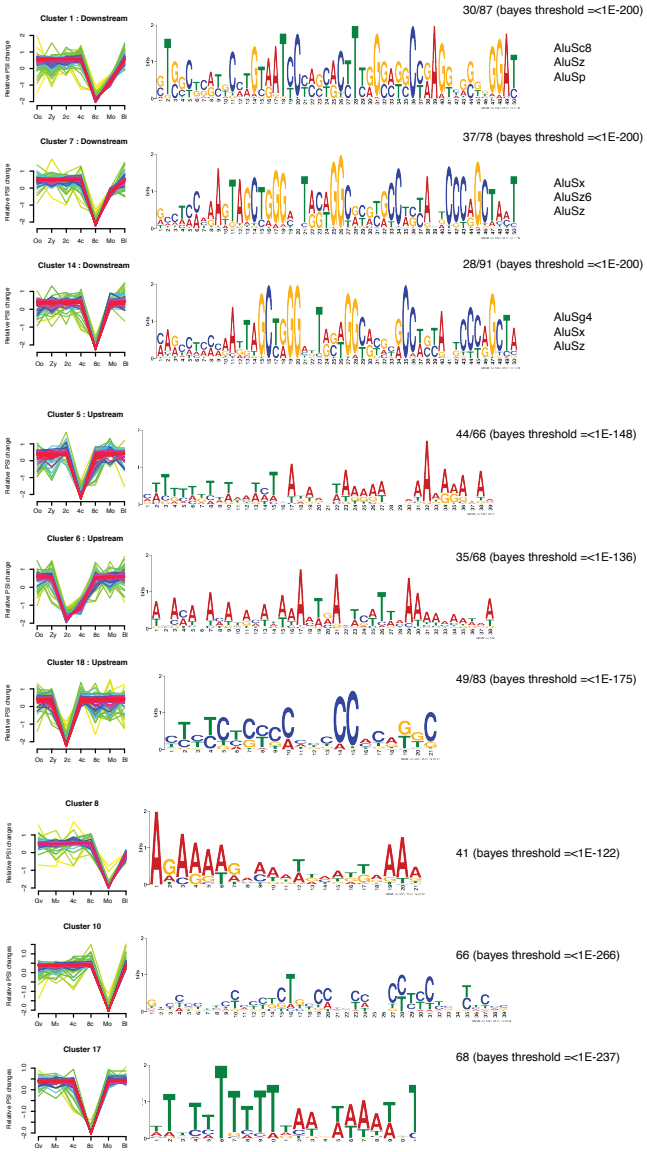
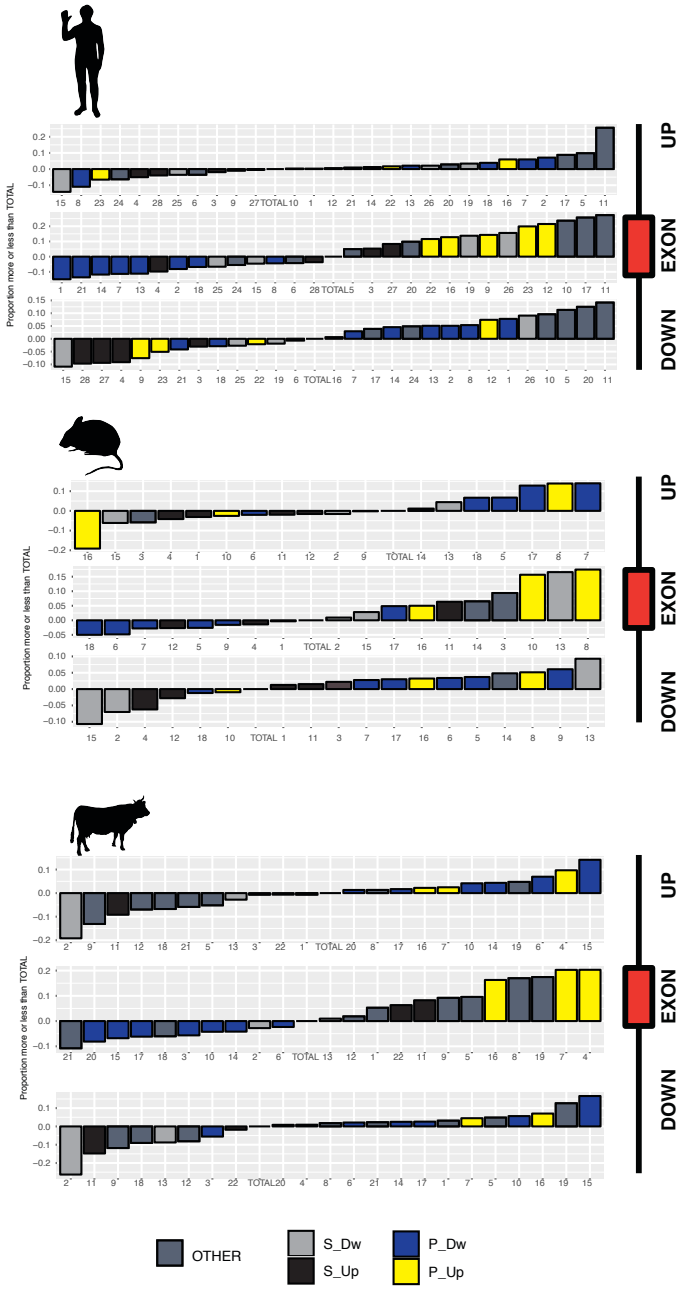
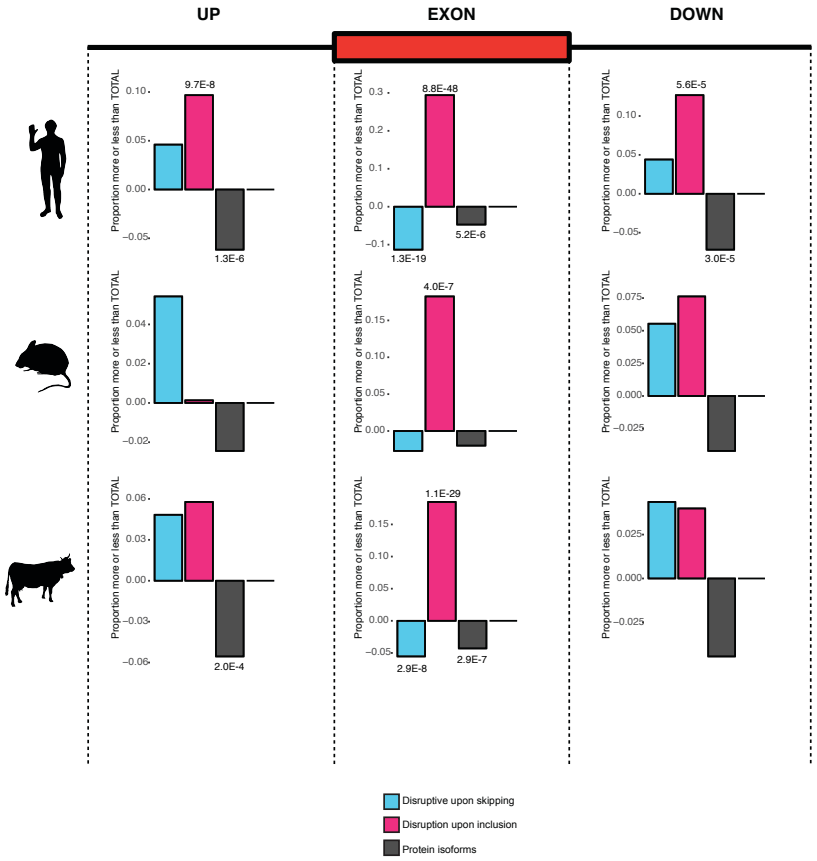


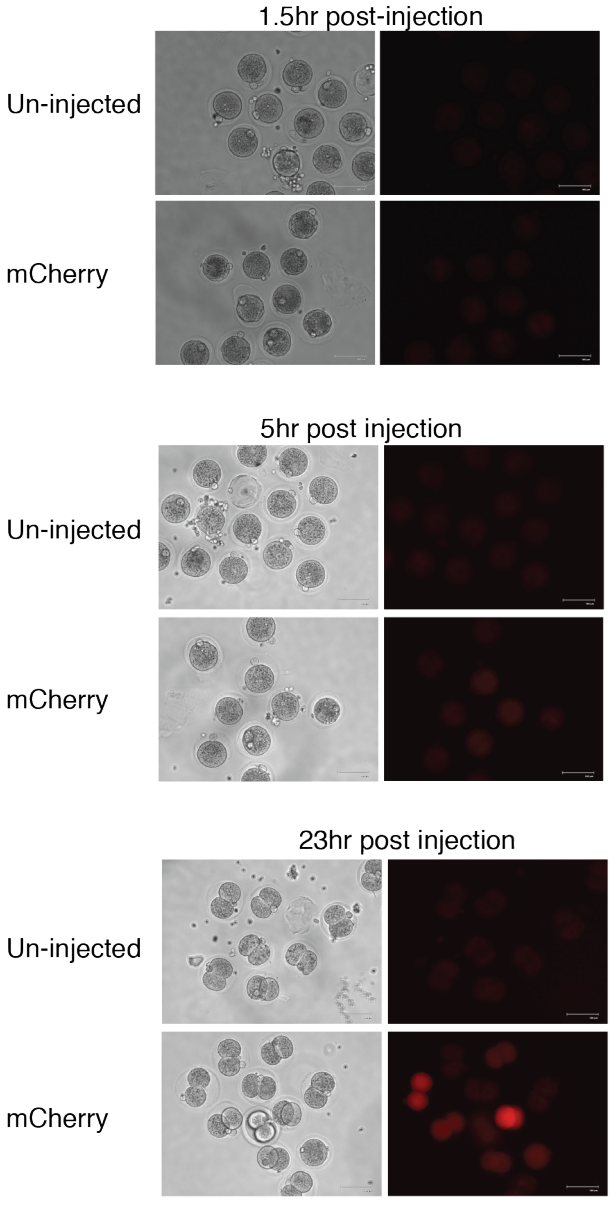
Figure S3.2.iii. Cluster specific transposable element enrichment. Y-axis: Proportion of Transposable elements different to the average (those in all Mfuzz clusters). In the upstream, exonic or downstream regions. 500nt size window, unless exon or intron is smaller than 500nt. Colours of the bars represent the cluster type, either peak/shift and up/down.



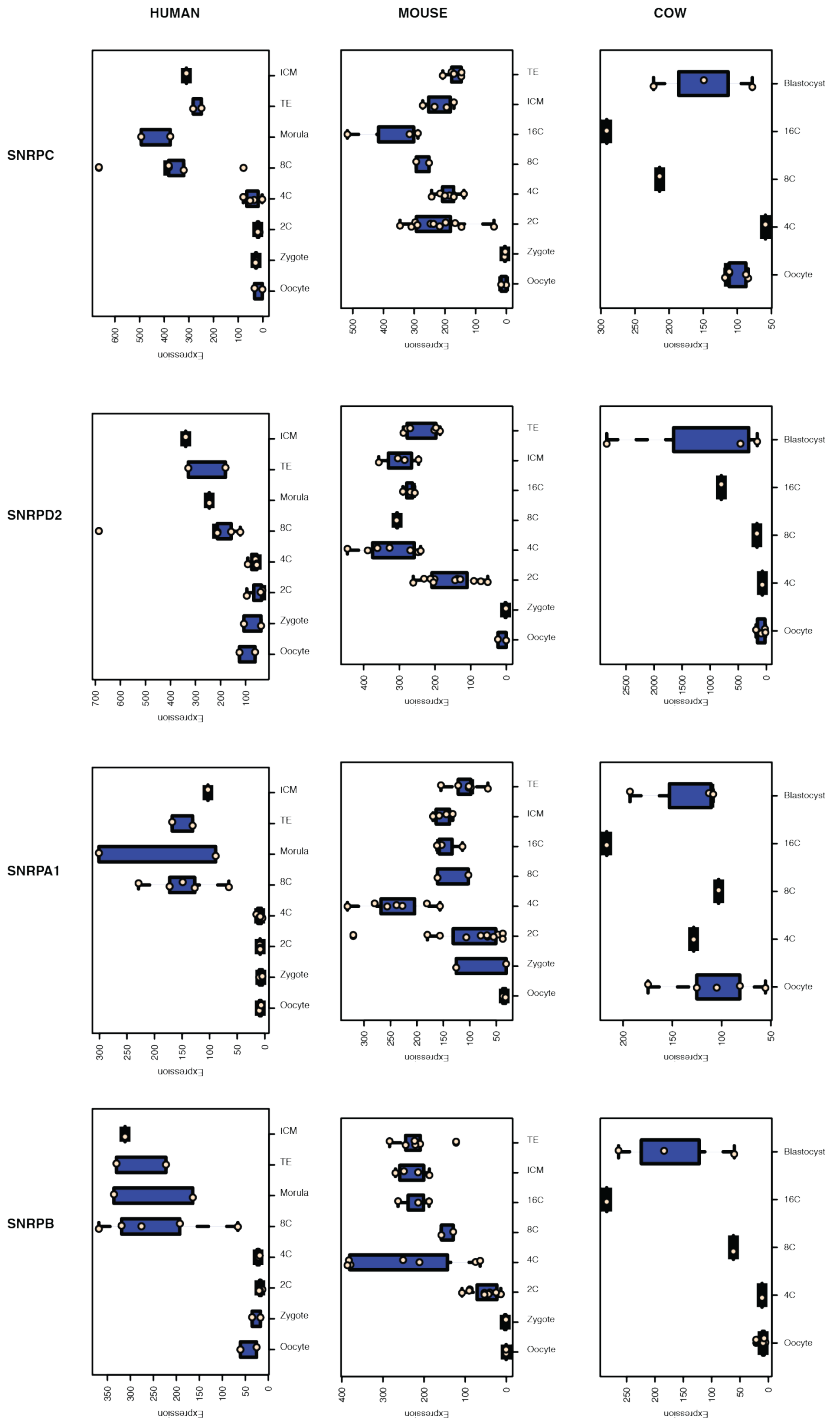
Supplementary Figure 3.2.iv. Enrichment and depletion of transposable elements in the three species, divided into up, down and exonic regions, with each bar representing the exon skipping events with their predicted fate, either protein coding, disruptive upon inclusion or exclusion. Binomial p.values show significance of set against the background of all events in an Mfuzz cluster (ie. had coverage across the time course).



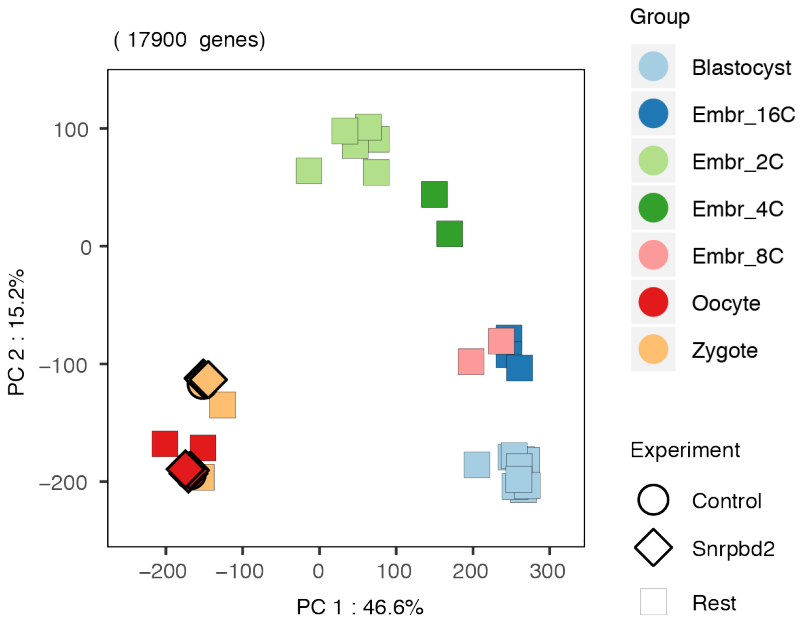
Supp. Figure 4.3.a. Snrpb/d2 mRNA microinjection with mCherry. This was performed to ensure that microinjection led to higher levels of mRNA in the embryo.



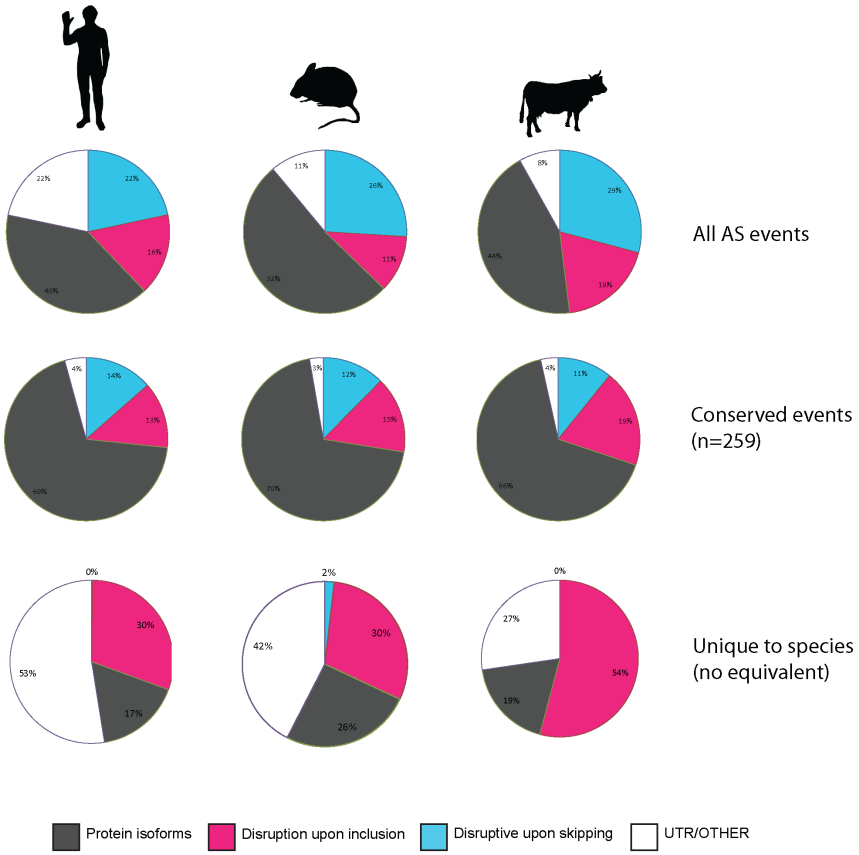
Supp. Fig 3.4. SNRP genes differentially expressed during early development. Expression in cRPKM values.



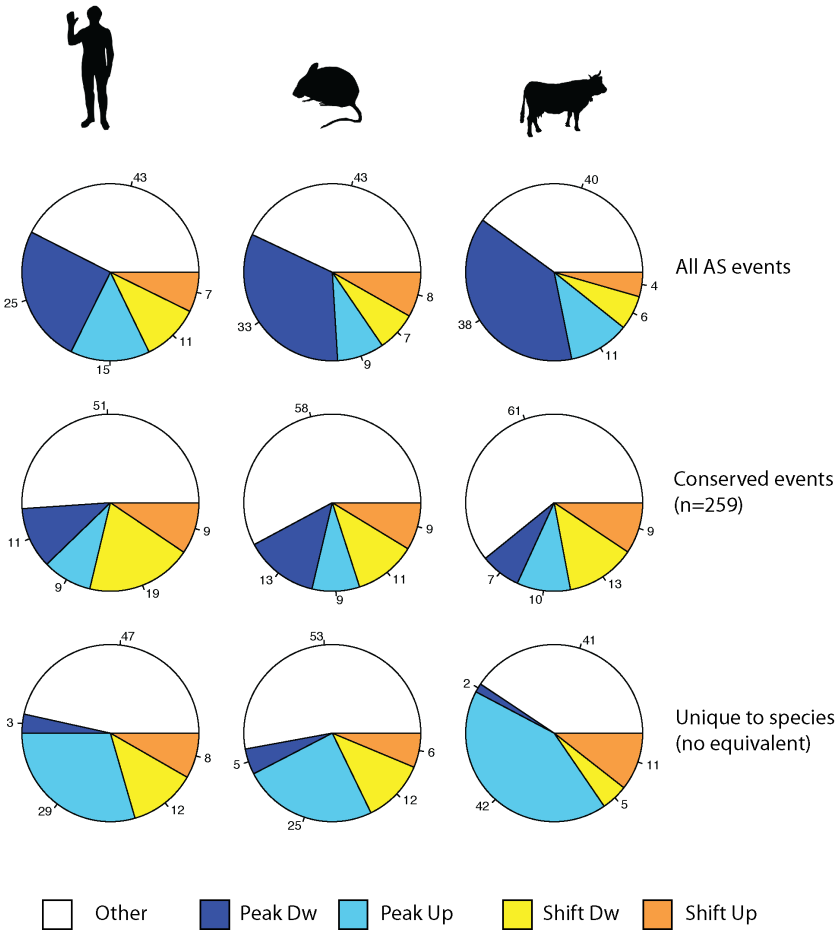
Supp. figure 4.3. PCA of Snrpb/d2 injected embryos. Gene expression values for all early developmental merged samples in mouse and our new experimental samples (circle=control; diamond=overexpression Snrpb/d2 or double overexpression).



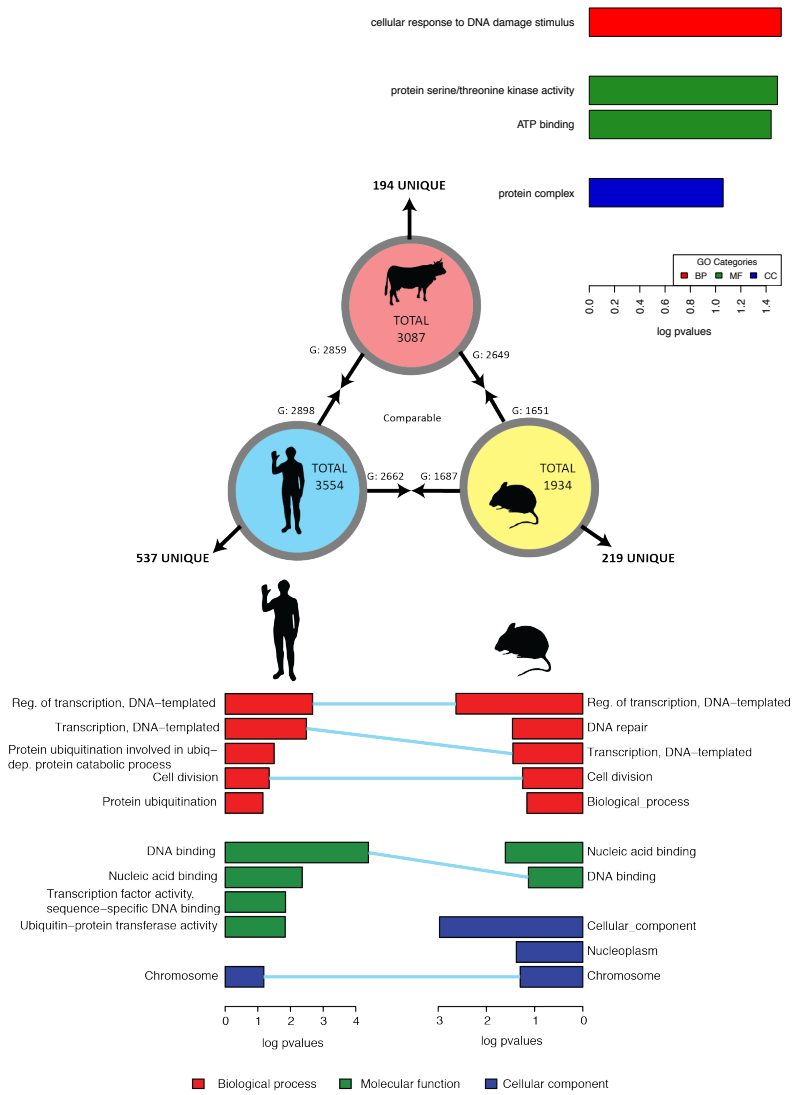
Supp. Figure. 5.1.a Pie charts showing the proportion of events that were protein coding (grey), disruption upon inclusion (pink), disruption upon exclusion (blue) and in the UTR or unassigned (white).



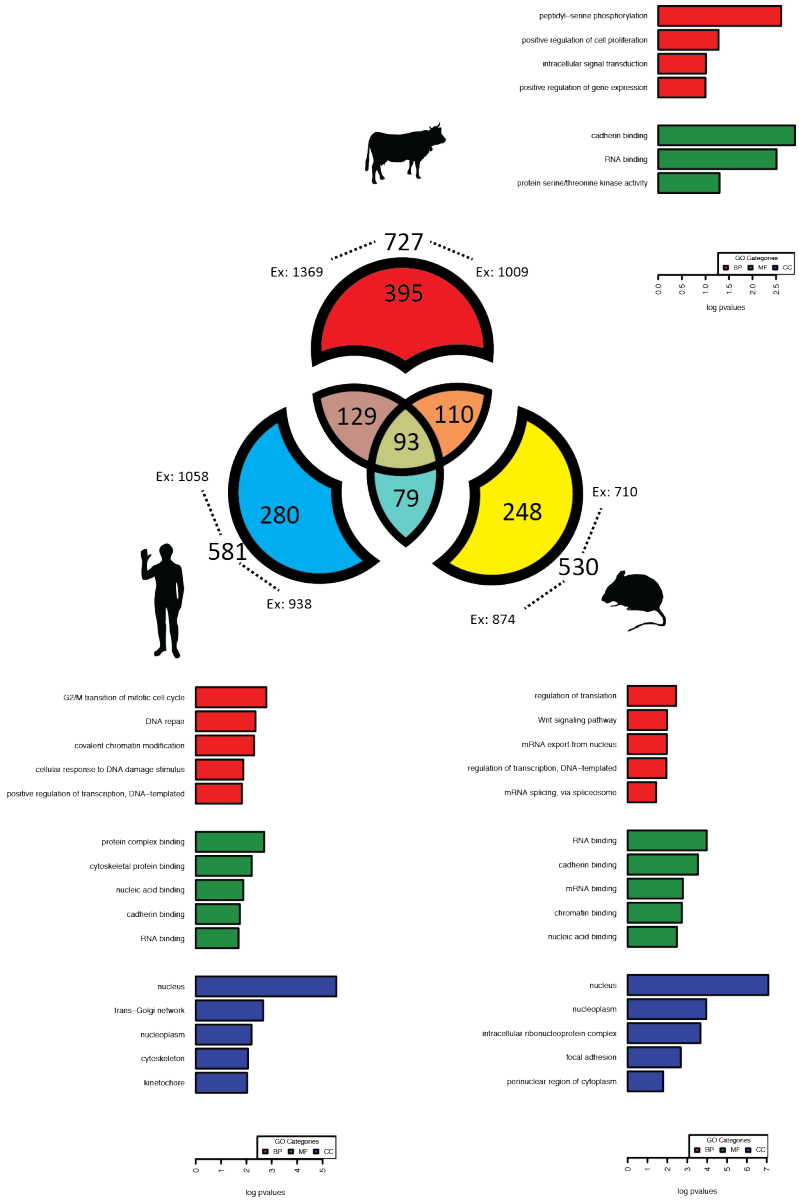
Supp. Figure. 5.1.b Pie charts showing the proportion of events in each cluster type. Shown for all exon-ski[ping events in the three species (top), those conserved (middle) and those unique to each species (at genomic level; bottom)



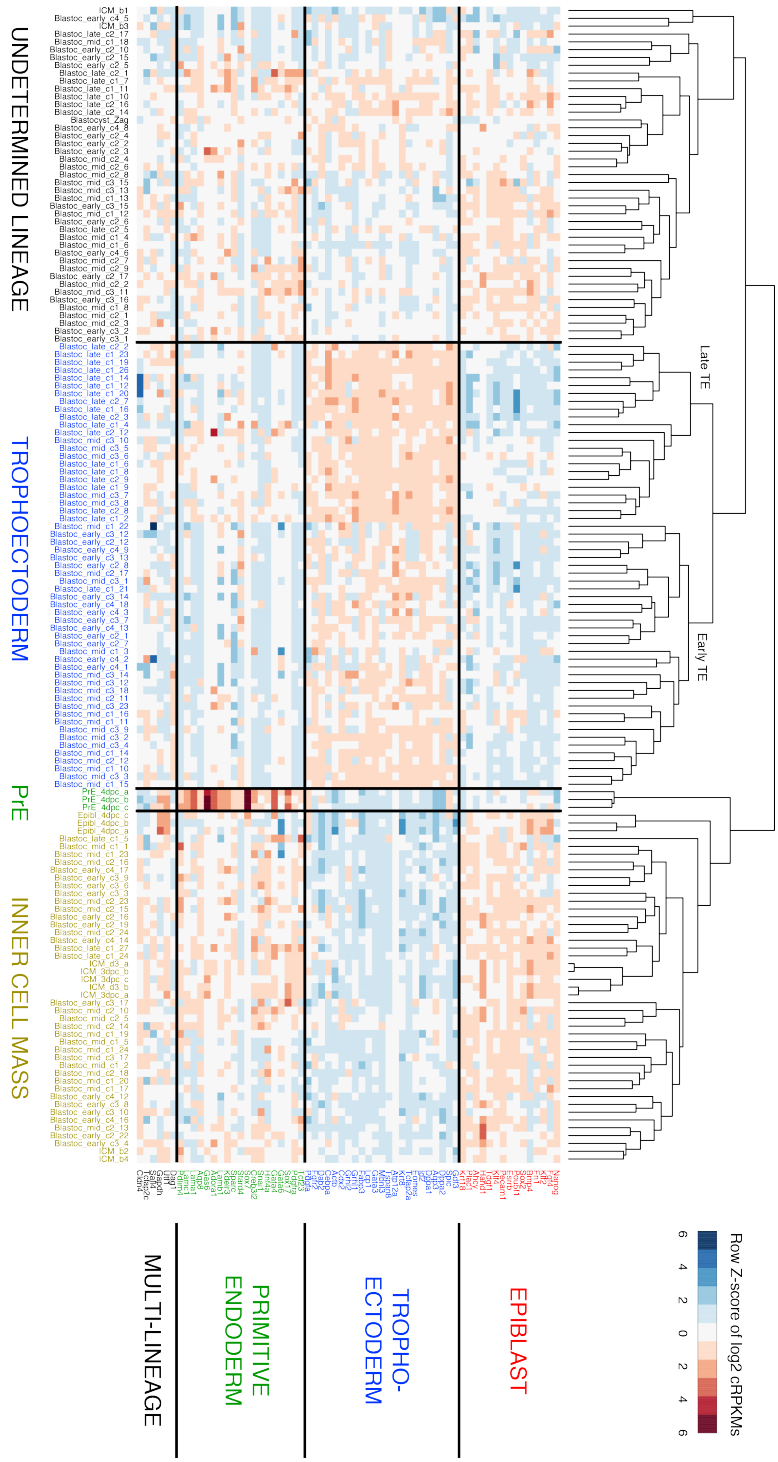
Supp. Figure 5.2.a. Species-specific events. Discarding the total number of events in each species, that had genomic similarity to the other two species, we could calculate the number of species-specific events. Species-specific exon skipping events have significant overlap in gene ontology terms in human and mouse. Showing the log binomial p.values (<0.05) for 537 human-specific exons, and 219 mouse-specific exons, using a background of all events with coverage in any transition. Blue lines show overlap in terms.



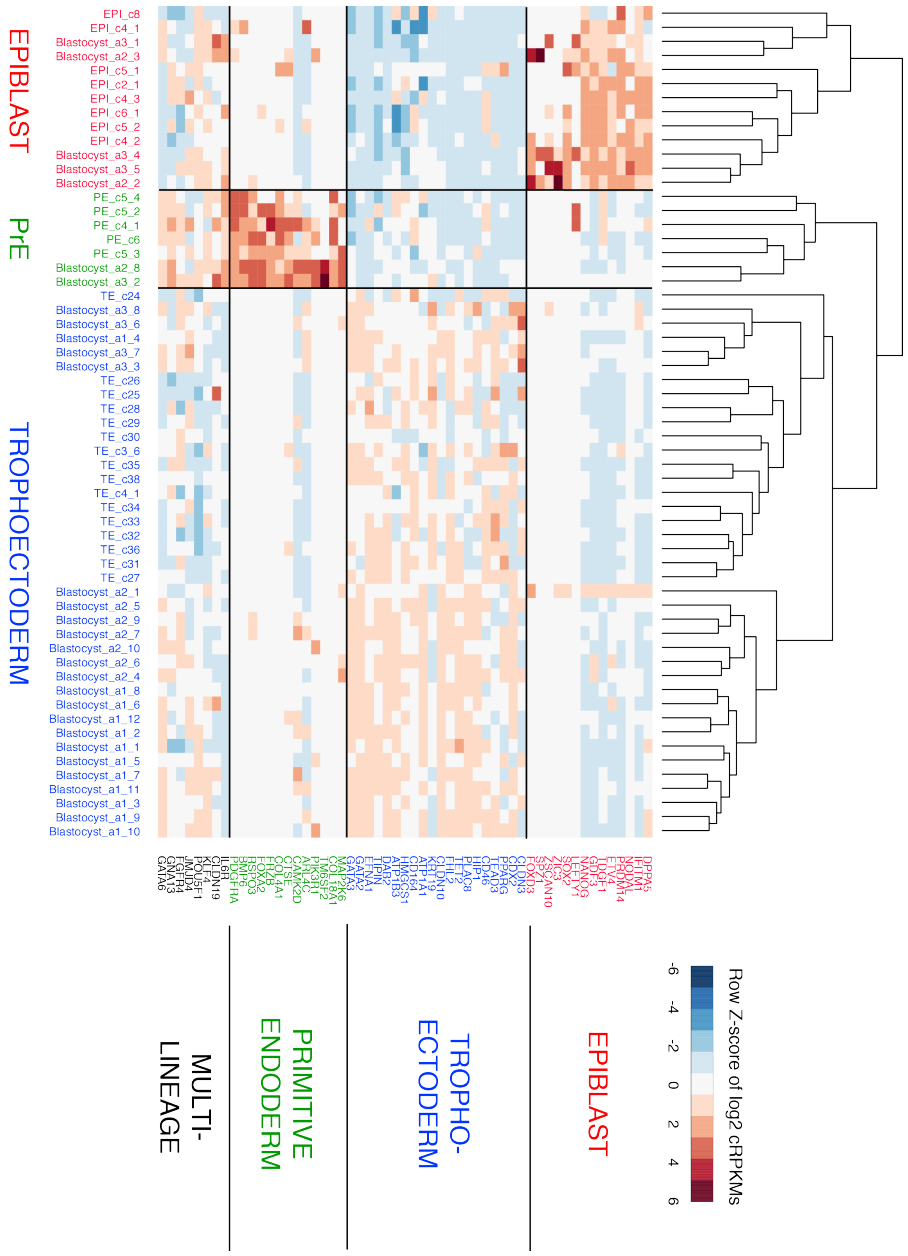
Supp. Figure 5.2.b. Species-specific events measurable in all species. Gene ontology analysis of events detected in the three species, but alternative in one species.



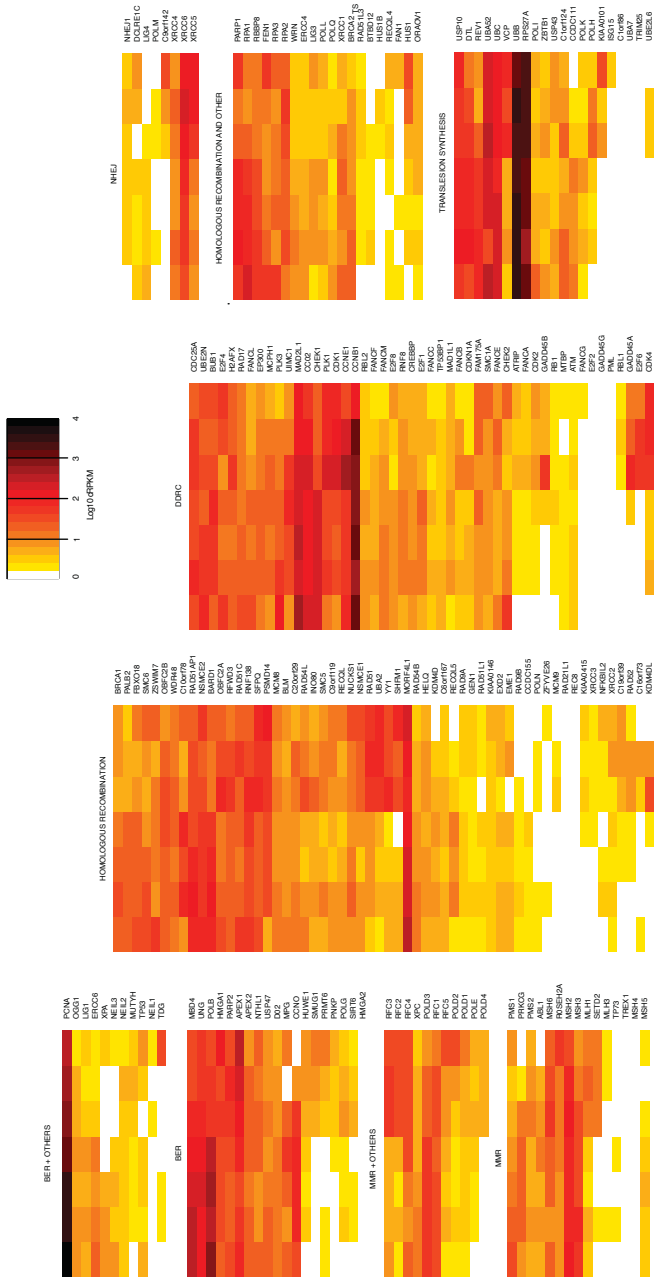
Supplementary Figure 6.1. (Part A). Lineage specific clustering of blastomeres: Gene expression values in cRPKM for lineage markers in Mouse. Blue is low expression and red high. Assumed fate is given to cells by the clustering, into trophectoderm (blue), inner cell mass (yellow),



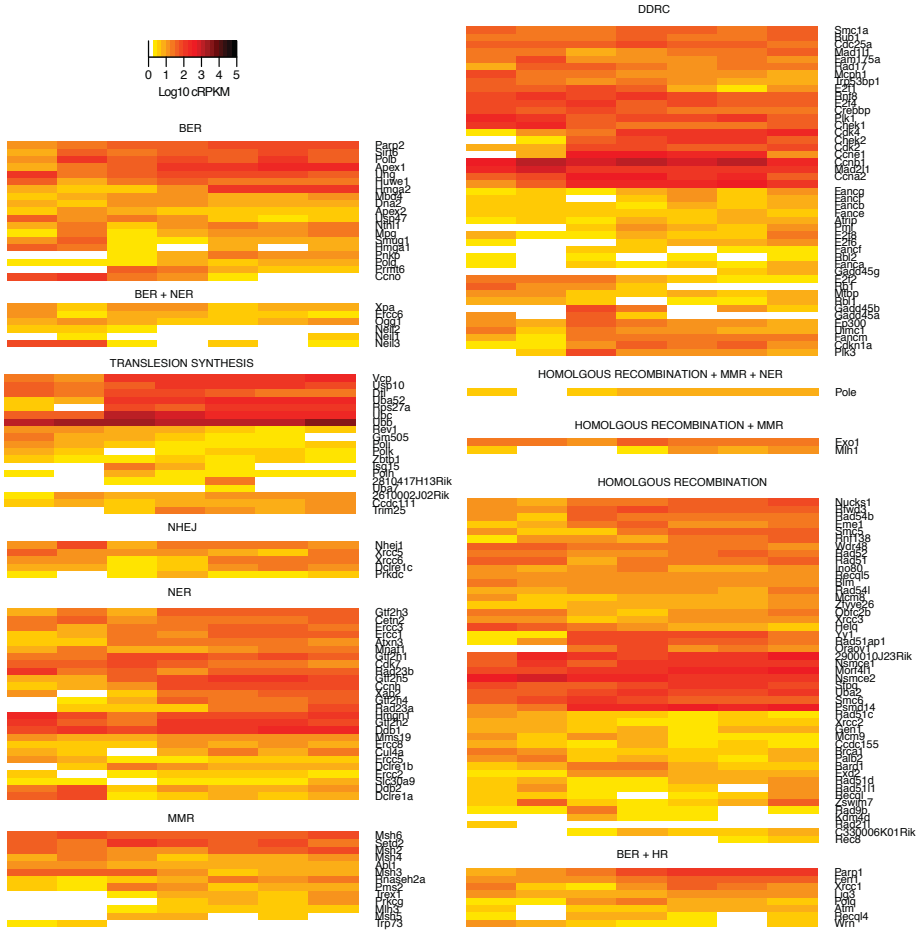
Supplementary Figure 6.1. (Part B). Lineage specific clustering of blastomeres: Gene expression values in cRPKM for lineage markers in Human. Same as mouse, except red is epiblast. We did not get so many unknown cells in human, most likely due to better read coverage of samples.



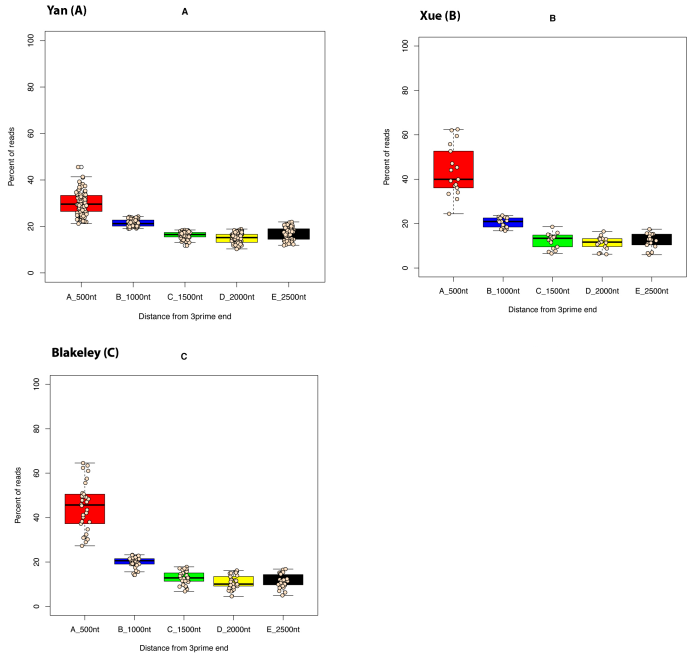
Supplementary figure 4.5 (Part A). DNA damage repair gene expression in human. Values in cRPKM (normalised counts for gene expression). For each gene left to right, shows average levels from the oocyte, zygote, 2cell, 4-cell, 8-cell, 16-cell and blastocyst.



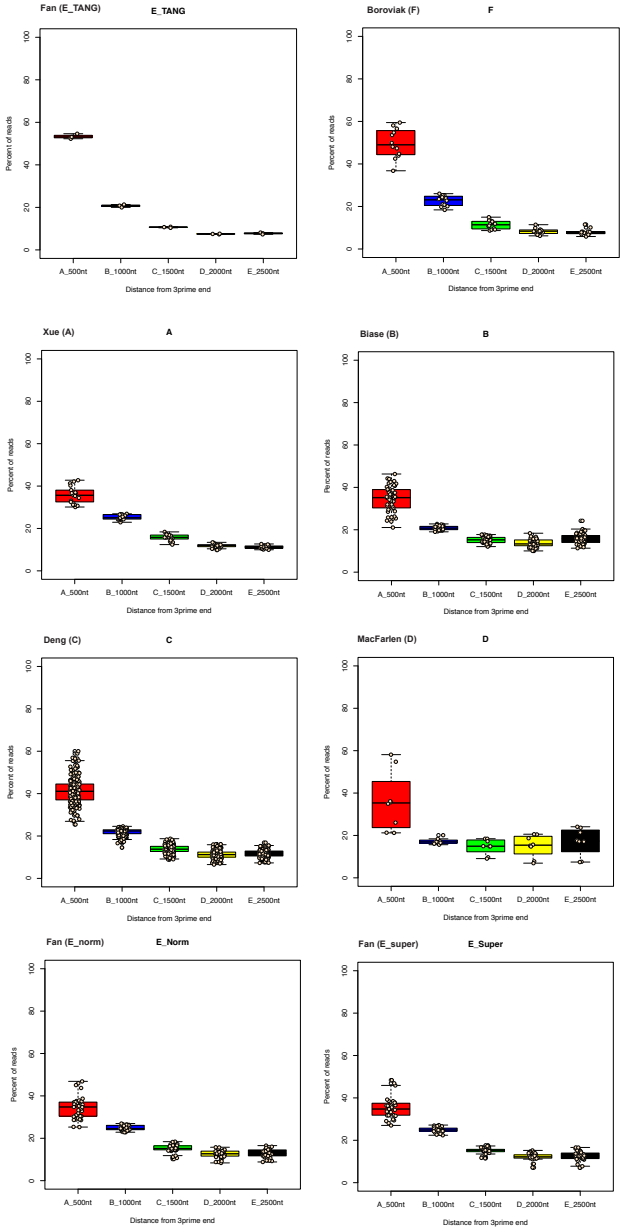
Supplementary figure 4.5. (Part B). DNA damage repair gene expression in mouse. Values in cRPKM (normalised counts for gene expression). For each gene left to right, shows average levels from the oocyte, zygote, 2cell, 4-cell, 8-cell, 16-cell and blastocyst.



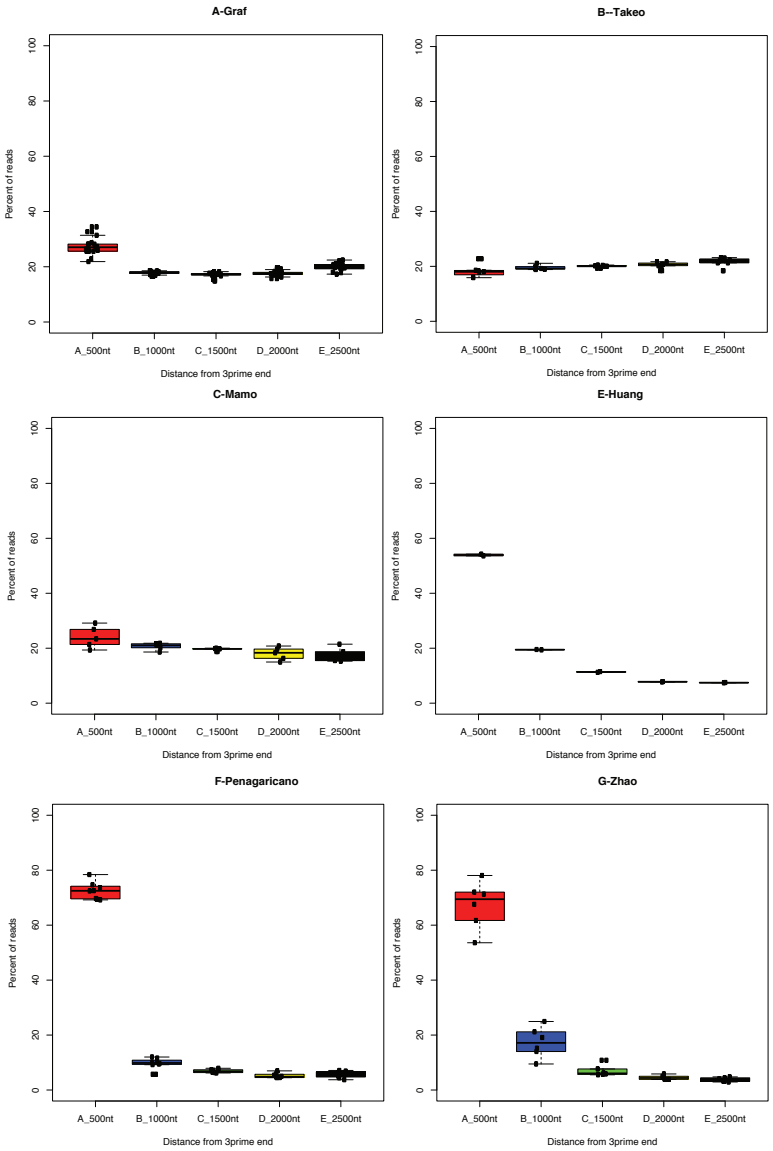
Supp. Figure D.1 (Human). Showing the percentage of reads mapping to the first 500nt from the 3' end, in 500nt bins toward the 5' end.



Supp. Figure D.1 (Mouse). Showing the percentage of reads mapping to the first 500nt from the 3' end, in 500nt bins toward the 5' end.



Supp. Figure D.1 (Cow). Showing the percentage of reads mapping to the first 500nt from the 3' end, in 500nt bins toward the 5' end.



ANNEX

Annex 1. Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals.

Maeso I, Dunwell TL, Wyatt CDR, Marlétaz F, Veto B, Bernal JA, et al. Evolutionary origin and functional divergence of totipotent cell homeobox genes in eutherian mammals. BMC Biol. 2016 Jun 13;14(1). DOI: 10.1186/s12915-016-0267-0

Annex 2: Deconstructing Superorganisms and Societies to Address Big Questions in Biology.

Kennedy P, Baron G, Qiu B, Freitak D, Helanterä H, Hunt ER, et al. Deconstructing Superorganisms and Societies to Address Big Questions in Biology. Trends in Ecology and Evolution. 2017;32(11):861–72. DOI: 10.1016/j.tree.2017.08.004

Annex 3. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies.

Patalano S, Vlasova A, Wyatt C, Ewels P, Camara F, Ferreira PG, et al. Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. Proc Natl Acad Sci U S A. 2015 Nov 10;112(45):13970–5. DOI: 10.1073/pnas.1515937112

Annex 4. The bladed Bangiales (Rhodophyta) of the South Eastern Pacific: Molecular species delimitation reveals extensive diversity

Guillemin ML, Contreras-Porcia L, Ramírez ME, Macaya EC, Contador CB, Woods H, et al. The bladed Bangiales (Rhodophyta) of the South Eastern Pacific: Molecular species delimitation reveals extensive diversity. Mol Phylogenet Evol. 2016 Jan 1;94(Pt B):814–26. DOI: 10.1016/j.ympev.2015.09.027

Annex 5. The alternative splicing factor Nova2 regulates vascular development and lumen formation

Giampietro C, Deflorian G, Gallo S, Di Matteo A, Pradella D, Bonomi S, et al. The alternative splicing factor Nova2 regulates vascular development and lumen formation. Nat Commun. 2015 Oct 8;6. DOI: 10.1038/ncomms9479

Annex 6: Amphioxus functional genomics and the origins of vertebrate gene

Marlétaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, et al. Amphioxus functional genomics and the origins of vertebrate gene regulation. Nature. 2018 Dec 6;564(7734):64–70. DOI: 10.1038/s41586-018-0734-6

BIBLIOGRAPHY

Aanes, Håvard, Olga Østrup, Ingrid S. Andersen, Lars F. Moen, Sinnakaruppan Mathavan, Philippe Collas, and Peter Alestrom. 2013. “Differential Transcript Isoform Usage Pre- and Post-Zygotic Genome Activation in Zebrafish.” *BMC Genomics*. doi:10.1186/1471-2164-14-331.

Abe, Ken-ichiro, Ryoma Yamamoto, Vedran Franke, Minjun Cao, Yutaka Suzuki, Masataka G Suzuki, Kristian Vlahovicek, Petr Svoboda, Richard M Schultz, and Fugaku Aoki. 2015. “The First Murine Zygotic Transcription Is Promiscuous and Uncoupled from Splicing and 3′ Processing.” *Embo J* 34 (11): 1523–37. doi:10.15252/embj.201490648.

Abzhanov, Arhat. 2013. “Von Baer’s Law for the Ages: Lost and Found Principles of Developmental Evolution.” *Trends in Genetics*. doi:10.1016/j.tig.2013.09.004.

Ahuja, Akshay K., Karolina Jodkowska, Federico Teloni, Anna H. Bizard, Ralph Zellweger, Raquel Herrador, Sagrario Ortega, et al. 2016. “A Short G1 Phase Imposes Constitutive Replication Stress and Fork Remodelling in Mouse Embryonic Stem Cells.” *Nature Communications*. doi:10.1038/ncomms10660.

Aitken, John, and Helen Fisher. 1994. “Reactive Oxygen Species Generation and Human Spermatozoa: The Balance of Benefit and Risk.” *BioEssays*. doi:10.1002/bies.950160409.

Aitken, R. J., and G. N. De Iuliis. 2009. “On the Possible Origins of DNA Damage in Human Spermatozoa.” *Molecular Human Reproduction*. doi:10.1093/molehr/gap059.

Alamancos, Gael P., Amadís Pagès, Juan L. Trincado, Nicolás Bellora, Eduardo Eyras, A. Pages, Juan L. Trincado, Nicolás Bellora, and Eduardo Eyras. 2014. “SUPPA: A Super-Fast Pipeline for Alternative Splicing Analysis from RNA-Seq.” *BioRxiv*. doi:10.1101/008763.

Alberts, B, A Johnson, J Lewis, M Raff, K Roberts, and P And Walter. 2008. *Molecular Biology of the Cell*. *Amino Acids*. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

Altschuler, Yoram, Shana M. Barbas, Laura J. Terlecky, Kitty Tang, Stephen Hardy, Keith E. Mostov, and Sandra L. Schmid. 1998. “Redundant and Distinct Functions for Dynamin-1 and Dynamin-2 Isoforms.” *Journal of Cell Biology*. doi:10.1083/jcb.143.7.1871.

Ames, B N, M K Shigenaga, and T M Hagen. 1993. “Oxidants, Antioxidants, and the Degenerative Diseases of Aging.” *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.90.17.7915.

Aoki, Fugaku, Diane M. Worrad, and Richard M. Schultz. 1997. “Regulation of Transcriptional Activity during the First and Second Cell Cycles in the Preimplantation Mouse Embryo.” *Developmental Biology*. doi:10.1006/dbio.1996.8466.

Artus, Jérôme, and Michel Cohen-Tannoudji. 2008. “Cell Cycle Regulation during Early Mouse Embryogenesis.” *Molecular and Cellular Endocrinology*. doi:10.1016/j.mce.2007.11.008.

Austin, C. R. 1956. “Cortical Granules in Hamster Eggs.” *Experimental Cell Research*. doi:10.1016/0014-4827(56)90025-8.

Ayoub, Nabieh, Anand D. Jeyasekharan, Juan A. Bernal, and Ashok R. Venkitaraman. 2008. “HP1-β Mobilization Promotes Chromatin Changes That Initiate the DNA Damage Response.” *Nature*. doi:10.1038/nature06875.

Babariya, D., E. Fragouli, S. Alfarawati, K. Spath, and D. Wells. 2017. “The Incidence and Origin of Segmental Aneuploidy in Human Oocytes and Preimplantation Embryos.” *Human Reproduction*. doi:10.1093/humrep/dex324.

Bailey, Timothy L. 2011. “DREME: Motif Discovery in Transcription Factor ChIP-Seq Data.”

Bioinformatics. doi:10.1093/bioinformatics/btr261.

Bailey, Timothy L, and Charles Elkan. 1994. "Fitting a Mixture Model by Expectation Maximization to Discover Motifs in Biopolymers." *Proc. 2nd Int. Conf. on Intelligent Systems for Molecular Biology*. doi:citeulike-article-id:878292.

Bang, Marie Louise, Thomas Centner, Friderike Fornoff, Adam J. Geach, Michael Gotthardt, Mark McNabb, Christian C. Witt, et al. 2001. "The Complete Gene Sequence of Titin, Expression of an Unusual ≈700-KDa Titin Isoform, and Its Interaction with Obscurin Identify a Novel Z-Line to I-Band Linking System." *Circulation Research*. doi:10.1161/hh2301.100981.

Barbosa-Morais, Nuno L, Manuel Irimia, Qun Pan, Hui Y Xiong, Serge Gueroussov, Leo J Lee, Valentina Slobodeniuc, et al. 2012. "The Evolutionary Landscape of Alternative Splicing in Vertebrate Species." *Science (New York, NY)* 338 (6114): 1587–93. doi:10.1126/science.1230612.

Bazrgar, Masood, Hamid Gourabi, Poopak Eftekhari Yazdi, Hamed Vazirinasab, Mostafa Fakhri, Fatemeh Hassani, and Mojtaba Rezazadeh Valojerdi. 2014. "DNA Repair Signalling Pathway Genes Are Overexpressed in Poor-Quality Pre-Implantation Human Embryos with Complex Aneuploidy." *European Journal of Obstetrics Gynecology and Reproductive Biology*. doi:10.1016/j.ejogrb.2014.01.010.

Bebee, Thomas W., Juw Won Park, Katherine I. Sheridan, Claude C. Warzecha, Benjamin W. Cieply, Alex M. Rohacek, Yi Xing, and Russ P. Carstens. 2015. "The Splicing Regulators Esrp1 and Esrp2 Direct an Epithelial Splicing Program Essential for Mammalian Development." *ELife*. doi:10.7554/eLife.08954.

Bebenek, Katarzyna, Lars C. Pedersen, and Thomas A. Kunkel. 2011. "Replication Infidelity via a Mismatch with Watson–Crick Geometry." *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1012825108.

Beerman, Isabel, Jun Seita, Matthew A. Inlay, Irving L. Weissman, and Derrick J. Rossi. 2014. "Quiescent Hematopoietic Stem Cells Accumulate DNA Damage during Aging That Is Repaired upon Entry into Cell Cycle." *Cell Stem Cell*. doi:10.1016/j.stem.2014.04.016.

Benayoun, Bérénice A., Elizabeth A. Pollina, Duygu Ucar, Salah Mahmoudi, Kalpana Karra, Edith D. Wong, Keerthana Devarajan, et al. 2014. "H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency." *Cell*. doi:10.1016/j.cell.2014.06.027.

Bernstein, Emily, Sang Yong Kim, Michelle A. Carmell, Elizabeth P. Murchison, Heather Alcorn, Mamie Z. Li, Alea A. Mills, Stephen J. Elledge, Kathryn V. Anderson, and Gregory J. Hannon. 2003. "Dicer Is Essential for Mouse Development." *Nature Genetics*. doi:10.1038/ng1253.

Beyer, Ann L., Amy H. Bouton, and Oscar L. Miller. 1981. "Correlation of HnRNP Structure and Nascent Transcript Cleavage." *Cell*. doi:10.1016/0092-8674(81)90299-3.

Bhargava, Vipul, Steven R Head, Phillip Ordoukhanian, Mark Mercola, and Shankar Subramaniam. 2014. "Technical Variations in Low-Input RNA-Seq Methodologies." *Scientific Reports* 4 (January). Macmillan Publishers Limited. All rights reserved: 3678. http://dx.doi.org/10.1038/srep03678.

Bianchi, Enrica, Brendan Doe, David Goulding, and Gavin J. Wright. 2014. "Juno Is the Egg Izumo Receptor and Is Essential for Mammalian Fertilization." *Nature*. doi:10.1038/nature13203.

Biase, Fernando H., Xiaoyi Cao, and Sheng Zhong. 2014. "Cell Fate Inclination within 2-Cell and 4-Cell Mouse Embryos Revealed by Single-Cell RNA Sequencing." *Genome Research* 24 (11): 1787–96. doi:10.1101/gr.177725.114.

Blair, John D., Dirk Hockemeyer, Jennifer A. Doudna, Helen S. Bateup, and Stephen N. Floor. 2017. "Widespread Translational Remodeling during Human Neuronal Differentiation." *Cell Reports*. doi:10.1016/j.celrep.2017.10.095.

Blakeley, Paul, Norah M E Fogarty, Ignacio del Valle, Sissy E Wamaitha, Tim Xiaoming Hu, Kay Elder, Philip Snell, Leila Christie, Paul Robson, and Kathy K Niakan. 2015. "Defining the Three Cell Lineages of the Human Blastocyst by Single-Cell RNA-Seq." *Development* , August. doi:10.1242/dev.123547.

Bleil, Jeffrey D., Clifford F. Beall, and Paul M. Wassarman. 1981. "Mammalian Sperm-Egg Interaction: Fertilization of Mouse Eggs Triggers Modification of the Major Zona Pellucida Glycoprotein, ZP2." *Developmental Biology*. doi:10.1016/0012-1606(81)90329-8.

Blythe, Shelby A., and Eric F. Wieschaus. 2015. "Zygotic Genome Activation Triggers the DNA Replication Checkpoint at the Midblastula Transition." *Cell*. doi:10.1016/j.cell.2015.01.050.

Boag, Peter R, Paul Francis Harrison, Adele Barugahare, Andrew D Pattison, Angavai Swaminathan, Kirill Tsyganov, Eva Heinz, Gregory Davis, David R Powell, and Traude Helene Beilharz. 2018. "Widespread Cytoplasmic Polyadenylation Programs Asymmetry in the Germline and Early Embryo." *BioRxiv*, January. http://biorxiv.org/content/early/2018/10/04/428540.abstract.

Bogliotti, Yanina S., and Pablo J. Ross. 2012. "Mechanisms of Histone H3 Lysine 27 Trimethylation Remodeling during Early Mammalian Development." *Epigenetics*. doi:10.4161/epi.21615.

Boroviak, Thorsten, Remco Loos, Patrick Lombard, Junko Okahara, R??diger Behr, Erika Sasaki, Jennifer Nichols, Austin Smith, and Paul Bertone. 2015. "Lineage-Specific Profiling Delineates the Emergence and Progression of Naive Pluripotency in Mammalian Embryogenesis." *Developmental Cell* 35 (3): 366–82. doi:10.1016/j.devcel.2015.10.011.

Boysen, Gunnar, Brian F. Pachkowski, Jun Nakamura, and James A. Swenberg. 2009. "The Formation and Biological Significance of N7-Guanine Adducts." *Mutation Research - Genetic Toxicology and Environmental Mutagenesis* 678 (2): 76–94. doi:10.1016/j.mrgentox.2009.05.006.

Brachet, J., H. Denis, and F. de Vitry. 1964. "The Effects of Actinomycin D and Puromycin on Morphogenesis in Amphibian Eggs and Acetabularia Mediterranea." *Developmental Biology*. doi:10.1093/clind/15.Supplement_1.S195.

Braude, Peter, Virginia Bolton, and Stephen Moore. 1988. "Human Gene Expression First Occurs between the Four- and Eight-Cell Stages of Preimplantation Development." *Nature* 332 (6163): 459–61. doi:10.1038/332459a0.

Browne Harvey, Ethel. 1936. "PARTHENOGENETIC MEROGONY OR CLEAVAGE WITH OUT NUCLEI IN ARBACIA PUNCTULATA." *Biological Bulletin*. doi:10.2307/1537411.

Bush, Stephen J., Lu Chen, Jaime M. Tovar-Corona, and Araxi O. Urrutia. 2017. "Alternative Splicing and the Evolution of Phenotypic Novelty." *Philosophical Transactions of the Royal Society B: Biological Sciences*. doi:10.1098/rstb.2015.0474.

Butuči, Melina, Ashley B. Williams, Matthew M. Wong, Brendan Kramer, and W. Matthew Michael. 2015. "Zygotic Genome Activation Triggers Chromosome Damage and Checkpoint Signaling in C. Elegans Primordial Germ Cells." *Developmental Cell* 34 (1): 85–95. doi:10.1016/j.devcel.2015.04.019.

Calegari, F. 2003. "An Inhibition of Cyclin-Dependent Kinases That Lengthens, but Does Not Arrest, Neuroepithelial Cell Cycle Induces Premature Neurogenesis." *Journal of Cell Science*.

doi:10.1242/jcs.00825.

Callebaut, Marc. 2008. “Historical Evolution of Preformistic versus Neoformistic (Epigenetic) Thinking in Embryology.” *Belgian Journal of Zoology*.

Camous, S., Y. Heyman, W. Meziou, and Y. Menezo. 2004. “Cleavage beyond the Block Stage and Survival after Transfer of Early Bovine Embryos Cultured with Trophoblastic Vesicles.” *Reproduction*. doi:10.1530/jrf.0.0720479.

Candeias, Marco M, Masatoshi Hagiwara, and Michiyuki Matsuda. 2016. “Cancer-specific Mutations in P53 Induce the Translation of Δ160p53 Promoting Tumorigenesis.” *EMBO Reports*. doi:10.15252/embr.201541956.

Cao, Fan, Yiwen Fang, Hong Kee Tan, Yufen Goh, Jocelyn Yeen Hui Choy, Bryan Thean Howe Koh, Jiong Hao Tan, et al. 2017. “Super-Enhancers and Broad H3k4me3 Domains Form Complex Gene Regulatory Circuits Involving Chromatin Interactions.” *Scientific Reports*. doi:10.1038/s41598-017-02257-3.

Cao, Hong, Fabiana Garcia, and Mark A. McNiven. 1998. “Differential Distribution of Dynamin Isoforms in Mammalian Cells.” *Molecular Biology of the Cell*. doi:10.1091/mbc.9.9.2595.

Carrillo Oesterreich, Fernando, Stephan Preibisch, and Karla M. Neugebauer. 2010. “Global Analysis of Nascent Rna Reveals Transcriptional Pausing in Terminal Exons.” *Molecular Cell*. doi:10.1016/j.molcel.2010.11.004.

Carroll, John, and Petros Marangos. 2013. “The DNA Damage Response in Mammalian Oocytes.” *Frontiers in Genetics*. doi:10.3389/fgene.2013.00117.

Causier, Barry, Zhen Li, Riet De Smet, James P.B. Lloyd, Yves Van De Peer, and Brendan Davies. 2017. “Conservation of Nonsense-Mediated MRNA Decay Complex Components Throughout Eukaryotic Evolution.” *Scientific Reports*. doi:10.1038/s41598-017-16942-w.

Chang, Judy C, and Yuet Wai Kan. 1979. “B0 Thalassemia, a Nonsense Mutation in Man.” *Proceedings of the National Academy of Science of United States of America*. doi:10.1073/pnas.76.6.2886.

Cheeseman, Liam P., Jérôme Boulanger, Lisa M. Bond, and Melina Schuh. 2016. “Two Pathways Regulate Cortical Granule Translocation to Prevent Polyspermy in Mouse Oocytes.” *Nature Communications*. doi:10.1038/ncomms13726.

Chen, Geng, Jiwei Chen, Jianmin Yang, Long Chen, Xiongfei Qu, Caiping Shi, Baitang Ning, et al. 2017. “Significant Variations in Alternative Splicing Patterns and Expression Profiles between Human-Mouse Orthologs in Early Embryos.” *Science China Life Sciences*. doi:10.1007/s11427-015-0348-5.

Chhangawala, Sagar, Gabe Rudy, Christopher E. Mason, and Jeffrey A. Rosenfeld. 2015. “The Impact of Read Length on Quantification of Differentially Expressed Genes and Splice Junction Detection.” *Genome Biology*. doi:10.1186/s13059-015-0697-y.

Ciccia, Alberto, and Stephen J. Elledge. 2010. “The DNA Damage Response: Making It Safe to Play with Knives.” *Molecular Cell*. doi:10.1016/j.molcel.2010.09.019.

Ciemerych, Maria A., Bernard Maro, and Jacek Z. Kubiak. 1999. “Control of Duration of the First Two Mitoses in a Mouse Embryo.” *Zygote*. doi:10.1017/S0967199499000696.

Cieply, Benjamin, Juw Won Park, Angela Nakauka-Ddamba, Thomas W. Bebee, Yang Guo, Xuequn Shang, Christopher J. Lengner, Yi Xing, and Russ P. Carstens. 2016. “Multiphasic and Dynamic Changes in Alternative Splicing during Induction of Pluripotency Are Coordinated by Numerous RNA-Binding Proteins.” *Cell Reports*. doi:10.1016/j.celrep.2016.03.025.

Coers, Jörn, Christina Ranft, and Radek C. Skoda. 2004. “A Truncated Isoform of C-Mpl with an Essential C-Terminal Peptide Targets the Full-Length Receptor for Degradation.” *Journal of Biological Chemistry*. doi:10.1074/jbc.M401386200.

Condic, Maureen L. 2013. “Totipotency: What It Is and What It Is Not.” *Stem Cells and Development*. doi:10.1089/scd.2013.0364.

Conn, Christopher W., Andrea L. Lewellyn, and James L. Maller. 2004. “The DNA Damage Checkpoint in Embryonic Cell Cycles Is Dependent on the DNA-to-Cytoplasmic Ratio.” *Developmental Cell*. doi:10.1016/j.devcel.2004.07.003.

Cooper, Geoffrey M, and Robert E Hausman. 2007. *The Cell: A Molecular Approach*. *Sinauer Associates*. doi:10.1017/CBO9781107415324.004.

Correa, Bruna R, Patricia Rosa de Araujo, Mei Qiao, Suzanne C Burns, Chen Chen, Richard Schlegel, Seema Agarwal, Pedro A F Galante, and Luiz O F Penalva. 2016. “Functional Genomics Analyses of RNA-Binding Proteins Reveal the Splicing Regulator SNRPB as an Oncogenic Candidate in Glioblastoma.” *Genome Biology* 17 (1): 125. doi:10.1186/s13059-016-0990-4.

Corrie, Pippa G. 2008. “Cytotoxic Chemotherapy: Clinical Aspects.” *Medicine* 36 (1). Elsevier: 24–28. doi:10.1016/j.mpmed.2007.10.012.

Crespo-Hernández, Carlos E., Kimberly De La Harpe, and Bern Kohler. 2008. “Ground-State Recovery Following UV Excitation Is Much Slower in G•C-DNA Duplexes and Hairpins than in Mononucleotides.” *Journal of the American Chemical Society* 130 (33): 10844–45. doi:10.1021/ja802183s.

Crick, F. 1970. “Central Dogma of Molecular Biology.” *Nature*. doi:10.1038/227561a0.

Crick, Francis. 1958. “On Protein Synthesis.” *The Symposia of the Society for Experimental Biology*. doi:10.1038/227561a0.

Cui, Jun, Caroline V. Sartain, Jeffrey A. Pleiss, and Mariana F. Wolfner. 2013. “Cytoplasmic Polyadenylation Is a Major MRNA Regulator during Oogenesis and Egg Activation in Drosophila.” *Developmental Biology*. doi:10.1016/j.ydbio.2013.08.013.

Cui, Peng, Qiang Lin, Feng Ding, Chengqi Xin, Wei Gong, Lingfang Zhang, Jianing Geng, et al. 2010. “A Comparison between Ribo-Minus RNA-Sequencing and PolyA-Selected RNA-Sequencing.” *Genomics*. doi:10.1016/j.ygeno.2010.07.010.

Dahl, John Arne, Inkyung Jung, Håvard Aanes, Gareth D. Greggains, Adeel Manaf, Mads Lerdrup, Guoqiang Li, et al. 2016. “Broad Histone H3K4me3 Domains in Mouse Oocytes Modulate Maternal-to-Zygotic Transition.” *Nature* 537 (7621): 548–52. doi:10.1038/nature19360.

Daniels, Rob, James Adjaye, Virginia Bolton, and Marilyn Monk. 1998. “Detection of a Novel Splice Variant of the Hypoxanthine-Guanine Phosphoribosyl Transferase Gene in Human Oocytes and Preimplantation Embryos: Implications for a RT-PCR-Based Preimplantation Diagnosis of Lesch-Nyhan Syndrome.” *Molecular Human Reproduction*. doi:10.1093/molehr/4.8.785.

Dasika, G K, S C Lin, S Zhao, P Sung, A Tomkinson, and E Y Lee. 1999. “DNA Damage-Induced Cell Cycle Checkpoints and DNA Strand Break Repair in Development and Tumorigenesis.” *Oncogene* 18 (55): 7883–99. doi:10.1038/sj.onc.1203283.

de Lamirande, Eve, and Claude Cagnon. 1993. “Human Sperm Hyperactivation and Capacitation as Parts of an Oxidative Process.” *Free Radical Biology and Medicine*. doi:10.1016/0891-5849(93)90006-G.

De Paepe, Caroline, Greet Cauffman, An Verloes, Johan Sterckx, Paul Devroey, Herman Tournaye, Inge

Liebaers, and Hilde Van de Velde. 2013. “Human Trophectoderm Cells Are Not yet Committed.” *Human Reproduction (Oxford, England)* 28 (3). England: 740–49. doi:10.1093/humrep/des432.

Deng, Qiaolin, Daniel Ramsköld, Björn Reinius, and Rickard Sandberg. 2014. “Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells.” *Science (New York, N.Y.)* 343 (6167): 193–96. doi:10.1126/science.1245316.

Derijck, Alwin A.H.A., Godfried W. van der Heijden, Maud Giele, Marielle E.P. Philippens, Casandra C.A.W. van Bavel, and Peter de Boer. 2006. “ΓH2AX Signalling during Sperm Chromatin Remodelling in the Mouse Zygote.” *DNA Repair*. doi:10.1016/j.dnarep.2006.05.043.

Deutsch, Daniela R., Thomas Fröhlich, Kathrin A. Otte, Andrea Beck, Felix A. Habermann, Eckhard Wolf, and Georg J. Arnold. 2014. “Stage-Specific Proteome Signatures in Early Bovine Embryo Development.” *Journal of Proteome Research*. doi:10.1021/pr500550t.

Di Giammartino, Dafne Campigli, Kensei Nishida, and James L. Manley. 2011. “Mechanisms and Consequences of Alternative Polyadenylation.” *Molecular Cell*. doi:10.1016/j.molcel.2011.08.017.

Dixon, Jesse R., Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S. Liu, and Bing Ren. 2012. “Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions.” *Nature*. doi:10.1038/nature11082.

Dizdaroglu, M. 1992. “Oxidative Damage to DNA in Mammalian Chromatin.” *Mutation Research* 275 (3–6): 331–42. doi:0921-8734(92)90036-O [pii].

Do, Dang Vinh, Bernhard Strauss, Engin Cukuroglu, Iain Macaulay, Keng Boon Wee, Tim Xiaoming Hu, Ruiz De Los Mozos Igor, et al. 2018. “SRSF3 Maintains Transcriptome Integrity in Oocytes by Regulation of Alternative Splicing and Transposable Elements.” *Cell Discovery*. doi:10.1038/s41421-018-0032-3.

Dominguez, Daniel, Peter Freese, Maria S. Alexis, Amanda Su, Myles Hochman, Tsultrim Palden, Cassandra Bazile, et al. 2018. “Sequence, Structure, and Context Preferences of Human RNA Binding Proteins.” *Molecular Cell*. doi:10.1016/j.molcel.2018.05.001.

Donà, G., C. Fiore, E. Tibaldi, F. Frezzato, A. Andrisani, G. Ambrosini, D. Fiorentin, D. Armanini, L. Bordin, and G. Clari. 2011. “Endogenous Reactive Oxygen Species Content and Modulation of Tyrosine Phosphorylation during Sperm Capacitation.” *International Journal of Andrology*. doi:10.1111/j.1365-2605.2010.01097.x.

Ducibella, T., T. Ukena, M. Karnovsky, and E. Anderson. 1977. “Changes in Cell Surface and Cortical Cytoplasmic Organization during Early Embryogenesis in the Preimplantation Mouse Embryo.” *Journal of Cell Biology*. doi:10.1083/jcb.74.1.153.

Duncan, Bruce K., and Jeffrey H. Miller. 1980. “Mutagenic Deamination of Cytosine Residues in DNA.” *Nature*. doi:10.1038/287560a0.

Echard, A., F. J.M. Opdam, H. J.P.C. de Leeuw, F. Jollivet, P. Savelkoul, W. Hendriks, J. Voorberg, B. Goud, and J. A.M. Fransen. 2000. “Alternative Splicing of the Human Rab6A Gene Generates Two Close but Functionally Different Isoforms.” *Molecular Biology of the Cell*. doi:10.1091/mbc.11.11.3819.

Edelmann, Winfried, Paula E. Cohen, Michael Kane, Kirkland Lau, Bernice Morrow, Samuel Bennett, Asad Umar, et al. 1996. “Meiotic Pachytene Arrest in MLH1-Deficient Mice.” *Cell*. doi:10.1016/S0092-8674(00)81312-4.

Eliasson, R. 1966. “Cholesterol in Human Semen.” *Biochem. J*. doi:10.1042/bj0980242.

Fabian, L., X. Xia, D. V. Venkitaramani, K. M. Johansen, J. Johansen, D. J. Andrew, and A. Forer. 2007. "Titin in Insect Spermatocyte Spindle Fibers Associates with Microtubules, Actin, Myosin and the Matrix Proteins Skeletor, Megator and Chromator." *Journal of Cell Science*. doi:10.1242/jcs.03465.

Falco, Geppino, Sung Lim Lee, Ilaria Stanghellini, Uwem C. Bassey, Toshio Hamatani, and Minoru S.H. Ko. 2007. "Zscan4: A Novel Gene Expressed Exclusively in Late 2-Cell Embryos and Embryonic Stem Cells." *Developmental Biology*. doi:10.1016/j.ydbio.2007.05.003.

Fan, Xiaoying, Xiannian Zhang, Xinglong Wu, Hongshan Guo, Yuqiong Hu, Fuchou Tang, and Yanyi Huang. 2015. "Single-Cell RNA-Seq Transcriptome Analysis of Linear and Circular RNAs in Mouse Preimplantation Embryos." *Genome Biology* 16 (1): 1–17. doi:10.1186/s13059-015-0706-1.

Filion, Tera M., Meng Qiao, Prachi N. Ghule, Matthew Mandeville, Andre J. Van Wijnen, Janet L. Stein, Jane B. Lian, Dario C. Altieri, and Gary S. Stein. 2009. "Survival Responses of Human Embryonic Stem Cells to DNA Damage." *Journal of Cellular Physiology*. doi:10.1002/jcp.21735.

Finch-Edmondson, Megan L., Robyn P. Strauss, Joshua S. Clayton, George C. Yeoh, and Bernard A. Callus. 2016. "Splice Variant Insertions in the C-Terminus Impairs YAP's Transactivation Domain." *Biochemistry and Biophysics Reports*. doi:10.1016/j.bbrep.2016.02.015.

Fleming, T, M Hay, Q Javed, and S Citi. 1993. "Localisation of Tight Junction Protein Cingulin Is Temporally and Spatially Regulated during Early Mouse Development." *Development*.

Flyamer, Ilya M., Johanna Gassler, Maxim Imakaev, Hugo B. Brandão, Sergey V. Ulianov, Nezar Abdennur, Sergey V. Razin, Leonid A. Mirny, and Kikuë Tachibana-Konwalski. 2017. "Single-Nucleus Hi-C Reveals Unique Chromatin Reorganization at Oocyte-to-Zygote Transition." *Nature*. doi:10.1038/nature21711.

Fox, Catherine Anne, M D Sheets, E Wahle, and M Wickens. 1992. "Polyadenylation of Maternal MRNA during Oocyte Maturation: Poly(A) Addition in Vitro Requires a Regulated RNA Binding Activity and a Poly(A) Polymerase." *The EMBO Journal*.

Froggatt, N J, C Brassett, D J Koch, D G Evans, S V Hodgson, B A Ponder, and E R Maher. 1996. "Mutation Screening of MSH2 and MLH1 MRNA in Hereditary Non-Polyposis Colon Cancer Syndrome." *Journal of Medical Genetics*. doi:10.1136/jmg.33.9.726.

Fu, Dragony, and Kathleen Collins. 2007. "Purification of Human Telomerase Complexes Identifies Factors Involved in Telomerase Biogenesis and Telomere Length Regulation." *Molecular Cell*. doi:10.1016/j.molcel.2007.09.023.

Futschik, Matthias E, and Bronwyn Carlisle. 2005. "Noise-Robust Soft Clustering of Gene Expression Time-Course Data." *Journal of Bioinformatics and Computational Biology* 3 (4): 965–88. doi:10.1142/S0219720005001375.

Gabut, Mathieu, Payman Samavarchi-Tehrani, Xinchen Wang, Valentina Slobodeniuc, Dave O'Hanlon, Hoon Ki Sung, Manuel Alvarez, et al. 2011. "An Alternative Splicing Switch Regulates Embryonic Stem Cell Pluripotency and Reprogramming." *Cell*. doi:10.1016/j.cell.2011.08.023.

Gaffney, Christian J., Tsutomu Oka, Virginia Mazack, Dror Hilman, Uri Gat, Tomoki Muramatsu, Johji Inazawa, et al. 2012. "Identification, Basic Characterization and Evolutionary Analysis of Differentially Spliced MRNA Isoforms of Human YAP1 Gene." *Gene*. doi:10.1016/j.gene.2012.08.025.

Gao, Yawei, Xiaoyu Liu, Bin Tang, Chong Li, Zhaohui Kou, Lin Li, Wenqiang Liu, et al. 2017. "Protein Expression Landscape of Mouse Embryos during Pre-Implantation Development." *Cell Reports* 21

(13): 3957–69. doi:10.1016/j.celrep.2017.11.111.

García-López, Jesús, and Jesús del Mazo. 2012. "Expression Dynamics of MicroRNA Biogenesis during Preimplantation Mouse Development." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. doi:10.1016/j.bbagrm.2012.03.007.

Gasior, Stephen L., Timothy P. Wakeman, Bo Xu, and Prescott L. Deininger. 2006. "The Human LINE-1 Retrotransposon Creates DNA Double-Strand Breaks." *Journal of Molecular Biology*. doi:10.1016/j.jmb.2006.01.089.

Gassler, Johanna, Hugo B Brandão, Maxim Imakaev, Ilya M Flyamer, Sabrina Ladstätter, Wendy A Bickmore, Jan-Michael Peters, Leonid A Mirny, and Kikuë Tachibana. 2017. "A Mechanism of Cohesin-dependent Loop Extrusion Organizes Zygotic Genome Architecture." *The EMBO Journal*. doi:10.15252/embj.201798083.

Gatewood, J M, G R Cook, R Balhorn, E M Bradbury, and C W Schmid. 1987. "Sequence-Specific Packaging of DNA in Human Sperm Chromatin." *Science (New York, N.Y.)* 236 (4804): 962–64. doi:10.1126/science.3576213.

Gawecka, Joanna E., Joel Marh, Michael Ortega, Yasuhiro Yamauchi, Monika A. Ward, and W. Steven Ward. 2013. "Mouse Zygotes Respond to Severe Sperm DNA Damage by Delaying Paternal DNA Replication and Embryonic Development." *PLoS ONE*. doi:10.1371/journal.pone.0056385.

Ge, Steven Xijin. 2017. "Exploratory Bioinformatics Investigation Reveals Importance of 'Junk' DNA in Early Embryo Development." *BMC Genomics*. doi:10.1186/s12864-017-3566-0.

Gilbert, Scott F. 2000. "Oogenesis." In *Developmental Biology. 6th Edition.*

Giraldez, Antonio J., Yuichiro Mishima, Jason Rihel, Russell J. Grocock, Stijn Van Dongen, Kunio Inoue, Anton J. Enright, and Alexander F. Schier. 2006. "Zebrafish MiR-430 Promotes Deadenylation and Clearance of Maternal MRNAs." *Science*. doi:10.1126/science.1122689.

Goddard, M J, and H P Pratt. 1983. "Control of Events during Early Cleavage of the Mouse Embryo: An Analysis of the '2-Cell Block'." *Journal of Embryology and Experimental Morphology*. doi:10.1080/01425690701737481.

Gohr, André, and Manuel Irimia. 2019. "Matt: Unix Tools for Alternative Splicing Analysis." *Bioinformatics (Oxford, England)*. doi:10.1093/bioinformatics/bty606.

Goll, Mary Grace, and Timothy H. Bestor. 2005. "EUKARYOTIC CYTOSINE METHYLTRANSFERASES." *Annual Review of Biochemistry*. doi:10.1146/annurev.biochem.74.010904.153721.

Graf, Alexander, Stefan Krebs, Valeri Zakhartchenko, Björn Schwalb, Helmut Blum, and Eckhard Wolf. 2014. "Fine Mapping of Genome Activation in Bovine Embryos by RNA Sequencing." *Proceedings of the National Academy of Sciences* 111 (11): 4139–44. doi:10.1073/pnas.1321569111.

Graves, Robert J., Benjamin F.L. Li, and Peter F. Swann. 1989. "Repair of O6-Methylguanine, O6-Ethylguanine, O6-Isopropylguanine, and O4-Methylthymine in Synthetic Oligodeoxynucleotides by Escherichia Coli Ada Gene O6-Alkylguanine-DNA-Alkyltransferase." *Carcinogenesis*. doi:10.1093/carcin/10.4.661.

Green, Justin B., Cary D. Gardner, Robin P. Wharton, and Aneel K. Aggarwal. 2003. "RNA Recognition via the SAM Domain of Smaug." *Molecular Cell*. doi:10.1016/S1097-2765(03)00178-3.

Grunwald, Hannah A., Valentino M. Gantz, Gunnar Poplawski, Xiang Ru S. Xu, Ethan Bier, and Kimberly

L. Cooper. 2019. “Super-Mendelian Inheritance Mediated by CRISPR–Cas9 in the Female Mouse Germline.” *Nature*. doi:10.1038/s41586-019-0875-2.

Gu, Ling, Qiang Wang, and Qing Yuan Sun. 2010. “Histone Modifications during Mammalian Oocyte Maturation: Dynamics, Regulation and Functions.” *Cell Cycle*. doi:10.4161/cc.9.10.11599.

Guilgur, Leonardo Gastón, Pedro Prudêncio, Daniel Sobral, Denisa Liszekova, André Rosa, and Rui Gonçalo Martinho. 2014. “Requirement for Highly Efficient Pre-MRNA Splicing during Drosophila Early Embryonic Development.” *ELife*. doi:10.7554/eLife.02181.001.

Guirouilh-Barbat, Josée, Sylvie Huck, Pascale Bertrand, Livia Pirzio, Chantal Desmaze, Laure Sabatier, and Bernard S. Lopez. 2004. “Impact of the KU80 Pathway on NHEJ-Induced Genome Rearrangements in Mammalian Cells.” *Molecular Cell*. doi:10.1016/j.molcel.2004.05.008.

Guo, G, F von Meyenn, M Rostovskaya, J Clarke, S Dietmann, D Baker, A Sahakyan, et al. 2017. “Epigenetic Resetting of Human Pluripotency.” *Development*. doi:10.1242/dev.146811.

Guo, Guoji, Mikael Huss, Guo Qing Tong, Chaoyang Wang, Li Li Sun, Neil D. Clarke, and Paul Robson. 2010. “Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote to Blastocyst.” *Developmental Cell* 18 (4): 675–85. doi:10.1016/j.devcel.2010.02.012.

Gurdon, J. B. 1962. “The Developmental Capacity of Nuclei Taken from Intestinal Epithelium Cells of Feeding Tadpoles.” *Journal of Embryology and Experimental Morphology*. doi:10.1002/stem.684.

Haber, James E. 1999. “DNA Recombination: The Replication Connection.” *Trends in Biochemical Sciences*. doi:10.1016/S0968-0004(99)01413-9.

Hammoud, Saher Sue, David A. Nix, Haiying Zhang, Jahnvi Purwar, Douglas T. Carrell, and Bradley R. Cairns. 2009. “Distinctive Chromatin in Human Sperm Packages Genes for Embryo Development.” *Nature* 460 (7254): 473–78. doi:10.1038/nature08162.

Han, Hong, Manuel Irimia, P. Joel Ross, Hoon Ki Sung, Babak Alipanahi, Laurent David, Azadeh Golipour, et al. 2013. “MBNL Proteins Repress ES-Cell-Specific Alternative Splicing and Reprogramming.” *Nature*. doi:10.1038/nature12270.

Handyside, A H. 1978. “Time of Commitment of inside Cells Isolated from Preimplantation Mouse Embryos.” *J. Embryol. Exp. Morph*.

Harton, Gary L., and Helen G. Tempest. 2012. “Chromosomal Disorders and Male Infertility.” *Asian Journal of Andrology*. doi:10.1038/aja.2011.66.

Hartshorne, G. M., S. Lyrakou, H. Hamoda, E. Oloto, and F. Ghafari. 2009. “Oogenesis and Cell Death in Human Prenatal Ovaries: What Are the Criteria for Oocyte Selection?” *Molecular Human Reproduction*. doi:10.1093/molehr/gap055.

Herr, Patrick, Bastiaan Evers, Christina Bauerschmidt, Daniel Ebner, Cecilia Lundin, Timea Palmai-Pallag, Oliver Mortusewicz, et al. 2015. “A Genome-Wide IR-Induced RAD51 Foci RNAi Screen Identifies CDC73 Involved in Chromatin Remodeling for DNA Repair.” *Cell Discovery*. doi:10.1038/celldisc.2015.34.

Heyn, Patricia, Martin Kircher, Andreas Dahl, Janet Kelso, Pavel Tomancak, Alex T. Kalinka, and Karla M. Neugebauer. 2014. “The Earliest Transcribed Zygotic Genes Are Short, Newly Evolved, and Different across Species.” *Cell Reports*. doi:10.1016/j.celrep.2013.12.030.

Hindley, Christopher, and Anna Philpott. 2013. “The Cell Cycle and Pluripotency.” *Biochemical Journal*. doi:10.1042/BJ20121627.

Hinrichs, A S, D Karolchik, Robert Baertsch, G P Barber, Gill Bejerano, H Clawson, M Diekhans, et al.

2006. “The UCSC Genome Browser Database: Update 2006.” *Nucleic Acids Research*. doi:10.1093/nar/gkj144.

Hnilicová, Jarmila, Samira Hozeifi, Eva Dušková, Jaroslav Icha, Tereza Tománková, and David Staněk. 2011. “Histone Deacetylase Activity Modulates Alternative Splicing.” *PLoS ONE*. doi:10.1371/journal.pone.0016727.

Ho, L., R. Jothi, J. L. Ronan, G. R. Crabtree, K. Zhao, and K. Cui. 2009. “An Embryonic Stem Cell Chromatin Remodeling Complex, EsBAF, Is an Essential Component of the Core Pluripotency Transcriptional Network.” *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.0812888106.

Holway, Antonia H., Seung Hwan Kim, Adriana La Volpe, and W. Matthew Michael. 2006. “Checkpoint Silencing during the DNA Damage Response in Caenorhabditis Elegans Embryos.” *Journal of Cell Biology*. doi:10.1083/jcb.200512136.

Hosseini, S. M., I. Dufort, J. Caballero, F. Moulavi, H. R. Ghanaei, and M. A. Sirard. 2015. “Transcriptome Profiling of Bovine Inner Cell Mass and Trophectoderm Derived from in Vivo Generated Blastocysts.” *BMC Developmental Biology*. doi:10.1186/s12861-015-0096-3.

Houbaviy, Hristo B., Michael F. Murray, and Phillip A. Sharp. 2003. “Embryonic Stem Cell-Specific MicroRNAs.” *Developmental Cell*. doi:10.1016/S1534-5807(03)00227-2.

Houghton, Franchesca D. 2006. “Energy Metabolism of the Inner Cell Mass and Trophectoderm of the Mouse Blastocyst.” *Differentiation*. doi:10.1111/j.1432-0436.2006.00052.x.

Houghton, Franchesca D., Peter G. Humpherson, Judith A. Hawkhead, Christine J. Hall, and Henry J. Leese. 2003. “Na+, K+, ATPase Activity in the Human and Bovine Preimplantation Embryo.” *Developmental Biology*. doi:10.1016/j.ydbio.2003.07.014.

Houseley, Jonathan, and David Tollervey. 2009. “The Many Pathways of RNA Degradation.” *Cell*. doi:10.1016/j.cell.2009.01.019.

Huang, Wen, and Hasan Khatib. 2010. “Comparison of Transcriptomic Landscapes of Bovine Embryos Using RNA-Seq.” *BMC Genomics*. doi:10.1186/1471-2164-11-711.

Huang, Xian Ju, Xuguang Wang, Xueshan Ma, Shao Chen Sun, Xiaolong Zhou, Chengcheng Zhu, and Honglin Liu. 2014. “EZH2 Is Essential for Development of Mouse Preimplantation Embryos.” *Reproduction, Fertility and Development*. doi:10.1071/RD13169.

Hubé, Florent, and Claire Francastel. 2015. “Mammalian Introns: When the Junk Generates Molecular Diversity.” *International Journal of Molecular Sciences*. doi:10.3390/ijms16034429.

Huff, Jason T., and Daniel Zilberman. 2014. “Dnmt1-Independent CG Methylation Contributes to Nucleosome Positioning in Diverse Eukaryotes.” *Cell*. doi:10.1002/ejic.201600990.

Hug, Clemens B., Alexis G. Grimaldi, Kai Kruse, and Juan M. Vaquerizas. 2017. “Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription.” *Cell*. doi:10.1016/j.cell.2017.03.024.

Hug, Clemens B., and Juan M. Vaquerizas. 2018. “The Birth of the 3D Genome during Early Embryonic Development.” *Trends in Genetics*. doi:10.1016/j.tig.2018.09.002.

Hwang, Young Sun, Minseok Seo, Sohyun Bang, Heebal Kim, and Jae Yong Han. 2018. “Transcriptional and Translational Dynamics during Maternal-To-Zygotic Transition in Early Chicken Development.” *FASEB Journal*. doi:10.1096/fj.201700955R.

Ichijima, Yosuke, Ryo Sakasai, Naoyuki Okita, Kinji Asahina, Shuki Mizutani, and Hirobumi Teraoka.

2005. “Phosphorylation of Histone H2AX at M Phase in Human Cells without DNA Damage Response.” *Biochemical and Biophysical Research Communications*. doi:10.1016/j.bbrc.2005.08.164.

Ikawa, Masahito, Naokazu Inoue, Adam M. Benham, and Masaru Okabe. 2010. “Fertilization: A Sperm’s Journey to and Interaction with the Oocyte.” *Journal of Clinical Investigation*. doi:10.1172/JCI41585.

Inoue, Azusa, Shogo Matoba, and Yi Zhang. 2012. “Transcriptional Activation of Transposable Elements in Mouse Zygotes Is Independent of Tet3-Mediated 5-Methylcytosine Oxidation.” *Cell Research*. doi:10.1038/cr.2012.160.

Inoue, Naokazu, Masahito Ikawa, Ayako Isotani, and Masaru Okabe. 2005. “The Immunoglobulin Superfamily Protein Izumo Is Required for Sperm to Fuse with Eggs.” *Nature*. doi:10.1038/nature03362.

Irimia, Manuel, and Benjamin J. Blencowe. 2012. “Alternative Splicing: Decoding an Expansive Regulatory Layer.” *Current Opinion in Cell Biology*. doi:10.1016/j.ceb.2012.03.005.

Irimia, Manuel, Robert J. Weatheritt, Jonathan D. Ellis, Neelroop N. Parikshak, Thomas Gonatopoulos-Pournatzis, Mariana Babor, Mathieu Quesnel-Vallières, et al. 2014. “A Highly Conserved Program of Neuronal Microexons Is Misregulated in Autistic Brains.” *Cell* 159 (7). Elsevier: 1511–23. doi:10.1016/j.cell.2014.11.035.

Ivashkevich, Alesia N., Olga A. Martin, Andrea J. Smith, Christophe E. Redon, William M. Bonner, Roger F. Martin, and Pavel N. Lobachevsky. 2011. “ΓH2AX Foci as a Measure of DNA Damage: A Computational Approach to Automatic Analysis.” *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*. doi:10.1016/j.mrfmmm.2010.12.015.

Izsvák, Zsuzsanna, Yongming Wang, and Zoltán Ivics. 2009. “Interactions of Transposons with the Cellular DNA Repair Machinery.” *Genome Dynamics and Stability*. doi:10.1007/7050_2008_043.

Jalkanen, Aimee L., Stephen J. Coleman, and Jeffrey Wilusz. 2014. “Determinants and Implications of MRNA Poly(A) Tail Size - Does This Protein Make My Tail Look Big?” *Seminars in Cell and Developmental Biology*. doi:10.1016/j.semcdb.2014.05.018.

Jaroudi, Souraya, and Sioban SenGupta. 2007. “DNA Repair in Mammalian Embryos.” *Mutation Research - Reviews in Mutation Research*. doi:10.1016/j.mrrev.2006.09.002.

Ji, Z., J. Y. Lee, Z. Pan, B. Jiang, and B. Tian. 2009. “Progressive Lengthening of 3’ Untranslated Regions of MRNAs by Alternative Polyadenylation during Mouse Embryonic Development.” *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.0900028106.

Johnson, Martin H. 2007. “No Title.” In *Essential Reproduction*, 188.

Jukam, David, S. Ali M. Shariati, and Jan M. Skotheim. 2017. “Zygotic Genome Activation in Vertebrates.” *Developmental Cell*. doi:10.1016/j.devcel.2017.07.026.

Jumaa, Hassan, Grace Wei, and Peter J. Nielsen. 1999. “Blastocyst Formation Is Blocked in Mouse Embryos Lacking the Splicing Factor SRp20.” *Current Biology*. doi:10.1016/S0960-9822(99)80394-7.

Kalmbach, Keri, Leroy G. Robinson, Fang Wang, Lin Liu, and David Keefe. 2014. “Telomere Length Reprogramming in Embryos and Stem Cells.” *BioMed Research International*. doi:10.1155/2014/925121.

Kane, Donald A, and Charles B Kimmel. 1993. “The Zebrafish Midblastula Transition.” *Development*.

Kanitz, Alexander, Afzal Pasha Syed, Keisuke Kaji, and Mihaela Zavolan. 2019. “Conserved Regulation of RNA Processing in Somatic Cell Reprogramming.” *BMC Genomics*. doi:10.1186/s12864-019-5438-2.

Kano, Hiroki, Irene Godoy, Christine Courtney, Melissa R. Vetter, George L. Gerton, Eric M. Ostertag, and Haig H. Kazazian. 2009. “L1 Retrotransposition Occurs Mainly in Embryogenesis and Creates Somatic Mosaicism.” *Genes and Development*. doi:10.1101/gad.1803909.

Karimi, Mobin A., Oscar A. Aguilar, Baixiang Zou, Michael H. Bachmann, James R. Carlyle, Cynthia L. Baldwin, and Taku Kambayashi. 2014. “A Truncated Human NKG2D Splice Isoform Negatively Regulates NKG2D-Mediated Function.” *The Journal of Immunology*. doi:10.4049/jimmunol.1400920.

Kasowitz, Seth D., Jun Ma, Stephen J. Anderson, N. Adrian Leu, Yang Xu, Brian D. Gregory, Richard M. Schultz, and P. Jeremy Wang. 2018. “Nuclear M6A Reader YTHDC1 Regulates Alternative Polyadenylation and Splicing during Mouse Oocyte Development.” *PLoS Genetics*. doi:10.1371/journal.pgen.1007412.

Ke, Yuwen, Yanan Xu, Xuepeng Chen, Songjie Feng, Zhenbo Liu, Yaoyu Sun, Xuelong Yao, et al. 2017. “3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis.” *Cell*. doi:10.1016/j.cell.2017.06.029.

Kim, Daehwan, Geo Pertea, Cole Trapnell, Harold Pimentel, Ryan Kelley, and Steven L Salzberg. 2013. “TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions.” *Genome Biology* 14 (4): R36. doi:10.1186/gb-2013-14-4-r36.

Kim, Euikyung, Lei Du, David B. Bregman, and Stephen L. Warren. 1997. “Splicing Factors Associate with Hyperphosphorylated RNA Polymerase II in the Absence of Pre-MRNA.” *Journal of Cell Biology*. doi:10.1083/jcb.136.1.19.

Kim, Jin Moon, Honglin Liu, Mayuko Tazaki, Masao Nagata, and Fugaku Aoki. 2003. “Changes in Histone Acetylation during Mouse Oocyte Meiosis.” *Journal of Cell Biology*. doi:10.1083/jcb.200303047.

Kim, Shinseog, Ufuk Günesdogan, Jan J. Zylicz, Jamie A. Hackett, Delphine Cougot, Siqin Bao, Caroline Lee, et al. 2014. “PRMT5 Protects Genomic Integrity during Global DNA Demethylation in Primordial Germ Cells and Preimplantation Embryos.” *Molecular Cell*. doi:10.1016/j.molcel.2014.10.003.

Kleinridders, Andre, Hans-Martin Pogoda, Sigrid Irlenbusch, Neil Smyth, Csaba Koncz, Matthias Hammerschmidt, and Jens C Bruning. 2009. “PLRG1 Is an Essential Regulator of Cell Proliferation and Apoptosis during Vertebrate Development and Tissue Homeostasis.” *Molecular and Cellular Biology*.

Ko, M. S.H. 2016. “Zygotic Genome Activation Revisited: Looking Through the Expression and Function of Zscan4.” In *Current Topics in Developmental Biology*. doi:10.1016/bs.ctdb.2016.04.004.

Kondo, Natsuko, Akihisa Takahashi, Koji Ono, and Takeo Ohnishi. 2010. “DNA Damage Induced by Alkylating Agents and Repair Pathways.” *Journal of Nucleic Acids*. doi:10.4061/2010/543531.

Koyama, H, H Suzuki, X Yang, S Jiang, and R H Foote. 1994. “Analysis of Polarity of Bovine and Rabbit Embryos by Scanning Electron Microscopy.” *Biology of Reproduction*. doi:10.1095/biolreprod50.1.163.

Kumar, Lokesh, and Matthias E Futschik. 2007. “Mfuzz: A Software Package for Soft Clustering of

Microarray Data." *Bioinformation* 2 (1): 5–7. doi:10.6026/97320630002005.

Kuo, Alex J., Peggie Cheung, Kaifu Chen, Barry M. Zee, Mitomu Kioi, Josh Lauring, Yuanxin Xi, et al. 2011. "NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming." *Molecular Cell*. doi:10.1016/j.molcel.2011.08.042.

Kurosaki, T., and L. E. Maquat. 2016. "Nonsense-Mediated MRNA Decay in Humans at a Glance." *Journal of Cell Science*. doi:10.1242/jcs.181008.

Labbé, Roselyne M., Manuel Irimia, Ko W. Currie, Alexander Lin, Shu Jun Zhu, David D R Brown, Eric J. Ross, et al. 2012. "A Comparative Transcriptomic Analysis Reveals Conserved Features of Stem Cell Pluripotency in Planarians and Mammals." *Stem Cells* 30 (8): 1734–45. doi:10.1002/stem.1144.

Labeit, Siegfried, Bernhard Kolmerer, and Wolfgang A. Linke. 1997. "The Giant Protein Titin: Emerging Roles in Physiology and Pathophysiology." *Circulation Research*. doi:10.1161/01.RES.80.2.290.

Laberge, Rémi-Martin, and Guylain Boissonneault. 2005. "On the Nature and Origin of DNA Strand Breaks in Elongating Spermatids1." *Biology of Reproduction*. doi:10.1095/biolreprod.104.036939.

Ladstätter, Sabrina, and Kikuë Tachibana-Konwalski. 2016. "A Surveillance Mechanism Ensures Repair of DNA Lesions during Zygotic Reprogramming." *Cell*. doi:10.1016/j.cell.2016.11.009.

Laporta, J, a Driver, and H Khatib. 2011. "Short Communication: Expression and Alternative Splicing of POU1F1 Pathway Genes in Preimplantation Bovine Embryos." *Journal of Dairy Science*. doi:10.3168/jds.2011-4144.

Latham, Keith E, James I Garrels, Cecile Chang, and Davor Solter. 1991. "Quantitative Analysis of Protein Synthesis in Mouse Embryos. I. Extensive Reprogramming at the One-and Two-Cell Stages." *Development*. doi:10.1017/S0007114508047685.

Lawrimore, Josh, Timothy M. Barry, Raymond M. Barry, Alyssa C. York, Brandon Friedman, Diana M. Cook, Kristen Akialis, et al. 2017. "Microtubule Dynamics Drive Enhanced Chromatin Motion and Mobilize Telomeres in Response to DNA Damage." *Molecular Biology of the Cell* 28 (12): 1701–11. doi:10.1091/mbc.E16-12-0846.

Lee, Amy, and Ann A. Kiessling. 2017. "Early Human Embryos Are Naturally Aneuploid—Can That Be Corrected?" *Journal of Assisted Reproduction and Genetics*. doi:10.1007/s10815-016-0845-7.

Lemmers, Richard J L F, Daniel G Miller, and Silvere M van der Maarel. 1993. "Facioscapulohumeral Muscular Dystrophy." In , edited by Margaret P Adam, Holly H Ardinger, Roberta A Pagon, Stephanie E Wallace, Lora J H Bean, Karen Stephens, and Anne Amemiya. Seattle (WA).

Lepper, Christopher P., Martin A.K. Williams, David Penny, Patrick J.B. Edwards, and Geoffrey B. Jameson. 2018. "Effects of Pressure and PH on the Hydrolysis of Cytosine: Implications for Nucleotide Stability around Deep-Sea Black Smokers." *ChemBioChem* 19 (6): 540–44. doi:10.1002/cbic.201700555.

Li, Chao Bo, Zhen Dong Wang, Zhong Zheng, Li Li Hu, Shu Qi Zhong, and Lei Lei. 2011. "Number of Blastomeres and Distribution of Microvilli in Cloned Mouse Embryos during Compaction." *Zygote*. doi:10.1017/S0967199410000377.

Li, Junling, Zhiliang Chen, Lian Yong Gao, Angelo Colorni, Michal Ucko, Shengyun Fang, and Shao Jun Du. 2015. "A Transgenic Zebrafish Model for Monitoring Xbp1 Splicing and Endoplasmic Reticulum Stress in Vivo." *Mechanisms of Development*. doi:10.1016/j.mod.2015.04.001.

Li, Wentao, Jinchuan Hu, Ogun Adebali, Sheera Adar, Yanyan Yang, Yi-Ying Chiou, and Aziz Sancar.

2017. “Human Genome-Wide Repair Map of DNA Damage Caused by the Cigarette Smoke Carcinogen Benzo[a]Pyrene.” *Proceedings of the National Academy of Sciences*, 201706021. doi:10.1073/pnas.1706021114.

Li, Yuan, Li Xinxin, Feng Jiaojiao, Yin Chenyang, Yuan Fang, and Wang Xinru. 2016. “IRE1α Is Essential for Xenopus Pancreas Development.” *Journal of Biomedical Research*. doi:10.7555/jbr.28.20130076.

Lim, Jaechul, Mihye Lee, Ahyeon Son, Hyeshik Chang, and V. Narry Kim. 2016. “MTAIL-Seq Reveals Dynamic Poly(A) Tail Regulation in Oocyte-to-Embryo Development.” *Genes and Development*. doi:10.1101/gad.284802.116.

Lima, Sarah Azoubel, Laura B. Chipman, Angela L. Nicholson, Ying Hsin Chen, Brian A. Yee, Gene W. Yeo, Jeff Coller, and Amy E. Pasquinelli. 2017. “Short Poly(A) Tails Are a Conserved Feature of Highly Expressed Genes.” *Nature Structural and Molecular Biology*. doi:10.1038/nsmb.3499.

Lin, Chih Jen, Fong Ming Koh, Priscilla Wong, Marco Conti, and Miguel Ramalho-Santos. 2014. “Hira-Mediated H3.3 Incorporation Is Required for DNA Replication and Ribosomal RNA Transcription in the Mouse Zygote.” *Developmental Cell*. doi:10.1016/j.devcel.2014.06.022.

Lin, K.-T., R.-M. Lu, and W.-Y. Tarn. 2004. “The WW Domain-Containing Proteins Interact with the Early Spliceosome and Participate in Pre-MRNA Splicing In Vivo.” *Molecular and Cellular Biology*. doi:10.1128/MCB.24.20.9176-9185.2004.

Lindahl, Tomas, and Barbro Nyberg. 1974. “Heat-Induced Deamination of Cytosine Residues in Deoxyribonucleic Acid.” *Biochemistry* 13 (16): 3405–10. doi:10.1021/bi00713a035.

Lindeboom, Rik G.H., Fran Supek, and Ben Lehner. 2016. “The Rules and Impact of Nonsense-Mediated MRNA Decay in Human Cancers.” *Nature Genetics*. doi:10.1038/ng.3664.

Liu, Lin, Susan M. Bailey, Maja Okuka, Purificación Muñoz, Chao Li, Lingjun Zhou, Chao Wu, et al. 2007. “Telomere Lengthening Early in Development.” *Nature Cell Biology*. doi:10.1038/ncb1664.

Liu, Ming Jung, Alexander E. Seddon, Zing Tsung Yeh Tsai, Ian T. Major, Monique Floer, Gregg A. Howe, and Shin Han Shiu. 2015. “Determinants of Nucleosome Positioning and Their Influence on Plant Gene Expression.” *Genome Research*. doi:10.1101/gr.188680.114.

Liu, W.-M., R. T. K. Pang, P. C. N. Chiu, B. P. C. Wong, K. Lao, K.-F. Lee, and W. S. B. Yeung. 2012. “Sperm-Borne MicroRNA-34c Is Required for the First Cleavage Division in Mouse.” *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1110368109.

Liu, Xiaoyu, Chenfei Wang, Wenqiang Liu, Jingyi Li, Chong Li, Xiaochen Kou, Jiayu Chen, et al. 2016. “Distinct Features of H3K4me3 and H3K27me3 Chromatin Domains in Pre-Implantation Embryos.” *Nature*. doi:10.1038/nature19362.

Liu, Xinjian, Yujun He, Fang Li, Qian Huang, Takamitsu A. Kato, Russell P. Hall, and Chuan Yuan Li. 2015. “Caspase-3 Promotes Genetic Instability and Carcinogenesis.” *Molecular Cell*. doi:10.1016/j.molcel.2015.03.003.

Liu, Y.-W., M. C. Surka, T. Schroeter, V. Lukiyanchuk, and S. L. Schmid. 2008. “Isoform and Splice-Variant Specific Functions of Dynamin-2 Revealed by Analysis of Conditional Knock-Out Cells.” *Molecular Biology of the Cell*. doi:10.1091/mbc.E08-08-0890.

Lu, Falong, and Yi Zhang. 2015. “Cell Totipotency: Molecular Features, Induction, and Maintenance.” *National Science Review*. doi:10.1093/nsr/nwv009.

Lu, Xukun, Zhen-Ao Zhao, Xiaoqing Wang, Xiaoxin Zhang, Yanhua Zhai, Wenbo Deng, Zhaohong Yi,

and Lei Li. 2018. “Whole-Transcriptome Splicing Profiling of E7.5 Mouse Primary Germ Layers Reveals Frequent Alternative Promoter Usage during Mouse Early Embryogenesis.” *Biology Open*. doi:10.1242/bio.032508.

Lucas, Bronwyn A., Eitan Lavi, Lily Shiue, Hana Cho, Sol Katzman, Keita Miyoshi, Mikiko C. Siomi, Liran Carmel, Manuel Ares, and Lynne E. Maquat. 2018. “Evidence for Convergent Evolution of SINE-Directed Staufen-Mediated MRNA Decay.” *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1715531115.

Ma, Rujun, Jiaqi Zhang, Xiaohui Liu, Ling Li, Honglin Liu, Rong Rui, Ling Gu, and Qiang Wang. 2016. “Involvement of Rab6a in Organelle Rearrangement and Cytoskeletal Organization during Mouse Oocyte Maturation.” *Scientific Reports*. doi:10.1038/srep23560.

Macosko, Evan Z., Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, et al. 2015. “Highly Parallel Genome-Wide Expression Profiling of Individual Cells Using Nanoliter Droplets.” *Cell*. doi:10.1016/j.cell.2015.05.002.

Maeso, Ignacio, Thomas L Dunwell, Chris D R Wyatt, Ferdinand Marlétaz, Borbála Vető, Juan A Bernal, Shan Quah, Manuel Irimia, and Peter W H Holland. 2016. “Evolutionary Origin and Functional Divergence of Totipotent Cell Homeobox Genes in Eutherian Mammals.” *BMC Biology* 14 (1): 45. doi:10.1186/s12915-016-0267-0.

Makishima, Hideki, Valeria Visconte, Hirotoshi Sakaguchi, Anna M. Jankowska, Sarah Abu Kar, Andres Jerez, Bartlomiej Przychodzen, et al. 2012. “Mutations in the Spliceosome Machinery, a Novel and Ubiquitous Pathway in Leukemogenesis.” *Blood*. doi:10.1182/blood-2011-12-399774.

Manzanares, Miguel, and Tristan A. Rodriguez. 2013. “Development: Hippo Signalling Turns the Embryo inside Out.” *Current Biology*. doi:10.1016/j.cub.2013.05.064.

Mao, Zhiyong, Michael Bozzella, Andrei Seluanov, and Vera Gorbunova. 2008. “DNA Repair by Nonhomologous End Joining and Homologous Recombination during Cell Cycle in Human Cells.” *Cell Cycle (Georgetown, Tex.)*. doi:10.2964/jsik.kuni0223.

Maquat, Lynne E., Woan Yuh Tarn, and Olaf Isken. 2010. “The Pioneer Round of Translation: Features and Functions.” *Cell*. doi:10.1016/j.cell.2010.07.022.

Maquat, Lynne E, a J Kinniburgh, E a Rachmilewitz, and J Ross. 1981. “Unstable Beta-Globin MRNA in MRNA-Deficient Beta o Thalassemia.” *Cell*. doi:0092-8674(81)90396-2 [pii].

Marangos, Petros, and John Carroll. 2012. “Oocytes Progress beyond Prophase in the Presence of DNA Damage.” *Current Biology*. doi:10.1016/j.cub.2012.03.063.

Marangos, Petros, Michelle Stevense, Konstantina Niaka, Michaela Lagoudaki, Ibtissem Nabti, Rolf Jessberger, and John Carroll. 2015. “DNA Damage-Induced Metaphase i Arrest Is Mediated by the Spindle Assembly Checkpoint and Maternal Age.” *Nature Communications*. doi:10.1038/ncomms9706.

Marchetto, Maria C.N., Iñigo Narvaiza, Ahmet M. Denli, Christopher Benner, Thomas A. Lazzarini, Jason L. Nathanson, Apuã C.M. Paquola, et al. 2013. “Differential L1 Regulation in Pluripotent Stem Cells of Humans and Apes.” *Nature*. doi:10.1038/nature12686.

Marieb, Elaine. 2013. *No Title*. *Anatomy & Physiology*.

Martin, Kelsey C., and Anne Ephrussi. 2009. “MRNA Localization: Gene Expression in the Spatial Dimension.” *Cell*. doi:10.1016/j.cell.2009.01.044.

McCarthy, B. J., and J. J. Holland. 1965. “Denatured DNA as a Direct Template for in Vitro Protein

Synthesis." *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.54.3.880.

McCulloch, Scott D., and Thomas A. Kunkel. 2008. "The Fidelity of DNA Synthesis by Eukaryotic Replicative and Translesion Synthesis Polymerases." *Cell Research*. doi:10.1038/cr.2008.4.

McIlwain, D. R., Q. Pan, P. T. Reilly, A. J. Elia, S. McCracken, A. C. Wakeham, A. Itie-Youten, B. J. Blencowe, and T. W. Mak. 2010. "Smg1 Is Required for Embryogenesis and Regulates Diverse Genes via Alternative Splicing Coupled to Nonsense-Mediated MRNA Decay." *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1007336107.

Ménézo, Yves, Brian Dale, and Marc Cohen. 2010. "DNA Damage and Repair in Human Oocytes and Embryos: A Review." *Zygote (Cambridge, England)*. doi:10.1017/S0967199410000286.

Merkhofer, Evan C., Peter Hu, and Tracy L. Johnson. 2014. "Introduction to Cotranscriptional RNA Splicing." *Methods in Molecular Biology*. doi:10.1007/978-1-62703-980-2_6.

Montes, M., A. Cloutier, N. Sanchez-Hernandez, L. Michelle, B. Lemieux, M. Blanchette, C. Hernandez-Munain, B. Chabot, and C. Sune. 2012. "TCERG1 Regulates Alternative Splicing of the Bcl-x Gene by Modulating the Rate of RNA Polymerase II Transcription." *Molecular and Cellular Biology*. doi:10.1128/MCB.06255-11.

Moore, J K, and J E Haber. 1996. "Cell Cycle and Genetic Requirements of Two Pathways of Nonhomologous End-Joining Repair of Double-Strand Breaks in Saccharomyces Cerevisiae." *Molecular and Cellular Biology*. doi:10.1128/MCB.16.5.2164.

Mosley, Jonathan D., and Ruth A. Keri. 2006. "Splice Variants of MIAP1 Have an Enhanced Ability to Inhibit Apoptosis." *Biochemical and Biophysical Research Communications*. doi:10.1016/j.bbrc.2006.07.176.

Mu, Weipeng, Joshua Starmer, Della Yee, and Terry Magnuson. 2018. "EZH2 Variants Differentially Regulate Polycomb Repressive Complex 2 in Histone Methylation and Cell Differentiation." *Epigenetics & Chromatin* 11 (1): 71. doi:10.1186/s13072-018-0242-9.

Mu, X. F., X. L. Jin, M. M. J. Farnham, Y. Li, and C. O'Neill. 2011. "DNA Damage-Sensing Kinases Mediate the Mouse 2-Cell Embryo's Response to Genotoxic Stress." *Biology of Reproduction*. doi:10.1095/biolreprod.110.089334.

Muller, F. 2000. "The Nature and Mechanism of Superoxide Production by the Electron Transport Chain: Its Relevance to Aging." *Journal of the American Aging Association* 23 (4): 227–53. doi:10.1007/s11357-000-0022-9.

Mummery, C. 2003. "Differentiation of Human Embryonic Stem Cells to Cardiomyocytes: Role of Coculture With Visceral Endoderm-Like Cells." *Circulation*. doi:10.1161/01.CIR.0000068356.38592.68.

Muñoz, Manuel J., M. Soledad Pérez Santangelo, Maria P. Paronetto, Manuel de la Mata, Federico Pelisch, Stéphanie Boireau, Kira Glover-Cutter, et al. 2009. "DNA Damage Regulates Alternative Splicing through Inhibition of RNA Polymerase II Elongation." *Cell*. doi:10.1016/j.cell.2009.03.010.

Murray, Johanne M., and Antony M. Carr. 2018. "Integrating DNA Damage Repair with the Cell Cycle." *Current Opinion in Cell Biology*. doi:10.1016/j.ceb.2018.03.006.

Narykov, Oleksandr, Nathan Johnson, and Dmitry Korkin. 2018. "Determining Rewiring Effects of Alternatively Spliced Isoform s on Protein - Protein Interaction s Using a Computational Approach." *BioRxiv*. doi:10.1101/256834.

Newcombe, Robert G. 1998. "Interval Estimation for the Difference between Independent Proportions:

Comparison of Eleven Methods." *Statistics in Medicine*. doi:10.1002/(SICI)1097-0258(19980430)17:8<873::AID-SIM779>3.0.CO;2-I.

Ni, K., A. N. Spiess, H. C. Schuppe, and K. Steger. 2016. "The Impact of Sperm Protamine Deficiency and Sperm DNA Damage on Human Male Fertility: A Systematic Review and Meta-Analysis." *Andrology*. doi:10.1111/andr.12216.

Nitiss, John L. 1994. "Roles of DNA Topoisomerases in Chromosomal Replication and Segregation." *Advances in Pharmacology* 29 (PA): 103–34. doi:10.1016/S1054-3589(08)60542-6.

O'Neill, Chris. 2015. "No Title." *Animal Frontiers*, no. 5: 42–49.

Okeyo-Owuor, T., B. S. White, R. Chatrikhi, D. R. Mohan, S. Kim, M. Griffith, L. Ding, et al. 2015. "U2AF1 Mutations Alter Sequence Specificity of Pre-MRNA Binding and Splicing." *Leukemia*. doi:10.1038/leu.2014.303.

Opitz, C. A., M. Kulke, M. C. Leake, C. Neagoe, H. Hinssen, R. J. Hajjar, and W. A. Linke. 2003. "Damped Elastic Recoil of the Titin Spring in Myofibrils of Human Myocardium." *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.2133733100.

Owens, Nick D.L., Ira L. Blitz, Maura A. Lane, Ilya Patrushev, John D. Overton, Michael J. Gilchrist, Ken W.Y. Cho, and Mustafa K. Khokha. 2016. "Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development." *Cell Reports*. doi:10.1016/j.celrep.2015.12.050.

Ozawa, M., M. Hirabayashi, and Y. Kanai. 2002. "Developmental Competence and Oxidative State of Mouse Zygotes Heat-Stressed Maternally or in Vitro." *Reproduction*. doi:10.1530/rep.0.1240683.

Ozturk, Saffet, Berna Sozen, and Necdet Demir. 2014. "Telomere Length and Telomerase Activity during Oocyte Maturation and Early Embryo Development in Mammalian Species." *Molecular Human Reproduction*. doi:10.1093/molehr/gat055.

Padgett, Richard A. 2012. "New Connections between Splicing and Human Disease." *Trends in Genetics*. doi:10.1016/j.tig.2012.01.001.

Palmer, N., and P. Kaldis. 2016. "Regulation of the Embryonic Cell Cycle During Mammalian Preimplantation Development." In *Current Topics in Developmental Biology*. doi:10.1016/bs.ctdb.2016.05.001.

Pan, Qun, Ofer Shai, Leo J. Lee, Brendan J. Frey, and Benjamin J. Blencowe. 2008. "Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing." *Nature Genetics*. doi:10.1038/ng.259.

Papaioannou, V E. 1982. "Lineage Analysis of Inner Cell Mass and Trophectoderm Using Microsurgically Reconstituted Mouse Blastocysts." *Journal of Embryology and Experimental Morphology*.

Papasaikas, Panagiotis, J. Ramón Tejedor, Luisa Vigevani, and Juan Valcárcel. 2015. "Functional Splicing Network Reveals Extensive Regulatory Potential of the Core Spliceosomal Machinery." *Molecular Cell*. doi:10.1016/j.molcel.2014.10.030.

Parker, William B. 2009. "Enzymology of Purine and Pyrimidine Antimetabolites Used in the Treatment of Cancer." *Chemical Reviews* 109 (7): 2880–93. doi:10.1021/cr900028p.

Payer, Bernhard. 2016. "Developmental Regulation of X-Chromosome Inactivation." *Seminars in Cell and Developmental Biology*. doi:10.1016/j.semcdb.2016.04.014.

Peaston, Anne E., Alexei V. Evsikov, Joel H. Graber, Wilhelmine N. de Vries, Andrea E. Holbrook, Davor Solter, and Barbara B. Knowles. 2004. "Retrotransposons Regulate Host Genes in Mouse Oocytes

and Preimplantation Embryos." *Developmental Cell*. doi:10.1016/j.devcel.2004.09.004.

Peng, Aimin, Andrea L. Lewellyn, and James L. Maller. 2008. "DNA Damage Signaling in Early Xenopus Embryos." *Cell Cycle*. doi:10.4161/cc.7.1.5157.

Perumalsamy, Alagammal, Roxanne Fernandes, Ingrid Lai, Jacqui Detmar, Sue Varmuza, Robert F. Casper, and Andrea Jurisicova. 2010. "Developmental Consequences of Alternative Bcl-x Splicing during Preimplantation Embryo Development." *FEBS Journal*. doi:10.1111/j.1742-4658.2010.07554.x.

Petropoulos, Sophie, Daniel Edsgärd, Björn Reinius, Qiaolin Deng, Sarita Pauliina Panula, Simone Codeluppi, Alvaro Plaza Reyes, Sten Linnarsson, Rickard Sandberg, and Fredrik Lanner. 2016. "Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos." *Cell*, April. Elsevier. doi:10.1016/j.cell.2016.03.023.

Piccirillo, Ciriaco A., Eva Bjur, Ivan Topisirovic, Nahum Sonenberg, and Ola Larsson. 2014. "Translational Control of Immune Responses: From Transcripts to Translatomes." *Nature Immunology*. doi:10.1038/ni.2891.

Picelli, Simone, Omid R. Faridani, Åsa K. Björklund, Gösta Winberg, Sven Sagasser, and Rickard Sandberg. 2014. "Full-Length RNA-Seq from Single Cells Using Smart-Seq2." *Nature Protocols*. doi:10.1038/nprot.2014.006.

Pilzecker, Bas, Olimpia Alessandra Buoninfante, Paul van den Berk, Cesare Lancini, Ji-Ying Song, Elisabetta Citterio, and Heinz Jacobs. 2017. "DNA Damage Tolerance in Hematopoietic Stem and Progenitor Cells in Mice." *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1706508114.

Plachot, Michelle. 2000. "The Blastocyst." *Human Reproduction*. doi:10.1093/humrep/15.suppl_4.49.

Plant, Tony M., Anthony J. Zeleznik, and Peter O'Shaughnessy. 2015. "Knobil and Neill's Physiology of Reproduction." In *Knobil and Neill's Physiology of Reproduction*, 2684. doi:10.1016/B978-0-12-397175-3.00030-2.

Poruchynsky, Marianne S., Edina Komlodi-Pasztor, Shana Trostel, Julia Wilkerson, Marie Regairaz, Yves Pommier, Xu Zhang, et al. 2015. "Microtubule-Targeting Agents Augment the Toxicity of DNA-Damaging Agents by Disrupting Intracellular Trafficking of DNA Repair Proteins." *Proceedings of the National Academy of Sciences* 112 (5): 1571–76. doi:10.1073/pnas.1416418112.

Pratt, HPM, and CA Ziomek. 1982. "Compaction of the Mouse Embryo: An Analysis of Its Components." *Journal of Embryology …*.

Price, Brendan D., and Alan D. D'Andrea. 2013. "Chromatin Remodeling at DNA Double-Strand Breaks." *Cell*. doi:10.1016/j.cell.2013.02.011.

Qin, Han, Miroslav Hejna, Yanxia Liu, Michelle Percharde, Mark Wossidlo, Laure Blouin, Jens Durruthy-Durruthy, et al. 2016. "YAP Induces Human Naive Pluripotency." *Cell Reports*. doi:10.1016/j.celrep.2016.02.036.

Radford, Helois E., Hedda A. Meijer, and Cornelia H. de Moor. 2008. "Translational Control by Cytoplasmic Polyadenylation in Xenopus Oocytes." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. doi:10.1016/j.bbagrm.2008.02.002.

Ralston, Amy, and Janet Rossant. 2008. "Cdx2 Acts Downstream of Cell Polarization to Cell-Autonomously Promote Trophectoderm Fate in the Early Mouse Embryo." *Developmental Biology*. doi:10.1016/j.ydbio.2007.10.054.

Ram, Prahlad T., and Richard M. Schultz. 1993. “Reporter Gene Expression in G2 of the 1-Cell Mouse Embryo.” *Developmental Biology*. doi:10.1006/dbio.1993.1101.

Ratsima, Hery, Diego Serrano, Mirela Pascariu, and Damien D’Amours. 2016. “Centrosome-Dependent Bypass of the DNA Damage Checkpoint by the Polo Kinase Cdc5.” *Cell Reports* 14 (6): 1422–34. doi:10.1016/j.celrep.2016.01.014.

Ray, Debashish, Hilal Kazan, Kate B. Cook, Matthew T. Weirauch, Hamed S. Najafabadi, Xiao Li, Serge Gueroussov, et al. 2013. “A Compendium of RNA-Binding Motifs for Decoding Gene Regulation.” *Nature* 499 (7457): 172–77. doi:10.1038/nature12311.

Reima, I., E. Lehtonen, I. Virtanen, and J. E. Fléchon. 1993. “The Cytoskeleton and Associated Proteins during Cleavage, Compaction and Blastocyst Differentiation in the Pig.” *Differentiation*. doi:10.1111/j.1432-0436.1993.tb01586.x.

Resnick, Michael A. 1976. “The Repair of Double-Strand Breaks in DNA: A Model Involving Recombination.” *Journal of Theoretical Biology*. doi:10.1016/S0022-5193(76)80025-2.

Revil, Timothée, Daniel Gaffney, Christel Dias, Jacek Majewski, and Loydie A. Jerome-Majewska. 2010. “Alternative Splicing Is Frequent during Early Embryonic Development in Mouse.” *BMC Genomics*. doi:10.1186/1471-2164-11-399.

Richards, Adair D., and Alison Rodger. 2007. “Synthetic Metallomolecules as Agents for the Control of DNA Structure.” *Chemical Society Reviews*. doi:10.1039/b609495c.

Ringvoll, Jeanette, Line M. Nordstrand, Cathrine B. Vågbø, Vivi Talstad, Karen Reite, Per Arne Aas, Knut H. Lauritzen, et al. 2006. “Repair Deficient Mice Reveal MABH2 as the Primary Oxidative Demethylase for Repairing 1meA and 3meC Lesions in DNA.” *EMBO Journal*. doi:10.1038/sj.emboj.7601109.

Roberts, Adam, Cole Trapnell, Julie Donaghey, John L. Rinn, and Lior Pachter. 2011. “Improving RNA-Seq Expression Estimates by Correcting for Fragment Bias.” *Genome Biology*. doi:10.1186/gb-2011-12-3-r22.

Roberts, G. 1998. “Co-Transcriptional Commitment to Alternative Splice Site Selection.” *Nucleic Acids Research*. doi:10.1093/nar/26.24.5568.

Rooney, Sean, Frederick W. Alt, David Lombard, Scott Whitlow, Mark Eckersdorff, James Fleming, Sebastian Fugmann, David O. Ferguson, David G. Schatz, and JoAnn Sekiguchi. 2003. “Defective DNA Repair and Increased Genomic Instability in Artemis-Deficient Murine Cells.” *The Journal of Experimental Medicine*. doi:10.1084/jem.20021891.

Ross, Pablo J., Neli P. Ragina, Ramon M. Rodriguez, Amy E. Iager, Kannika Siripattarapravat, Nestor Lopez-Corrales, and Jose B. Cibelli. 2008. “Polycomb Gene Expression and Histone H3 Lysine 27 Trimethylation Changes during Bovine Preimplantation Development.” *Reproduction*. doi:10.1530/REP-08-0045.

Rossant, J., and P. P. L. Tam. 2009. “Blastocyst Lineage Formation, Early Embryonic Asymmetries and Axis Patterning in the Mouse.” *Development*. doi:10.1242/dev.017178.

Rothkamm, Kai, Ines Krüger, Larry H Thompson, Ines Kru, and Markus Lo. 2003. “Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle.” *Molecular and Cellular Biology*. doi:10.1128/MCB.23.16.5706.

SAKATANI, Miki. 2017. “Effects of Heat Stress on Bovine Preimplantation Embryos Produced *in Vitro*.”

Journal of Reproduction and Development. doi:10.1262/jrd.2017-045.

Sakatani, Miki, Shu Ichi Kobayashi, and Masashi Takahashi. 2004. "Effects of Heat Shock on in Vitro Development and Intracellular Oxidative State of Bovine Preimplantation Embryos." *Molecular Reproduction and Development*. doi:10.1002/mrd.20014.

Saltzman, Arneet L., Qun Pan, and Benjamin J. Blencowe. 2011. "Regulation of Alternative Splicing by the Core Spliceosomal Machinery." *Genes and Development*. doi:10.1101/gad.2004811.

Sánchez-Hernández, Noemí, Stéphanie Boireau, Ute Schmidt, Juan Pablo Muñoz-Cobo, Cristina Hernández-Munain, Edouard Bertrand, and Carlos Suñé. 2016. "The in Vivo Dynamics of TCERG1, a Factor That Couples Transcriptional Elongation with Splicing." *RNA*. doi:10.1261/rna.052795.115.

Sanchez, Roberto, and Ming Ming Zhou. 2011. "The PHD Finger: A Versatile Epigenome Reader." *Trends in Biochemical Sciences*. doi:10.1016/j.tibs.2011.03.005.

Sato, Miyuki, and Ken Sato. 2013. "Maternal Inheritance of Mitochondrial DNA by Diverse Mechanisms to Eliminate Paternal Mitochondrial DNA." *Biochimica et Biophysica Acta - Molecular Cell Research*. doi:10.1016/j.bbamcr.2013.03.010.

Saunders, Christopher M., Mark G. Larman, John Parrington, Llewellyn J. Cox, Jillian Royse, Lynda M. Blayney, Karl Swann, and F. Anthony Lai. 2002. "PLC{zeta}: A Sperm-Specific Trigger of Ca2+ Oscillations in Eggs and Embryo Development." *Development*.

Schaaper, R. M. 1993. "Base Selection, Proofreading, and Mismatch Repair during DNA Replication in Escherichia Coli." *Journal of Biological Chemistry*. doi:10.1074/jbc.274.3.1306.

Schultz, Richard M. 1993. "Regulation of Zygotic Gene Activation in the Mouse." *BioEssays*. doi:10.1002/bies.950150806.

———. 2002. "The Molecular Foundations of the Maternal to Zygotic Transition in the Preimplantation Embryo." *Human Reproduction Update*. doi:10.1093/humupd/8.4.323.

Schümperli, D., and R. S. Pillai. 2004. "The Special Sm Core Structure of the U7 SnRNP: Far-Reaching Significance of a Small Nuclear Ribonucleoprotein." *Cellular and Molecular Life Sciences*. doi:10.1007/s00018-004-4190-0.

Scott, Shaun P., and Tej K. Pandita. 2006. "The Cellular Control of DNA Double-Strand Breaks." *Journal of Cellular Biochemistry*. doi:10.1002/jcb.21067.

Shah, M Y, E Martinez-Garcia, J M Phillip, A B Chambliss, R Popovic, T Ezponda, E C Small, et al. 2016. "MMSET/WHSC1 Enhances DNA Damage Repair Leading to an Increase in Resistance to Chemotherapeutic Agents." *Oncogene* 35 (April). Macmillan Publishers Limited: 5905. https://doi.org/10.1038/onc.2016.116.

Sharova, Lioudmila V., Alexei A. Sharov, Timur Nedorezov, Yulan Piao, Nabeebi Shaik, and Minoru S H Ko. 2009. "Database for MRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells." *DNA Research*. doi:10.1093/dnares/dsn030.

Sharp, Andrew J., Elisavet Stathaki, Eugenia Migliavacca, Manisha Brahmachary, Stephen B. Montgomery, Yann Dupre, and Stylianos E. Antonarakis. 2011. "DNA Methylation Profiles of Human Active and Inactive X Chromosomes." *Genome Research*. doi:10.1101/gr.112680.110.

Shkreta, Lulzim, and Benoit Chabot. 2015. "The RNA Splicing Response to DNA Damage." *Biomolecules*. doi:10.3390/biom5042935.

Shum, Eleen Y., Samantha H. Jones, Ada Shao, Jennifer Dumdie, Matthew D. Krause, Wai Kin Chan, Chih Hong Lou, et al. 2016. "The Antagonistic Gene Paralogs Upf3a and Upf3b Govern Nonsense-Mediated RNA Decay." *Cell*. doi:10.1016/j.cell.2016.02.046.

Skirkanich, Jennifer, Guillaume Luxardi, Jing Yang, Laurent Kodjabachian, and Peter S. Klein. 2011. "An Essential Role for Transcription before the MBT in Xenopus Laevis." *Developmental Biology*. doi:10.1016/j.ydbio.2011.06.010.

Smit, AFA, R Hubley, and P Grenn. 2013. "RepeatMasker Open-4.0." *RepeatMasker Open-4.0.5*.

Smith, Zachary D., Michelle M. Chan, Kathryn C. Humm, Rahul Karnik, Shila Mekhoubad, Aviv Regev, Kevin Eggan, and Alexander Meissner. 2014. "DNA Methylation Dynamics of the Human Preimplantation Embryo." *Nature*. doi:10.1038/nature13581.

Smith, Zachary D., Michelle M. Chan, Tarjei S. Mikkelsen, Hongcang Gu, Andreas Gnirke, Aviv Regev, and Alexander Meissner. 2012. "A Unique Regulatory Phase of DNA Methylation in the Early Mammalian Embryo." *Nature*. doi:10.1038/nature10960.

Soumillon, Magali, Anamaria Necsulea, Manuela Weier, David Brawand, Xiaolan Zhang, Hongcang Gu, Pauline Barthès, et al. 2013. "Cellular Source and Mechanisms of High Transcriptome Complexity in the Mammalian Testis." *Cell Reports*. doi:10.1016/j.celrep.2013.05.031.

Squadrito, Mario Leonardo, Caroline Baer, Frédéric Burdet, Claudio Maderna, Gregor D. Gilfillan, Robert Lyle, Mark Ibberson, and Michele De Palma. 2014. "Endogenous RNAs Modulate MicroRNA Sorting to Exosomes and Transfer to Acceptor Cells." *Cell Reports*. doi:10.1016/j.celrep.2014.07.035.

Stanger, James D., Karen Stevenson, Aaron Lakmaker, and Robert Woolcott. 2001. "Pregnancy Following Fertilization of Zona-Free, Coronal Cell Intact Human Ova." *Human Reproduction*. doi:10.1093/humrep/16.1.164.

Stebbins-Boaz, B, L E Hake, and J D Richter. 1996. "CPEB Controls the Cytoplasmic Polyadenylation of Cyclin, Cdk2 and c-Mos MRNAs and Is Necessary for Oocyte Maturation in Xenopus." *The The European Molecular Biology Organization Journal*. doi:8665866.

Suh, Eun Kyung, Annie Yang, Arminja Kettenbach, Casimir Bamberger, Ala H. Michaelis, Zhou Zhu, Julia A. Elvin, Roderick T. Bronson, Christopher P. Crum, and Frank McKeon. 2006. "P63 Protects the Female Germ Line during Meiotic Arrest." *Nature*. doi:10.1038/nature05337.

Sun, Yehong, Tiyuan Li, Kewei Ma, Zhongkai Tian, Ying Zhu, Fuqiang Chen, and Gang Hu. 2009. "The Impacts of Ercc1 Gene Exon Viii Alternative Splicing on Cisplatin-Resistance in Ovarian Cancer Cells." *Cancer Investigation*. doi:10.3109/07357900902744536.

Sung, Hoon Back, Kyungho Lee, Elizabeth Vink, and Randal J. Kaufman. 2006. "Cytoplasmic IRE1α-Mediated XBP1 MRNA Splicing in the Absence of Nuclear Processing and Endoplasmic Reticulum Stress." *Journal of Biological Chemistry*. doi:10.1074/jbc.M602030200.

Sutovsky, P, R D Moreno, J Ramalho-Santos, T Dominko, C Simerly, and G Schatten. 2000. "Ubiquitinated Sperm Mitochondria, Selective Proteolysis, and the Regulation of Mitochondrial Inheritance in Mammalian Embryos." *Biology of Reproduction*. doi:10.1095/BIOLREPROD63.2.582.

Svoboda, Petr, and Matyas Flemr. 2010. "The Role of MiRNAs and Endogenous SiRNAs in Maternal-to-Zygotic Reprogramming and the Establishment of Pluripotency." *EMBO Reports*. doi:10.1038/embor.2010.102.

Swenberg, James A., Kun Lu, Benjamin C. Moeller, Lina Gao, Patricia B. Upton, Jun Nakamura, and Thomas B. Starr. 2011. "Endogenous versus Exogenous DNA Adducts: Their Role in Carcinogenesis, Epidemiology, and Risk Assessment." *Toxicological Sciences*. doi:10.1093/toxsci/kfq371.

Szabo, Stephen P, and Danton H O'Day. 1983. "THE FUSION OF SEXUAL NUCLEI." *Biological Reviews* 58 (3). Wiley/Blackwell (10.1111): 323–42. doi:10.1111/j.1469-185X.1983.tb00393.x.

Tadros, Wael, Aaron L. Goldman, Tomas Babak, Fiona Menzies, Leah Vardy, Terry Orr-Weaver, Timothy R. Hughes, J. Timothy Westwood, Craig A. Smibert, and Howard D. Lipshitz. 2007. "SMAUG Is a Major Regulator of Maternal MRNA Destabilization in Drosophila and Its Translation Is Activated by the PAN GU Kinase." *Developmental Cell*. doi:10.1016/j.devcel.2006.10.005.

Tanaka, K., T. Hiramoto, T. Fukuda, and K. Miyagawa. 2000. "A Novel Human Rad54 Homologue, Rad54B, Associates with Rad51." *Journal of Biological Chemistry*. doi:10.1074/jbc.M910306199.

Tang, Fuchou, Masahiro Kaneda, Dónal O'Carroll, Petra Hajkova, Sheila C. Barton, Y. Andrew Sun, Caroline Lee, Alexander Tarakhovsky, Kaiqin Lao, and M. Azim Surani. 2007. "Maternal MicroRNAs Are Essential for Mouse Zygotic Development." *Genes and Development*. doi:10.1101/gad.418707.

Tang, Wen, Ram Kannan, Marco Blanchette, and Peter Baumann. 2012. "Telomerase RNA Biogenesis Involves Sequential Binding by Sm and Lsm Complexes." *Nature*. doi:10.1038/nature10924.

Tanikawa, Michihiro. 2017. "Abstract 1430: The Spliceosome U2 SnRNP Factors Promote Genome Stability through Distinct Mechanisms; Transcription of Repair Factors and R-Loop Processing." *Cancer Research*. doi:10.1158/1538-7445.am2017-1430.

Tapial, Javier, Kevin C.H. Ha, Timothy Sterne-Weiler, André Gohr, Ulrich Braunschweig, Antonio Hermoso-Pulido, Mathieu Quesnel-Vallières, et al. 2017. "An Atlas of Alternative Splicing Profiles and Functional Associations Reveals New Regulatory Programs and Genes That Simultaneously Express Multiple Major Isoforms." *Genome Research* 27 (10): 1759–68. doi:10.1101/gr.220962.117.

Tarkowski, A. K., and J. Wróblewska. 1967. "Development of Blastomeres of Mouse Eggs Isolated at the 4- and 8-Cell Stage." *Journal of Embryology and Experimental Morphology*.

Tatsumi, Naoya, Nozomi Hojo, Hiroyuki Sakamoto, Rena Inaba, Nahoko Moriguchi, Keiko Matsuno, Mari Fukuda, et al. 2015. "Identification of a Novel C-Terminal Truncated WT1 Isoform with Antagonistic Effects against Major WT1 Isoforms." *PLoS ONE*. doi:10.1371/journal.pone.0130578.

Taylor, Deborah M., Pierre F. Ray, Asangla Ao, Robert M.L. Winston, and Alan H. Handyside. 1997. "Paternal Transcripts for Glucose-6-Phosphate Dehydrogenase and Adenosine Deaminase Are First Detectable in the Human Preimplantation Embryo at the Three- to Four-Cell Stage." *Molecular Reproduction and Development*. doi:10.1002/(SICI)1098-2795(199712)48:4<442::AID-MRD4>3.0.CO;2-Q.

Telford, Nancy A., Andrew J. Watson, and Gilbert A. Schultz. 1990. "Transition from Maternal to Embryonic Control in Early Mammalian Development: A Comparison of Several Species." *Molecular Reproduction and Development* 26 (1). United States: 90–100. doi:10.1002/mrd.1080260113.

Thermann, Rolf, Gabriele Neu-Yilik, Andrea Deters, Ute Frede, Kristina Wehr, Christian Hagemeier, Matthias W. Hentze, and Andreas E. Kulozik. 1998. "Binary Specification of Nonsense Codons by

Splicing and Cytoplasmic Translation." *EMBO Journal*. doi:10.1093/emboj/17.12.3484.

Tichy, Elisia D., and Peter J. Stambrook. 2008. "DNA Repair in Murine Embryonic Stem Cells and Differentiated Cells." *Experimental Cell Research*. doi:10.1016/j.yexcr.2008.02.007.

Tichy, Elisia D, Resmi Pillai, Li Deng, Li Liang, Jay Tischfield, Sandy J Schwemberger, George F Babcock, and Peter J Stambrook. 2010. "Mouse Embryonic Stem Cells, but Not Somatic Cells, Predominantly Use Homologous Recombination to Repair Double-Strand DNA Breaks." *Stem Cells and Development*. doi:10.1089/scd.2010.0058.

Tilgner, Hagen, Christoforos Nikolaou, Sonja Althammer, Michael Sammeth, Miguel Beato, Juan Valcárcel, and Roderic Guigó. 2009. "Nucleosome Positioning as a Determinant of Exon Recognition." *Nature Structural and Molecular Biology*. doi:10.1038/nsmb.1658.

Trapnell, Cole, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pachter. 2012. "Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks." *Nature Protocols* 7 (3): 562–78. doi:10.1038/nprot.2012.016.

Trapnell, Cole, Brian A Williams, Geo Pertea, Ali Mortazavi, Gordon Kwan, Marijke J van Baren, Steven L Salzberg, Barbara J Wold, and Lior Pachter. 2011. "Transcript Assembly and Abundance Estimation from RNA-Seq Reveals Thousands of New Transcripts and Switching among Isoforms." *Nature Biotechnology* 28 (5): 511–15. doi:10.1038/nbt.1621.Transcript.

Tress, Michael L., Federico Abascal, and Alfonso Valencia. 2017. "Alternative Splicing May Not Be the Key to Proteome Complexity." *Trends in Biochemical Sciences*. doi:10.1016/j.tibs.2016.08.008.

Trounson, A., C. Anderiesz, and G. Jones. 2001. "Maturation of Human Oocytes in Vitro and Their Developmental Competence." *Reproduction*. doi:10.1530/rep.0.1210051.

Tubbs, Anthony, and André Nussenzweig. 2017. "Endogenous DNA Damage as a Source of Genomic Instability in Cancer." *Cell*. doi:10.1016/j.cell.2017.01.002.

Van Den Abbeel, Etienne, Basak Balaban, Søren Ziebe, Kersti Lundin, Maria José Gómez Cuesta, Bjarke Mirner Klein, Lisbeth Helmgaard, and Joan Carles Arce. 2013. "Association between Blastocyst Morphology and Outcome of Single-Blastocyst Transfer." *Reproductive BioMedicine Online*. doi:10.1016/j.rbmo.2013.07.006.

van der Laan, Siem, Nikolay Tsanov, Carole Crozet, and Domenico Maiorano. 2013. "High Dub3 Expression in Mouse ESCs Couples the G1/S Checkpoint to Pluripotency." *Molecular Cell* 52 (3). United States: 366–79. doi:10.1016/j.molcel.2013.10.003.

Vandaele, L, M Thys, J Bijttebier, A Van Langendonckt, I Donnay, D Maes, E Meyer, and A Van Soom. 2010. "Short-Term Exposure to Hydrogen Peroxide during Oocyte Maturation Improves Bovine Embryo Development." *Reproduction (Cambridge, England)*. doi:10.1530/REP-09-0430.

Vassalli, J. D., J. Huarte, D. Belin, P. Gubler, A. Vassalli, M. L. O'Connell, L. A. Parton, R. J. Rickles, and S. Strickland. 1989. "Regulated Polyadenylation Controls MRNA Translation during Meiotic Maturation of Mouse Oocytes." *Genes and Development*. doi:10.1101/gad.3.12b.2163.

Vilenchik, M. M., and A. G. Knudson. 2003. "Endogenous DNA Double-Strand Breaks: Production, Fidelity of Repair, and Induction of Cancer." *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.2135498100.

Vitale, Ilio, Gwenola Manic, Ruggero De Maria, Guido Kroemer, and Lorenzo Galluzzi. 2017. "DNA Damage in Stem Cells." *Molecular Cell*. doi:10.1016/j.molcel.2017.04.006.

Vogel, Christine, and Edward M. Marcotte. 2012. “Insights into the Regulation of Protein Abundance from Proteomic and Transcriptomic Analyses.” *Nature Reviews Genetics* 13 (4): 227–32. doi:10.1038/nrg3185.

Wagner, Eric J., and Phillip B. Carpenter. 2012. “Understanding the Language of Lys36 Methylation at Histone H3.” *Nature Reviews Molecular Cell Biology*. doi:10.1038/nrm3274.

Walter, Marius, Aurélie Teissandier, Raquel Pérez-Palacios, and Déborah Bourc’His. 2016. “An Epigenetic Switch Ensures Transposon Repression upon Dynamic Loss of DNA Methylation in Embryonic Stem Cells.” *ELife*. doi:10.7554/eLife.11418.001.

Wang, Bin, Zhiling Li, Chao Wang, Man Chen, Jianfeng Xiao, Xiaoyan Wu, Wanfen Xiao, Yu Song, and Xiaoyan Wang. 2013. “Zygotic G2/M Cell Cycle Arrest Induced by ATM/Chk1 Activation and DNA Repair in Mouse Embryos Fertilized with Hydrogen Peroxide-Treated Epididymal Mouse Sperm.” *PLoS ONE*. doi:10.1371/journal.pone.0073987.

Wang, HaiYang, YiBo Luo, Ming-Hui Zhao, ZiLi Lin, Jeongwoo Kwon, Xiang-Shun Cui, and Nam-Hyung Kim. 2016. “DNA Double-Strand Breaks Disrupted the Spindle Assembly in Porcine Oocytes.” *Molecular Reproduction and Development*. doi:10.1002/mrd.22602.

Wang, K., R. Ramirez-Mitchell, and D. Palter. 1984. “Titin Is an Extraordinarily Long, Flexible, and Slender Myofibrillar Protein.” *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.81.12.3685.

Wang, Lu, Jun Zhang, Jialei Duan, Xinxing Gao, Wei Zhu, Xingyu Lu, Lu Yang, et al. 2014. “Programming and Inheritance of Parental DNA Methylomes in Mammals.” *Cell*. doi:10.1016/j.cell.2014.04.017.

Wang, Q C, J Liu, X Duan, X S Cui, N H Kim, B Xiong, and S C Sun. 2015. “The Dynamin 2 Inhibitor Dynasore Affects the Actin Filament Distribution during Mouse Early Embryo Development.” *The Journal of Reproduction and Development*. doi:10.1262/jrd.2014-079.

Wang, Qiao Chu, Jun Liu, Zhen Bo Wang, Yu Zhang, Xing Duan, Xiang Shun Cui, Nam Hyung Kim, and Shao Chen Sun. 2014. “Dynamin 2 Regulates Actin-Mediated Spindle Migration in Mouse Oocytes.” *Biology of the Cell*. doi:10.1111/boc.201400007.

Wang, Xinyi, Denghui Liu, Dajian He, Shengbao Suo, Xian Xia, Xiechao He, Jing Dong J. Han, and Ping Zheng. 2017. “Transcriptome Analyses of Rhesus Monkey Preimplantation Embryos Reveal a Reduced Capacity for DNA Double-Strand Break Repair in Primate Oocytes and Early Embryos.” *Genome Research* 27 (4): 567–79. doi:10.1101/gr.198044.115.

Wang, Yangming, Rostislav Medvid, Collin Melton, Rudolf Jaenisch, and Robert Blelloch. 2007. “DGCR8 Is Essential for MicroRNA Biogenesis and Silencing of Embryonic Stem Cell Self-Renewal.” *Nature Genetics*. doi:10.1038/ng1969.

Wang, Zhong Wei, Xue Shan Ma, Jun Yu Ma, Yi Bo Luo, Fei Lin, Zhen Bo Wang, Heng Yu Fan, Heide Schatten, and Qing Yuan Sun. 2013. “Laser Microbeam-Induced DNA Damage Inhibits Cell Division in Fertilized Eggs and Early Embryos.” *Cell Cycle*. doi:10.4161/cc.26327.

Ward, W. Steven. 2009. “Function of Sperm Chromatin Structural Elements in Fertilization and Development.” *Molecular Human Reproduction*. doi:10.1093/molehr/gap080.

Ward, W S, Donald S. Coffey, and Steven W. Ward. 1991. “DNA Packaging and Organization in Mammalian Spermatozoa: Comparison with Somatic Cells.” *Biology of Reproduction*. doi:10.1095/biolreprod44.4.569.

Watt, Fiona M, Ryan R Driskell, T. Aasen, D. J. Anderson, F. H. Gage, I. L. Weissman, P. Andrews, et al. 2010. "The Therapeutic Potential of Stem Cells." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. doi:10.1098/rstb.2009.0149.

Wiley, L M. 1988. "Trophectoderm: The First Epithelium to Develop in the Mammalian Embryo." *Scanning Microscopy* 2 (1). United States: 417–26.

Will, Cindy L., and Reinhard Lührmann. 2006. "Spliceosome Structure and Function." *Cold Spring Harbor Monograph Archive*. doi:10.1101/cshperspect.a003707.

Wilmut, Ian, Zhiyi Jing, Yu Zhang, Shaorong Gao, Mengqiu Dong, Xinzheng Guo, Shufang Wang, and Zhaohui Kou. 2010. "Proteome of Mouse Oocytes at Different Developmental Stages." *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1013185107.

Wong, Connie C., Kevin E. Loewke, Nancy L. Bossert, Barry Behr, Christopher J. De Jonge, Thomas M. Baer, and Renee A.Reijo Pera. 2010. "Non-Invasive Imaging of Human Embryos before Embryonic Genome Activation Predicts Development to the Blastocyst Stage." *Nature Biotechnology*. doi:10.1038/nbt.1686.

Wu, Xuemei, Maria M. Viveiros, John J. Eppig, Yuchen Bai, Susan L. Fitzpatrick, and Martin M. Matzuk. 2003. "Zygote Arrest 1 (Zar1) Is a Novel Maternal-Effect Gene Critical for the Oocyte-to-Embryo Transition." *Nature Genetics*. doi:10.1038/ng1079.

Xiao, Min, Jin Li, Wei Li, Yu Wang, Feizhen Wu, Yanping Xi, Lan Zhang, et al. 2017. "MicroRNAs Activate Gene Transcription Epigenetically as an Enhancer Trigger." *RNA Biology*. doi:10.1080/15476286.2015.1112487.

Xie, Dan, Chieh Chun Chen, Leon M. Ptaszek, Shu Xiao, Xiaoyi Cao, Fang Fang, Huck H. Ng, Harris A. Lewin, Chad Cowan, and Sheng Zhong. 2010. "Rewirable Gene Regulatory Networks in the Preimplantation Embryonic Development of Three Mammalian Species." *Genome Research*. doi:10.1101/gr.100594.109.

Xu, Xiuling, Shunlei Duan, Fei Yi, Alejandro Ocampo, Guang Hui Liu, and Juan Carlos Izpisua Belmonte. 2013. "Mitochondrial Regulation in Pluripotent Stem Cells." *Cell Metabolism*. doi:10.1016/j.cmet.2013.06.005.

Yamamori, Tohru, Hironobu Yasui, Masayuki Yamazumi, Yusuke Wada, Yoshinari Nakamura, Hideo Nakamura, and Osamu Inanami. 2012. "Ionizing Radiation Induces Mitochondrial Reactive Oxygen Species Production Accompanied by Upregulation of Mitochondrial Electron Transport Chain Function and Mitochondrial Content under Control of the Cell Cycle Checkpoint." *Free Radical Biology and Medicine* 53 (2): 260–70. doi:10.1016/j.freeradbiomed.2012.04.033.

Yan, Liying, Mingyu Yang, Hongshan Guo, Lu Yang, Jun Wu, Rong Li, Ping Liu, et al. 2013. "Single-Cell RNA-Seq Profiling of Human Preimplantation Embryos and Embryonic Stem Cells." *Nature Structural & Molecular Biology* 20 (9): 1131–39. doi:10.1038/nsmb.2660.

Yang, Qiyuan, Jimin Lin, Miao Liu, Ronghong Li, Bin Tian, Xue Zhang, Beiying Xu, et al. 2016. "Highly Sensitive Sequencing Reveals Dynamic Modifications and Activities of Small RNAs in Mouse Oocytes and Early Embryos." *Science Advances*. doi:10.1126/sciadv.1501482.

Yao, Yuan Q, David H Barlow, and Ian L Sargent. 2005. "Differential Expression of Alternatively Spliced Transcripts of HLA-G in Human Preimplantation Embryos and Inner Cell Masses." *Journal of Immunology* 175: 8379–85. doi:10.4049/jimmunol.175.12.8379.

Yonemasu, Rie, Mitsuyoshi Minami, Yoshimichi Nakatsu, Masayo Takeuchi, Isao Kuraoka, Yoichi

Matsuda, Yujiro Higashi, Hisato Kondoh, and Kiyoji Tanaka. 2005. “Disruption of Mouse XAB2 Gene Involved in Pre-MRNA Splicing, Transcription and Transcription-Coupled DNA Repair Results in Preimplantation Lethality.” *DNA Repair*. doi:10.1016/j.dnarep.2004.12.004.

Yoshimi, Akihide, and Omar Abdel-Wahab. 2017. “Molecular Pathways: Understanding and Targeting Mutant Spliceosomal Proteins.” *Clinical Cancer Research*. doi:10.1158/1078-0432.CCR-16-0131.

Yuan, Shuiqiao, Andrew Schuster, Chong Tang, Tian Yu, Nicole Ortogero, Jianqiang Bao, Huili Zheng, and Wei Yan. 2016. “Sperm-Borne MiRNAs and Endo-SiRNAs Are Important for Fertilization and Preimplantation Embryonic Development.” *Development*. doi:10.1242/dev.131755.

Yukawa, Masashi, Shoji Oda, Hiroshi Mitani, Masao Nagata, and Fugaku Aoki. 2007. “Deficiency in the Response to DNA Double-Strand Breaks in Mouse Early Preimplantation Embryos.” *Biochemical and Biophysical Research Communications*. doi:10.1016/j.bbrc.2007.04.162.

Zemach, Assaf, Ivy E. McDaniel, Pedro Silva, and Daniel Zilberman. 2010. “Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation.” *Science*. doi:10.1126/science.1186366.

Zhang, Jin Yu, Yun Fei Diao, Hong Rye Kim, and Dong Il Jin. 2012. “Inhibition of Endoplasmic Reticulum Stress Improves Mouse Embryo Development.” *PLoS ONE*. doi:10.1371/journal.pone.0040433.

Zhang, Maomao, Priyanka Kothari, Mary Mullins, and Michael A. Lampson. 2014. “Regulation of Zygotic Genome Activation and DNA Damage Checkpoint Acquisition at the Mid-Blastula Transition.” *Cell Cycle*. doi:10.4161/15384101.2014.967066.

Zhao, Bo, Wei Dao Zhang, Ying Liang Duan, Yong Qing Lu, Yi Xian Cun, Chao Hui Li, Kun Guo, et al. 2015. “Filia Is an ESC-Specific Regulator of DNA Damage Response and Safeguards Genomic Stability.” *Cell Stem Cell*. doi:10.1016/j.stem.2015.03.017.

Zhao, Shanrong, Ying Zhang, Ramya Gamini, Baohong Zhang, and David Von Schack. 2018. “Evaluation of Two Main RNA-Seq Approaches for Gene Quantification in Clinical RNA Sequencing: PolyA+ Selection versus RRNA Depletion.” *Scientific Reports*. doi:10.1038/s41598-018-23226-4.

Zhou, Bin Bing S., and Jiri Bartek. 2004. “Targeting the Checkpoint Kinases: Chemosensitization versus Chemoprotection.” *Nature Reviews Cancer*. doi:10.1038/nrc1296.

Zhou, Wei, Amanda L. Anderson, Adrian P. Turner, Geoffry N. De Iuliis, Adam McCluskey, Eileen A. McLaughlin, and Brett Nixon. 2017. “Characterization of a Novel Role for the Dynamin Mechanoenzymes in the Regulation of Human Sperm Acrosomal Exocytosis.” *Molecular Human Reproduction*. doi:10.1093/molehr/gax044.

CONTRIBUTIONS

Manuel Irimia envisioned the overall project, and performed some of the foundational bioinformatics work, that were the basis of my own work (including creating a vast-tools library for cow to measure splicing, and providing datasets [protein frame predictions etc.]). Barbara Pernaute (CRG) designed and conducted all the wet–lab experiments, with the assistance of Marta Miret Cuesta (CRG) and Lucia Goyeneche (Centro de esterilidad, Montevideo [former CRG]). Quirze Rovira Castellà (Max Planck Institute for Molecular Biomedicine, Munster [former CRG]) did a masters project in the lab, studying the relationship between transposable elements and alternative splicing during preimplantation development, providing the motivation to explore this interesting connection in this thesis. Sophie Bonnal (Valcarcel lab, CRG) and Manuel Irimia downloaded the RNA-Seq for RBP knockdown/overexpression comparions. Yamile Marquez provided the exon conservation keys and scientific advice.

ACKNOWLEDGEMENTS

This is a rare opportunity to thank all the people who have supported my career during the past four and a half years, and indeed before. I have been blessed with many amazing colleagues, understanding family and supporting friends, without which, it would have been impossible to get through this marathon, that is a PhD.

First of all, I need to give huge thanks my supervisor.
Manuel Irimia (Manu), who has been an unbelievable source of inspiration and ideas during the whole of my time at CRG. The enthusiasm and the ability to work on so many independent projects at the same time, is truly unbelievable. I thank you for all that you have taught me, not just the science, but being a scientist. I owe a great debt for your support all these years, and I hope I can repay this in the near future.
In addition, to **Barbara Pernaute,** who has been enormously generous in her time and support of my PhD project. Completing arduous rounds of experimental work (all the wet lab sections of this thesis). I thank her for her expert knowledge of embryology, and her dedication, without which it would have been impossible to complete this PhD. I cannot thank her enough for all her help.

In regards to this thesis directly, I have to give special mention to Marta Miret Cuesta, who also performed a multitude of mouse embryo experiments alongside Barbara, and to Quirze who did an excellent analysis of transposable elements and their relationship to splicing, which formed the basis for Chapter 3.2. To Sophie, who created the excellent RBP knockdown database I used to discover AS correlation with SNRPB. To Yamile, who has also contributed massively to my thesis project, she has always been there to help (on a myriad of issues), and I have always taken full advantage of her vast experience. In addition, I want to thank Juan Valcarcel, Ben Lehner and Tomas Marques, who offered their precious time to help evaluate my work, as part of my thesis advisory board. Finally, thanks to Imma, for her dedication to the student body, and her assistance throughout the PhD.

I also want to thank my thesis defence committee board for their time and willingness to evaluate my work. Taking away some of their precious time from their busy labs. To Bernhard Payer, Reini Luco and Mihaela Zavolan.

Next, I want to thank all the members of the lab, who have made my time here so special. Many have already left to far-flung places, but I hope we can always stay in contact. Andre, Antonio, Barbara, Beth, Cris, Demi, Fede, Jon, Laura, Lucia, Marta, Patryk, Quirze, Reza, Simon, Thom, Vicky and Yamile.

I also want to thank my former (Bristol university) and future supervisor (UCL), **Seirian Sumner**, who rescued me from a wasp-infested cider factory, to become part of her less wasp-infested laboratory. Here, I was able to get great experience, working on the transcriptomics and proteomics of the Polistes paper wasps. In addition, **Gary Barker** (Bristol University) gave me massive amounts of help and inspired me to become a bioinformatician.

To my family, you all have been an amazing support. To my parents, who have always strove to give me as many opportunities as possible in life. Supporting everything I do, understanding everything that I am, and helping me to appreciate and be thankful. To my brothers and sisters-in-law, I hope I can spend more time with you all now, especially if you need day-care (that's a promise in writing).

To Adria, thank you for always being there for me. Living in different countries is hard, yet you always are able to brighten the day from whatever distance. I am not sure I would have been able to get through it all without you. We have had such amazing times over the years, since we were acquainted at the Systems biology retreat in Cardona. With especial thanks to Carolina, who linked us up.

To Barcelona, what a perfect city to live in, I will find it difficult to ever leave. To Hospital clinic, to whom I potentially owe my life, thanks for the emergency surgery.

To all the people who I had the pleasure to get to know during my PhD, especially our PhD year. I watch our YouTube video with such fond memories so many years ago (https://www.youtube.com/watch?v=D00fpATMhbY). To Alessandro, for showing

me the best sides of Barcelona. To Bogumila, the best flat-mate/friend ever, thanks for everything, its been an amazing few years together, will be very sad to see you leave. To Birgit, the life and soul of any party, I miss you so much, too many amazing times to write here. I wish you all the best in your amazing new position in Vienna. To Cindy, you are the queen of style. To Claudia, to all the amazing times we have spent together, and for the endless help submitting this thesis. To Evan, who would have thought we would have got to the end after 4 years, so many fond memories in Eixample and beyond. Thanks for all the amazing times exploring Barcelona and your friendship. To Artem, Andrea, Anna, Dan, Elena, Gregoire, Ivica, Krystal, Micky, Moritz, Paul, Paula thank you all.

Lastly, to all those that I have forgotten in the race to finish the PhD.