MECHANISMS OF GERM CELL FORMATION DURING ZEBRAFISH EMBRYOGENESIS

by

FABIO MARCO D'ORAZIO

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences

College of Medical and Dental Sciences

The University of Birmingham

May 2019

Abstract

In metazoans, the germ fate is acquired during embryogenesis either via oocyte-inherited cytoplasmic aggregates or via chemical induction from the surrounding embryonic cells. Most of the model organisms, including *Caenorhabditis elegans, Drosophila melanogaster, Xenopus laevis and Danio rerio,* rely on maternal determinants necessary to generate the germ line of the offspring. Although it has been largely established that germ determinants are required for the formation of germ cells, the specific molecular mechanisms driving the onset of the germ line are still unclear. Germ granules have been implicated in transcriptional inhibition contributing to skipping somatic differentiation. Also, epigenetic reprogramming of the embryonic germ line has been shown in several model organisms. However, little is known about the role of the germ plasm in transcription and epigenetics.

Here, we show that the germ plasm and the epigenetic landscape of zebrafish primordial germ cells (PGCs) are tightly linked. The early germ line shows similar transcriptional timing, transcriptomic and chromatin profiles with the rest of the embryo and the germ fate is gradually acquired during the first day of development. A PGC-like chromatin profile is acquired while germ plasm re-localises within the cells and PGCs and somatic cells undergo significant epigenetic and transcriptional divergence.

By performing time series of chromatin and transcript profiles in the PGCs, we could identify candidate PGC-specific cis-regulatory elements and transcripts. We detect both hypermethylation and chromatin compaction around putative developmental enhancers indicating that the germ fate is acquired avoiding lineage differentiation.

Finally, to link epigenetic dynamics to germ plasm behaviour, we inhibited the translation of Tudor Domain 7 (Tdrd7), a germ-plasm-localised protein involved in structural organisation of the germ granules. The mutant embryos reprogram the PGC-specific chromatin state and resemble the somatic cells, suggesting that the germ plasm is primarily responsible for epigenetically preserving the pluripotent state of the PGCs.

Acknowledgements

Within the several people who I would like to acknowledge, my deepest thank goes to Professor Ferenc Mueller, my Phd supervisor, scientific mentor and guide. Thanks Ferenc for introducing me to the world of science, for the constant motivation and for trusting me throughout this project.

A big thank goes to Professor Boris Lenhard and Doctor Piotr Balwierz at the Imperial College London, who offered their help and greatly contribute to the completion of the project.

I must acknowledge the ZENCODE consortium for the excellent training and the many opportunities provided, the principal investigators involved and a special mention to the students. It was a pleasure to share this experience with all of you. I also acknowledge the Marie Curie Actions for funding my PhD.

I would like to thank Matthew MacKenzie, for its endless and constant support in PGC isolation. Thanks also to all the people in the sequencing and animal facilities at the University of Birmingham and to Professor Roland Dosch for donating transgenic zebrafish embryos.

Also, many thanks to my colleagues, who daily supported and advised me. Particularly, I thank Doctor Nan Li and Yavor Hadzhiev for sharing their invaluable knowledge and expertise.

A warm thank to my family: to my parents for their encouragement, motivation and love and to my brother and sister for transmitting their energy and happiness.

Finally, my gratitude goes to my amazing wife Veronica, who shared this experience with me from the beginning to the end and always believed in me. You are my constant reference point.

Table of Contents

Table of Figures

List of Tables

List of abbreviations

µl	microlitre
°C	Degrees Celsius
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
DAPI	4',6-diamidino-2-phenylindole
dNTP	Deoxyribonucleotide triphosphate
dpf	days post fertilisation
ESC	Embryonic stem cell
FACS	Fluorescence Activated Cell Sorting
FC	Fold change
hpf	hours post fertilisation
miRNAs	MicroRNAs
ml	Millilitre
mM	Millimolar
mRNA	Messenger ribonucleic acid
ng	Nanogram
nl	Nanolitre
PBS	Phosphate buffered saline
PCA	Principle component analysis
PCR	Polymerase chain reaction
PGC	Primordial germ cell
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RT	Room temperature

RT-PCR Reverse transcriptase-polymerase chain reaction

TSS Transcriptional start site

UTR Untranslated region

g Times gravity

1 Introduction

1.1 General introduction

All sexually-reproducing organisms are made of two fundamental cell types: somatic and germ cells. While the somatic cells differentiate and form tissues and organs needed for keeping the individual alive and functional, the germ cells are superfluous to the organism's survival. Instead, they supply the role of forming the offspring and propagating the species. Germ cells have the unique ability to perform meiosis, to retain totipotency although undergoing terminal differentiation and to fuse aploid nuclei in order to generate a zygote.

The lineage of the germ cells is the germ line and its developmental fate is usually established at the beginning of embryogenesis when the primordial germ cells (PGCs) arise. The correct development of the germ line requires several levels of regulation and defects of this embryonic process lead to adult sterility (Kobayashi et al., 2017; Strome and Lehmann, 2007; Strome and Updike, 2015). The acquisition of embryonic germ fate is established via highly-specialised post-transcriptional control under regulation of both cytoplasmic aggregates and epigenetic reprogramming (Seydoux and Braun, 2006). Although recent technological advances have explored the roles of germ cell-specific proteins and epigenetic landscapes, several aspects of PGC formation remain unknown. For example, the function of many germ line-specific factors, the link between cytoplasmic and nuclear germ cell regulation and the molecular mechanisms guiding epigenetic reprogramming are currently unclear.

We suggest *Danio rerio* (zebrafish) as a model for studying the relation between nuclear and cytoplasmic events required for the transition from embryonic stem cell (ESC) to germ cell. In fact, while many zebrafish germ factors are conserved among the animal kingdom, zebrafish PGCs mature quickly and are easily accessible due to the extrauterine development of the

embryo. In principle, these features boost PGCs retrieval and allow in-depth studies of the molecular steps triggering germ cell formation in a developing vertebrate.

1.2 Two main mechanisms allow germ cell formation in the metazoans

The germ line is one the first cell types to be specified during multicellular organism development and it can develop in two ways among the animal kingdom (Strome and Lehmann, 2007). Mammals, urodeles and some insects develop their PGCs after extracellular signals are released by neighbouring somatic cells, usually in a peripheral side of the embryo. This mechanism is known as epigenesis or induction of the germ line. In other organisms, such as *D. melanogaster*, *C. elegans*, *X. laevis* and *D. rerio*, the specification of the germ line requires maternal factors that form the germ plasm. The formation of the germ line through maternal germ plasm is known as preformation (Lehmann, 2016). The germ plasm is accumulated in the oocyte and segregated into the future embryonic germ cells after fertilisation. Inheritance of maternal determinants (germ plasm) allows the PGCs to be specified very early during development, even before the onset of zygotic genome activation (ZGA) (Eddy, 1975; Williamson and Lehmann, 1996). Interestingly, the transmission of the germ plasm throughout the subsequent generations makes the germ line continuous in every stage of an individual's lifetime (Figure 1.1).

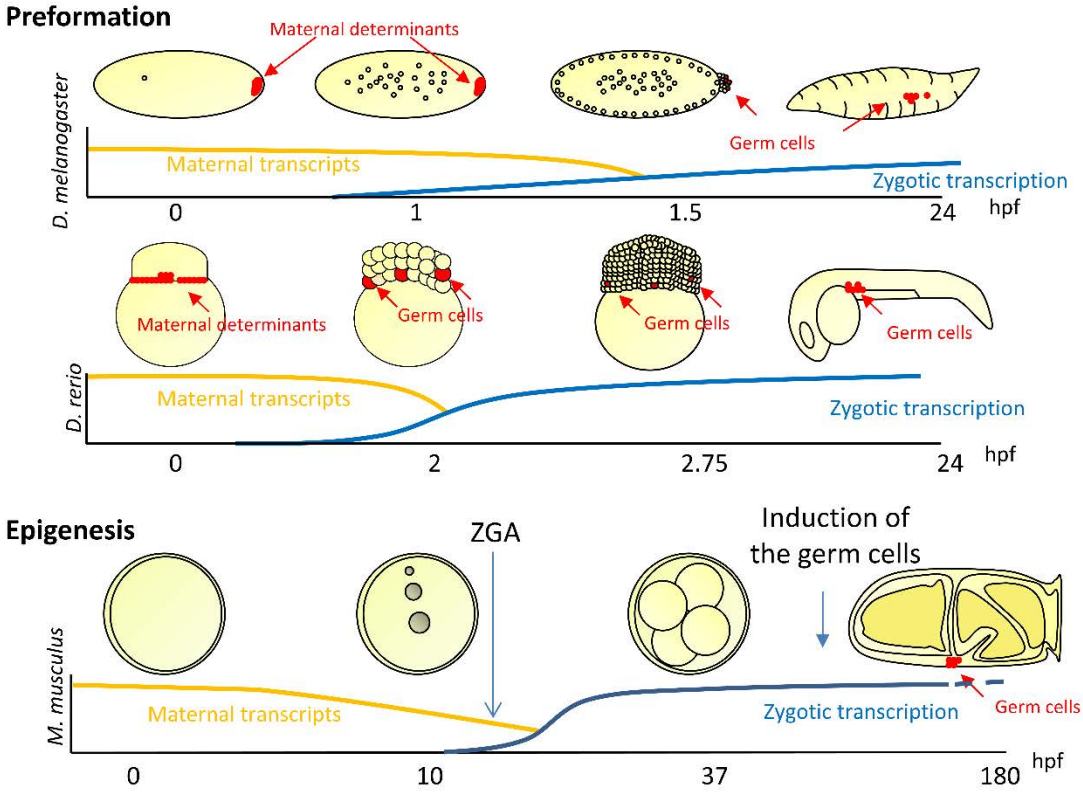

Figure 1. 1: Two modes of germ cell specification have been described among different organisms. Germ plasm-dependent preformation (top) and epigenesis (bottom) in three representative model organisms. Embryo schematics represent early developmental phases in fly, fish and mouse. Germ plasm and germ cells are in red and indicated by red arrows. Line chart depicts maternal (orange) and zygotic (blue) transcript levels over hours post fertilisation (hpf). Onset of ZGA in mouse is indicated by the blue arrow.

Preformation and epigenesis are utilised differently among the metazoans. Although most of the laboratory model organisms rely on preformation, within the bilateria epigenesis is more common (Extavour, 2007). Interestingly, the phylogenetic tree of PGC specification in the animal kingdom suggests that epigenesis is evolutionarily more ancestral (Figure 1.2). Moreover, organisms in which preformation occurs belong to evolutionary-distinct branches sharing ancestors in where the germ line was specified by induction, indicating that germ plasm-dependent PGC specification evolved independently among clades of animal. The fact that many vertebrates and invertebrates convergently evolved preformative acquisition of the germ line also suggests a selective advantage. In fact, it was proposed that germ plasm-driven germ cell formation eases the disengagement between the somatic and the germ line, therefore

favouring the early embryonic differentiation of somatic pathways (Johnson et al., 2011). In support of this suggestion, maternal acquisition of germ cell determinants seems to promote evolution. Consequently, teleosts, anurans, birds and ascidians underwent intense species branching in a relative short evolutionary window compared to organisms in which PGCs are derived by epigenesis (Evans et al., 2014; Johnson et al., 2011).

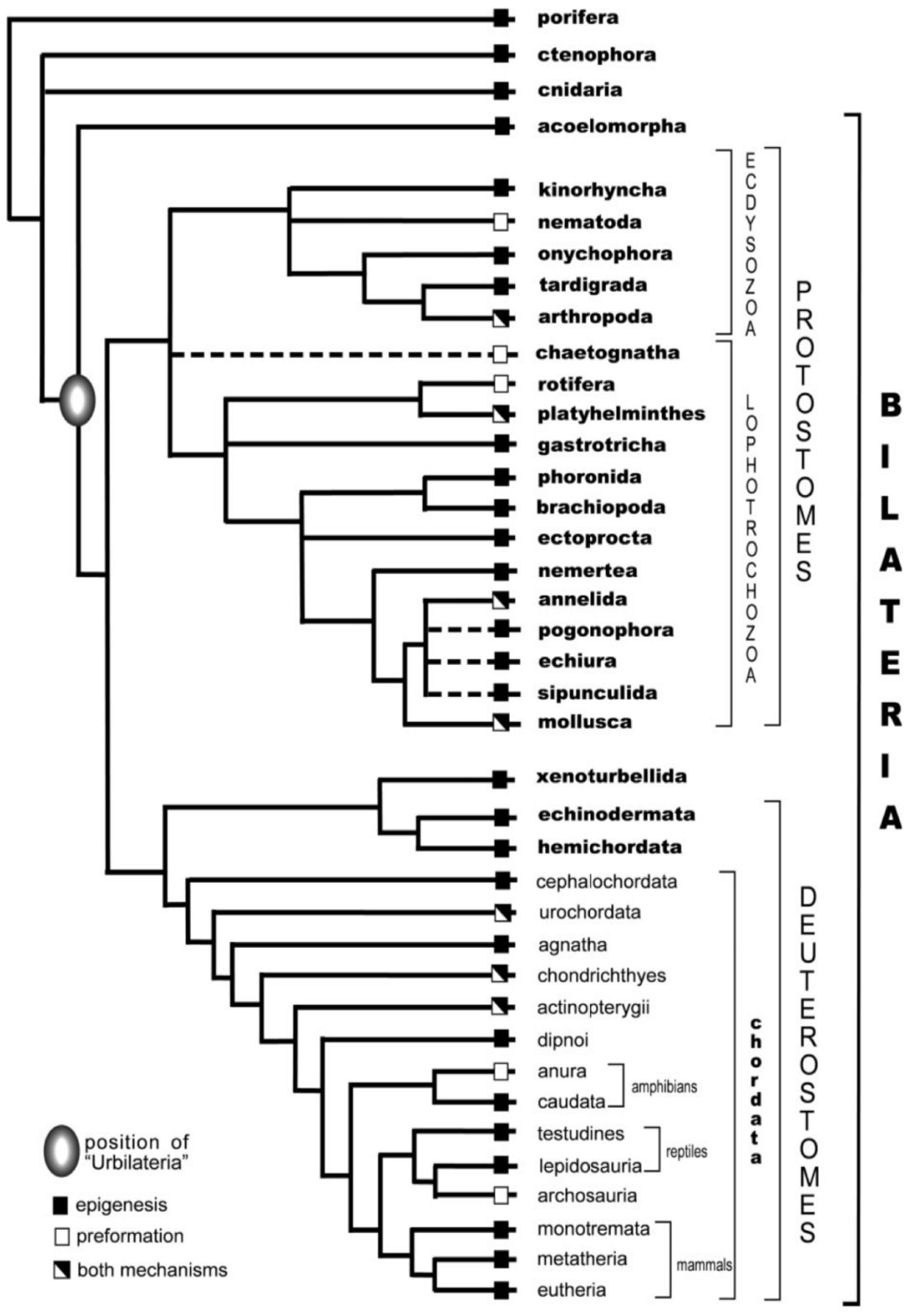

Figure 1. 2: Modes of PGC specification across Bilateria.
Phylogenetic dendogram of bilateria highlighting the modes of germ line specification. The position of the urbilatera is indicated by the white oval with shaded edges. Organisms forming the germ cells via epigenesis are labelled by a dark square, while organisms using preformation are shown as white squares. From (Extavour, 2007).

The two mechanisms of PGC formation employ different triggers: while embryonic expression of pluripotent-specific genes is required during epigenesis (Kurimoto et al., 2008), organisms preforming the germ line do not need embryonic transcription to initiate PGC formation, being instead driven by parental-inherited factors (Tam and Zhou, 1996). This is supported by several lines of evidence showing that early zygotic transcription is not required to induce the germ fate. When the nucleus of a germ plasm-carrying cell was manually removed, the germ plasm was still able to change cytoplasmic distribution from agglomerate to spread (Knaut et al., 2000). Moreover, it was shown that PGCs can begin migratory movements even in presence of the transcriptional inhibitor Alpha-amanitin, reinforcing the idea that zygotic transcription does not guide early phases of PGC development (Blaser et al., 2005). Notably, the first phases of germ cell development are characterised by a natural transcriptional inhibition or delay in ZGA (Hanyu-Nakamura et al., 2008; Knaut et al., 2000; Saitou et al., 2002; Yu et al., 2000). In many organisms, regardless whether PGCs are preformed or induced, an extended transcriptional silencing seems essential for protecting germ cells against the somatic transcriptional plan (Strome and Updike, 2015). The molecular and biological reasons for delaying genomic activation in PGCs are discussed later in this thesis.

1.2.1 Epigenesis of the germ line

As mentioned, most of the bilateria rely on germ plasm-independent PGC formation, a mechanism defined as epigenesis. This requires secretion of molecules from developing tissues that generate gradients aimed to trigger the activation of germ line-associated pathways (Figure 1.3). In mouse, the germ line specification is initiated by the expression of several transcription factors in few cells at the posterior streak (Ying and Zhao, 2001). In particular, at embryonic day 5.75 (E5.75), the bone morphogenetic protein 4 (Bmp4) is released by extra embryonic ectoderm in proximity to the site where PGCs arise. Bmp4 antagonises the activity of Wingless

and INT (Wnt) 3 and changes the binding affinity of the transcription factor Brachyury (Figure 1.4). In the somatic line, Brachyury activates transcription of mesodermal genes, however, it is hijacked towards germ line genes upon Bmp4/Brachyury cooperation (Lawson et al., 1999; Lolas et al., 2014). As well as Brachyury, three other transcription factors under the control of Wnt3/Bmp4 induce germ cell formation in the epiblast: Blimp1, PR domain zinc‑finger protein 14 (Prdm14) and activating enhancer-binding protein 2γ (Ap2γ) (Ohinata et al., 2005; Yamaji et al., 2008). Blimp1 and Ap2γ bind the DNA through a PR-domain and a Krüppel-type zinc fingers domain (Yu et al., 2000) and Blimp1 was shown to function epistatically to Ap2γ (Weber et al., 2010). Importantly, the PGCs of *Blimp1* knockout mice are unable to transcribe germ cell genes and express somatic genes, suggesting that *Blimp1* acts epistatically during germ fate acquisition (Kobayashi et al., 2017; Tang et al., 2016).

The PR-domain and the Kruppel domains are typically used by transcription factors (TFs) due to their binding specificity for selective DNA motifs. The Kruppel-like zinc finger domain is made of 28-30 amino acids stabilised by a zinc ion interacting with two cysteines and two histidines (Looman et al., 2002). This domain is repeated up to 30 times in a single Kruppel-associated box zinc finger (KRAB-ZF) protein, allowing high specificity for the DNA target. Similarly, the PR-domain characterises the PRDM family of proteins (PRDI-BF1 and RIZ homology domain) and is found on TFs in association with multiple zinc finger motifs (Huang et al., 1998).

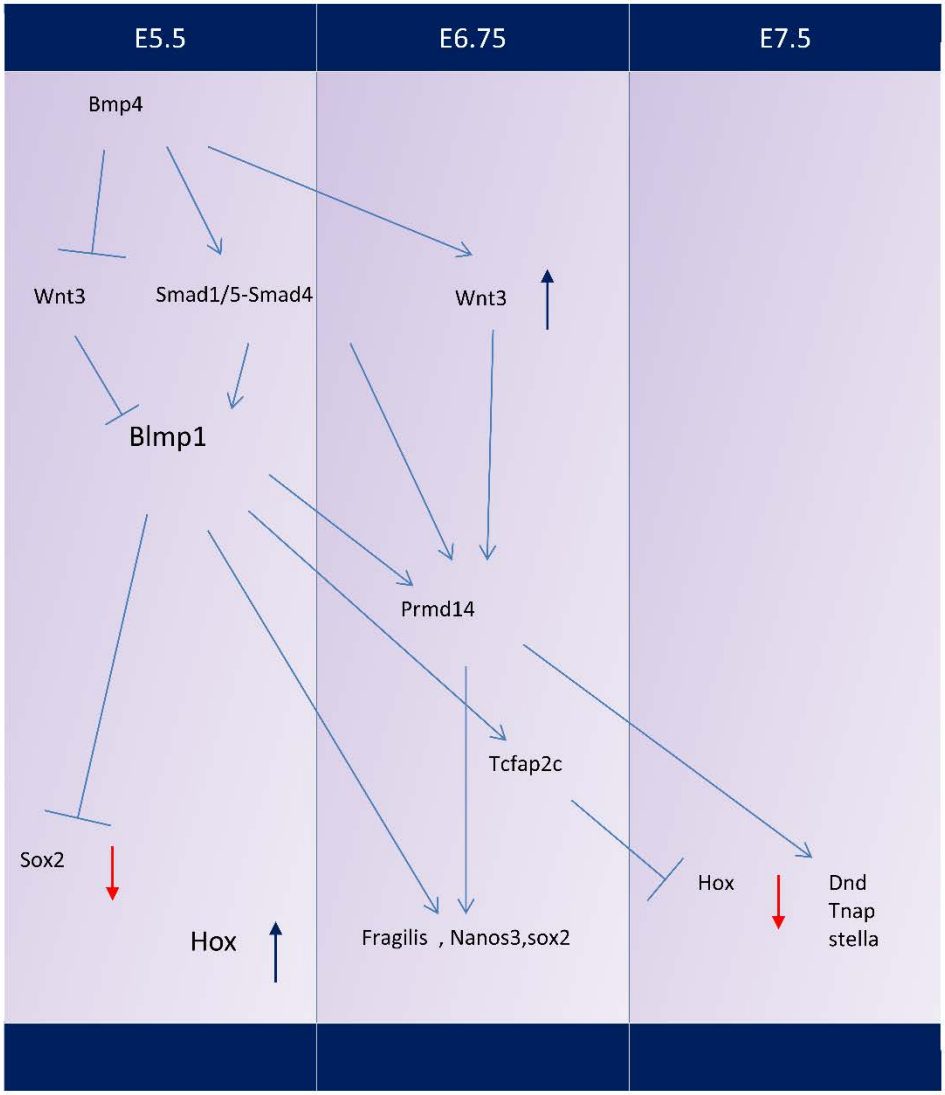

Figure 1. 3: Molecular activation of PGC pathways in mouse.
Pathways leading to PGC formation and migration in mouse from embryonic day 5.5 (E5.5) to embryonic day 7.5 (E7.5). Light blue arrows indicate gene activation. Red and dark blue arrows report down and upregulation respectively. At E5.5, BMP4 triggers activation of SMAD1-5 and SMAD4, while contrasting the activity of WNT3. This event liberates BLMP1, resulting in the upregulation of *hox* genes. At day E6.5, reactivation of WNT3 induces expression of PRMD14 and consequently of *fragilis*, *nanos3* and *sox2*. Later, *hox* genes are switched off and further germ line genes result upregulated (*dnd*, *tnap* and *stella*).

1.2.1.1 Activation of germ line-specific transcription in mouse

In mouse, PGC arise upon initiation of transcriptional programs triggered by Blimp1 (Ohinata et al., 2005). Between embryonic day 5.5 and 7.5 (E5.5 and E7.5), more than 800 genes are differentially regulated in PGCs compared to somatic cells. Two important markers for the

germ line, *fragilis* and *stella* are detected in the PGCs after E7.5 (Figure 1.3), although no effects on the germ cells are observed in the corresponding knock down (KD) animals (Lange et al., 2008; Payer et al., 2003). After specification, a significant cell and nuclear size increase of the murine PGCs was observed and linked to global hyper-transcriptional activity. Transcriptome analysis by qPCR and RNA-seq on E13.5 and E15.5 isolated germ cells and cell-number-normalization (CNN) indicated that proliferative PGCs (E13.5) show a 4-fold increase in total transcript levels compared to somatic cells while a 2-fold increase was observed in non-dividing PGCs (E15.5) (Percharde et al., 2017).

Gene expression analysis by RNA-seq of male and female embryonic germ cells between E11.5 and E18.5 has identified eleven germ cell-specific genes (Sabour et al., 2011). All of these genes showed significant enrichment in PGCs when compared to ESCs and higher levels in either both or male and female mature germ cells (Table 1.1).

Table 1.1: expression of PGC-specific genes in male and female mature germ/somatic cells.

Gene	Male germ cells	Female germ cells	Somatic cells
Fkbp6	High	Low	Low
Mov10l1	High	Low	Low
4930432K21Rik	High	Low	Low
Tex13	High	Low	Low
Akt3	High	High	Mild
Gm1673	High	High	Mild
Hba-a1	High	High	High
Pik3r3	High	High	Mild
Plcl2	High	High	Mild
Spo11	High	High	Mild
Tdrkh	High	High	Mild

1.2.1.2 PGC migration in mouse

The migration of the PGCs in mouse embryos starts when they passively leave the posterior streak to the hind-gut epithelium at E8.0. At this stage, the PGCs show mild motile features and start asynchronized movements towards the endoderm (Molyneaux et al., 2001; Richardson and Lehmann, 2010). At E9.0, a cluster of PGCs moves dorsally towards the genital ridge, however, it was shown that they can reach the final destination from different directions as happens in other animals. This occurs through chemotaxis, which drives the PGCs expressing CXCR4 towards SDF1 signals (Molyneaux et al., 2003). At this stage, PGC survival is strongly dependent on their position within the embryo and the surrounding stimuli and signals. Germ cells that fail to reach the genital ridge promptly undergo apoptosis due to lack of interaction between the germ cell receptor c-Kit and surrounding chemical stimuli (Godin et al., 1991; Molyneaux et al., 2001). PGC survival and migration are also dependent on FGF signalling. E10.5 PGCs express both *Fgfr2-IIIb* and *Fgfr1-IIIc,* which interact with FGF2 during migration. In fact, migration speed and relocation of PGCs treated with FGF2 and the FGF inhibitor SU5402 were increased and decreased respectively (Takeuchi et al., 2005). At the end of migration, the PGCs colonise the gonads where they are now collectively termed gonocytes. Here, male gonocytes experience mitotic arrest at E14.5, whereas female gonocytes undergo meiosis (Yamaguchi et al., 2005).

1.2.1.3 PGC development in human

Studies on the human embryonic germ line are limited by ethics and accessibility to human embryos, therefore most of the current knowledge is based on model organisms. In human, the PGCs form during gastrulation in the posterior epiblast two weeks after fertilisation. During the following two weeks, the PGCs develop separately from the somatic cells, acquiring distinct morphological and molecular features, until they migrate to the developing gonads, where they undergo sexual determination (Richardson and Lehmann, 2010).

Recent transcriptomic studies have revealed that human PGCs express several pluripotency factors, such as POU5F1 (OCT4), NANOG, ZFP42 (REX1), DPPA3 (STELLA), SALL4, and LIN28A, but not SOX2 as murine PGCs (Hayashi et al., 2011; Magnúsdóttir et al., 2013a; Ohinata et al., 2009). They also show several germ line marker genes, such as KIT, ALPL, TFAP2C, DND1, NANOS3, and TCL1A (Guo et al., 2015). SOX17 seems to play a central role for controlling the germ line fate during human embryogenesis (Irie et al., 2015). As well as mouse, human gonadal PGCs express the surface marker CD117 (Gaskell et al., 2004).

1.2.2 Preformation of the germ line

The term preformation of the germ line indicates that the embryonic fate of the germ cells is predetermined before fertilisation occurs. In this case, the germ line formation initiates during oocyte maturation, where proteins and transcripts are aggregated to form the germ plasm. The germ plasm is a phase-separated aggregate of proteins and RNAs that interact tightly to generate a compact, self-contained cellular structure (Eddy, 1975). Due to its recognisable shape and fairly big size, the germ plasm was identified for the first time in the 19th century (von Wittich, 1845; Hubbard, 1894) and even recent studies showed that the PGCs tend to be bigger than the same-stage somatic cells and to be less round (Braat et al., 1999). The germ plasm is a common feature of many animals, but its composition and initial localisation within the embryo show species-specific peculiarities.

After fertilisation, the germ plasm is dragged usually to the periphery of the developing embryo where it will define the germ cells (Strome and Wood, 1983). Germ plasm dynamics vary slightly among organisms and these will be discussed further during this thesis.

The contribution of germ components to the development of the PGCs was initially demonstrated by transplantation experiments. Removal of germ plasm inevitably impedes germ cell formation and leads to sterility (Buehr and Blackler, 1970). On the other hand, the

inception of extraembryonic germ plasm into blastomeres was sufficient to induce differentiation of PGCs and formation of mature gametes in *D. melanogaster* (Blackler, 1962; Illmensee and Mahowald, 1974). This work clearly demonstrated that ectopic germ plasm transplanted into somatic tissues of a surrogate embryo can convert them to germ cells. In addition, the transplantation of germ cells from the animal-pole of *X. laevis* embryos to any different location in a surrogate embryo could generate migrating PGCs (Tada et al., 2012). However, when *X. laevis* germ cells were transplanted from the gonadal ridge to blastocoel cavity (outside the gonads) of a host animal, they tended to differentiate into somatic tissues rather than developing into the germ line (Wylie et al., 1985). This suggested that the microenvironment and cell-cell signalling is required to raise and maintain germ cell identity. Taken together these results show that although the germ plasm is sufficient to form functional germ cells, other stimuli can heavily impact the fate of the germ line, which is therefore dependent on its location within the embryo.

1.2.2.1 Onset of germ cell development in the fruit fly D. melanogaster

In *D. melanogaster*, the germ plasm forms in the oocyte from a maternal polarised structure known as Balbiani body (Bb), which restricts several germ factors at the vegetal pole of the egg (Ephrussi and Lehmann, 1992). Upon fertilisation, nuclear divisions are not followed by cytokinesis and the embryo develops as a syncytium, where many nuclei share the same cytoplasm. Approximately, twenty nuclei situated at the embryonic posterior pole, soaked in the maternal localised germ plasm will develop into germ cells. At gastrulation (Stage 6), the midgut invagination carries the PGCs towards the lumen (Seifert and Lehmann, 2012). From this site, the PGCs migrate attracted by 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase (Hmgcr)-expressing mesoderm to guide them toward the developing gonads (Van Doren et al., 1998). The germ fate is regulated from this point at two levels: retention of pluripotency and germ cell identity through RNA-Binding-Proteins (RBPs) such as Oskar,

Vasa and Tudor (Ephrussi and Lehmann, 1992), and inhibition of somatic differentiation though *slam*, *nullo*, *bottleneck* (Lecuit et al., 2002; Schejter and Wieschaus, 1993).

1.2.2.2 Onset of germ cell development in the worm

In *C. Elegans* the two cell divisions that follow fertilisation produce four founder cells: ABa, ABb, EMS and P2 that generate the nervous system, hypodermis, muscle/intestine and muscle/germ line respectively. The germ line eventually originates from the P2 cell lineage, generating a germ cell precursor after two cell divisions, the P4. The P4 will exclusively give rise to germ cells (Figure 1.4).

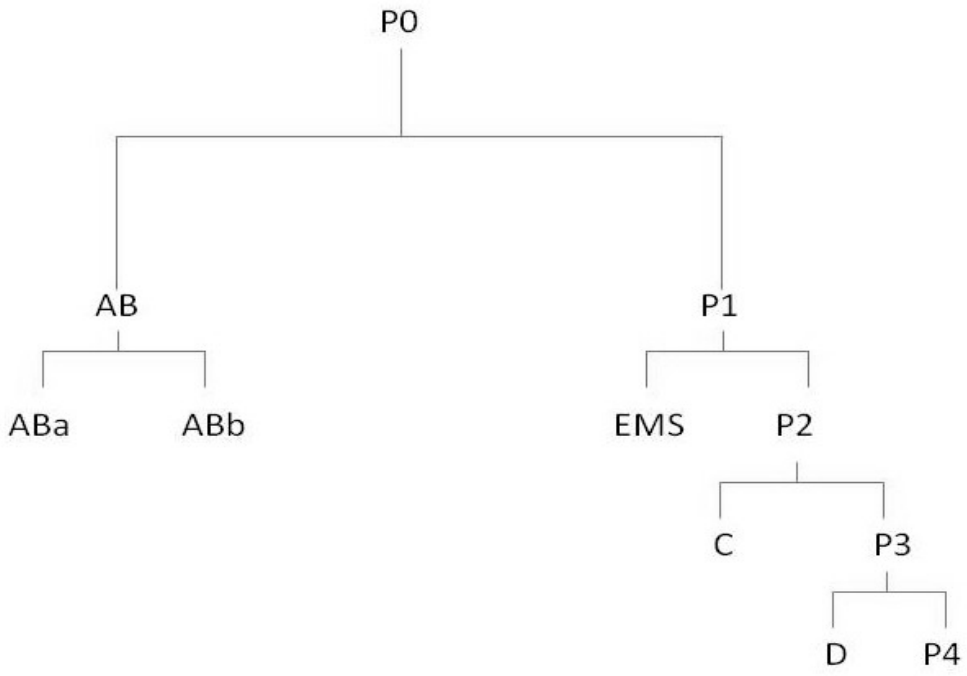

Figure 1. 4: Lineage specification in *C. elegans*.
The funder cell P0 (zygote) generates two ESCs upon division, AB and P1. AB generates ABa and ABb that form the nervous system, hypodermis respectively. When divides, P1 generates EMS (muscle/intestine) and P2 (muscle/germ line). Muscle and germ line will terminally separate after two cell divisions, with the formation of P4 blastomere that uniquely inherits the germ granules to form the germ line.

Similarly to *D. melanogaster*, in *C. elegans*, maternally-deposited P-granules (equivalent of germ granules or germ plasm) selectively segregate with the P lineage and are asymmetrically distributed at each cell division until formation of P4 (Strome and Wood, 1983).

Notably, it was shown that P-granules are necessary but not sufficient to form the germ cells in *C. elegans*. For example, *mes-1* mutants fail to exclusively distribute the P-granules to the blastomere P4. As a result, also the blastomere D (precursor of muscles) acquires P-granules. Interestingly, D still commits towards somatic differentiation regardless the presence of P-granules, suggesting that extra signalling is needed to form the PGCs (Strome et al., 1995). One could hypothesise that P-granules carry maternal factors in charge of initiating the germ fate in any blastomere, however, microenvironment and cell-to-cell signalling would also be required.

Potential maternal factors responsible for epigenetic reprogramming of the germ line were identified through the study of a family of histone methyltransferases (MES). In particular, the enzymes maternally-inherited MES-2/3/6 (the worm polycomb repression complex 2) and MES-4 cooperate and compete to control gene expression in the germ line by catalysing histone modifications. The proteins MES-2/3/6 generate repressive H3K27me3 on somatic genes and on the X chromosome in the germ cells (Bender et al., 2004), keeping these repressed. On the other hand, MES-4 induces H3K36me3 on germ line genes and repels repressive H3K27me3 (Bender et al., 2006; Rechtsteiner et al., 2010).

Modes of germ fate acquisition and protection will be mentioned later in this thesis.

1.3 Features of PGC development in zebrafish

1.3.1 Zebrafish as model for molecular visualisation and medical genomics

Zebrafish (*Danio rerio*) is a fresh water fish, which has become an excellent tool for the research community in the recent years. It belongs to the Cyprinadae family and it naturally lives in East-Asian rivers. Zebrafish is characterised by small size also in adulthood (usually < 120 mm) and by characteristic horizontal stripes awarding it its popular name. Zebrafish became a popular research model in the mid-70s, when George Streisinger proposed zebrafish in replacement of the widely used *D. melanogaster* and *X. laevis* models. To date, over 1000 laboratories around the world perform research using zebrafish as a model organism and nearly 30000 publications have taken advantage of this model system (Varga et al., 2016).

There are several advantages to using the zebrafish model. Firstly, it is easy and inexpensive to maintain due to its small size and the limited amount of water it requires. The embryo is relatively large compared to the adult body size and it develops naturally outside the mother. These are two fundamental features that make a zebrafish embryo easy to manipulate and microinject. Additionally, the embryo is transparent through most of its embryonic and larval stages making it extremely suitable for imaging.

The zebrafish generation time is relatively short ranging between 2 and 4 months and comparable to the mouse model, which generation time is approximately 3 months. However, while the gestational time of a mouse is 20 days, zebrafish development occurs outside the uterus and gastrulation and organogenesis are achieved much faster. After fertilisation, the initial pace of cell division is 15 minutes and gastrulation onsets only 4 hours post fertilisation (hpf). At 8 hpf, the neural tube is closed and somitogenesis begins. Already 20 hours after fertilisation, most of the primordial organs are formed: from the brain, composed of fore-, mid- and hind-brain, fully-developed eyes, the heart and the blood stream, as well as the notochord, the myocytes and the germ cells (Kimmel et al., 1995). At this stage, the germ cells have

migrated to the genital ridge and distribute in two clusters along the right-left axis (Yoon et al., 1997). Germ cell development is very dynamic and trackable during the first hours of zebrafish embryogenesis and it was intensively studied in the recent years.

1.3.2 Development of the PGCs in zebrafish

Zebrafish PGCs are an attractive biological system and have been heavily used in the past years to study patterns of gene expression and developmental processes. Several visualisation methods were utilised, focusing mostly on their localisation and migration. The identification of PGC-localised proteins allowed the generation of several transgenic lines expressing germ cell factors tagged with fluorescent proteins (Blaser et al., 2006; Riemer et al., 2015). The development of in-situ hybridization techniques and generation of transgenic lines expressing fluorescently-tagged Vasa and Nanos heavily contributed to deciphering mechanisms of PGC formation and migration (Braat et al., 1999; Köprunner et al., 2001; Yoon et al., 1997). Moreover, several germ cell markers were firstly discovered in zebrafish and later identified in other organisms. Examples are *dead-end* (*dnd*) (Weidinger et al., 2003), which is conserved in several species, and the zebrafish-specific *bucky ball* (*buc*) (Bontems et al., 2009; Marlow and Mullins, 2008). Moreover, zebrafish has greatly contributed to elucidating several aspects of chemokine-driven germ cell development (Doitsidou et al., 2002; Knaut et al., 2003). Nevertheless, the small number of embryonic PGCs has limited biochemical and genome-wide studies so far. Hence, our current knowledge on germ fate-driving molecular pathways, metabolism and transcriptional regulation is still limited.

As previously mentioned, the germ plasm is an agglomerate of maternal inherited proteins and mRNA, important for specifying the germ fate. The localisation of the germ factors in zebrafish occurs during oogenesis, when several ribonucleoparticles (RNPs), including maternal mRNA encoding the proteins Dazl and Bruno-like, are accumulated in the vegetal pole (Maegawa et

al., 1999; Suzuki et al., 2000), whereas some others, like Nanos3, Dead-end (Dnd) and Vasa are found in the animal pole (Braat et al., 1999; Köprunner et al., 2001). As soon as fertilisation occurs, the animal and the vegetal germ plasms co-localise along the cleavage planes and form four distinct agglomerates at the 4-cell-stage (Theusch et al., 2006) (Figure 1.5B, C and D). The early translocation of the germ plasm is driven by cytoskeletal microtubules that separate the four masses by recruitment of germ plasm RNPs to the division furrow (Eno and Pelegri, 2016). At this point, the four germ plasm agglomerates stick to the division planes until they are internalised into the PGCs at the 32-cell stage (Figure 1.5G). From this time, the PGCs undergo asymmetric divisions for eight cell cycles and will remain at the four poles of the blastodisc until the transitions from subcellular to cellular distribution of germ granules occurring between the sphere and the dome stage.

After ZGA, the asymmetric cell divisions stop and the number of PGCs increases. At the beginning of gastrulation, the PGCs leave the four poles by a first step of migration: at the 70% epiboly stage, all the PGCs are found in the hypoblast (Yoon et al., 1997). PGC migration continues during the following stages and the onset of somitogenesis corresponds to the alignment of the PGCs on the anteroposterior axis on both sides of the first somite (Weidinger et al., 1999).

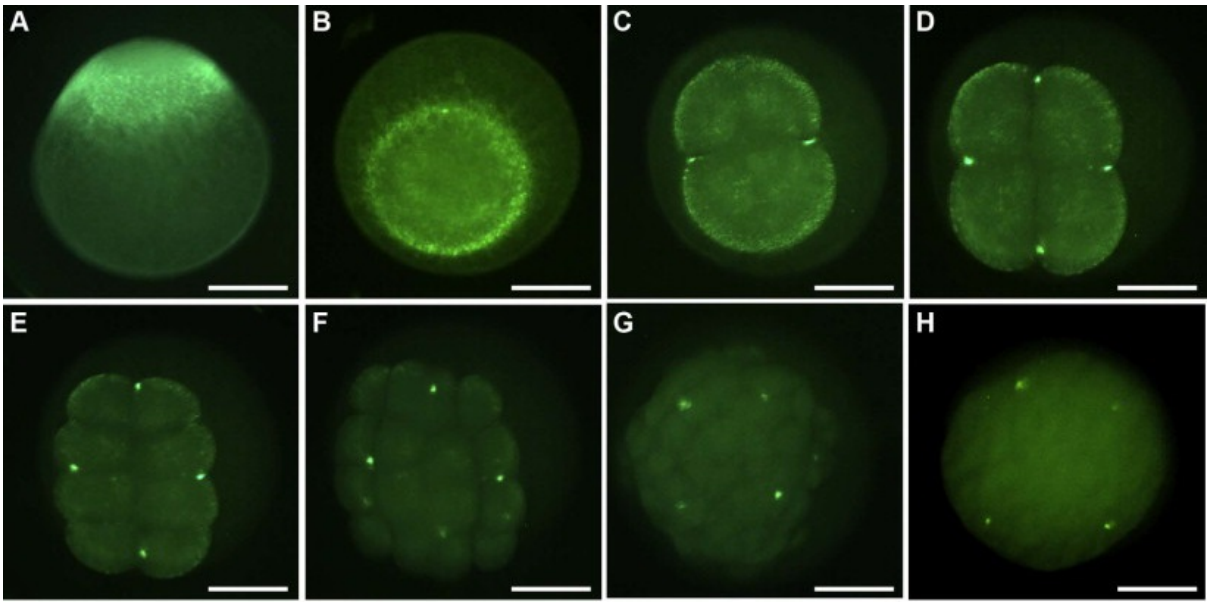

Figure 1. 5: Germ plasm visualisation through Buc-GFP transgene.
Confocal microscopy showing animal views of developing zebrafish embryos (Riemer et al., 2015). (A-B) Germ plasm material is localised at the edges of the fertilised cell and (C-D) segregates with the cleavage furrows following division symmetry during the first two cell cycles. (E-F) The germ plasm remains extracellular during the following two divisions (G) until it ingresses the cells at the 32-cell stage.

1.3.3 Many zebrafish germ plasm-localised factors are conserved among organisms

The early specification of the PGCs requires maternally supplied factors, in particular mRNAs and RBPs accumulated in the germ plasm. Until the zygotic genome activation, these molecules contained within the germ plasm are the exclusive determinants of the future germ line and required for a correct formation and survival of the PGCs. Although many germ factors were identified and associated with germ cell features, most of their functions are still not well understood. Several germ plasm-relegated mRNAs are important in the inhibition of somatic fate and favour the differentiation and maintenance of the germ cells while others are crucial for migration. It is important to highlight that many germ plasm-contained factors have RNA binding and/or RNA processing ability, indicating that the germ plasm acts as a post-transcriptional-regulatory factory in where selective RNAs are processed.

Nanos, for example, is one of the most known germ cell factors, conserved from *D. melanogaster* to human. Nanos is a RNA-binding-Zinc-Finger-Protein discovered in *D. melanogaster* and acts as a translational repressor (Kobayashi et al., 1996). In *D. melanogaster*, *nanos* is not required for the specification of the PGCs but for their maintenance. In fact, *nanos* was shown to prevent somatic and oocyte differentiation in the PGC by inhibiting expression of meiosis specific proteins (Forbes and Lehmann, 1998; Suzuki et al., 2010). In addition to this, lack of *nanos* blocks migration of the PGCs towards the gonads. There are three *nanos* homologs known in zebrafish, *nanos1*, *nanos2* and *nanos3* and show redundant functions (Beer and Draper, 2013; Draper et al., 2007; Köprunner et al., 2001). *Nanos3* (initially defined *nanos1*) was identified by sequence similarities with the fly homolog and in-situ hybridization screening. *Nanos1* mRNA is detectable from 1-cell stage until 5 days post fertilisation (dpf) in the zebrafish embryo and it shows co-localisation with the germ plasm marker *vasa*. When the translation of *nanos3* is inhibited, the migration and survival of the PGCs are affected (Köprunner et al., 2001).

A similar phenotype caused by lack of Nanos3 is observed in *piwi*$^{-/-}$ mutants, which also fail to generate functional germ cells (Houwing et al., 2007). Piwi belongs to the Argonaute protein family, which groups proteins presenting PAZ and Piwi amino acidic domains and it is involved in several germ cell-specific functions. Piwi protein was initially identified in the germ plasm of *D. melanogaster* and it was associated with germ fate maintenance and RNA processing (Cox et al., 1998). In mammals, Piwi protein is also known to interact with several small RNAs (piRNAs) and to block deleterious expression of transposons in the germ line (Brennecke et al., 2007; Houwing et al., 2007). Zebrafish Piwi (Ziwi) was shown to co-localise within the germ plasm during early development. Interestingly, loss of Ziwi was reported to induce germ cell apoptosis, adult infertility and transposon activation (Houwing et al., 2007).

As well as Piwi, Dead-end (Dnd) is an RNA-binding-protein detected exclusively in the germ plasm. This protein is required for initiating the migration of the PGCs and it is involved in E-caderin formation during PGC migration. Another suggested function of Dnd is to antagonize the inhibitory activity of some miRNAs, allowing PGC-specific genes to be transcribed (Kedde et al., 2007). Moreover, when morpholinos targeting Dead-end are used, PGC migration and proliferation are affected, meaning fewer PGCs are observed or found at non-canonical positions (Weidinger et al., 2003). Until recently, it was thought that lack of Dnd protein led to germ cell death as most of the known germ cell markers is lost upon morpholino treatment. However, by tracking the PGCs via a marker independent from the germ fate, it was recently shown that Dnd-lacking PGCs do not die, but differentiate towards other cell types upon failure on reaching the genital ridge (Gross-Thebing et al., 2017).

Another germ cell protein conserved among clades is Vasa. Vasa was the first germ cell marker identified in zebrafish and studies on its localisation pioneered the understanding of the germ line development in fish (Yoon et al., 1997). *Vasa* mRNA encodes for an RNA-helicase belonging to the DEAD-box family and it is strictly required for the transcription of *nanos* and *gurken* genes in *D. melanogaster* (Hay et al., 1998; Styhler et al., 1999). Like *nanos*, *gurken* was described to control oocyte polarity in *D. melanogaster* under the control of *vasa* (Tomancak et al., 1998).

The discovery of the *buckyball* gene (*buc*) has shed light on the mechanisms behind oocyte polarisation and germ plasm organisation. Bucky ball was identified as an oocyte polarising factor during mutagenesis screening on zebrafish embryos (Bontems et al., 2009). *Buc*$^{-/-}$ embryos showed a round phenotype, without any sign of Axio-Ventral polarisation, lack of any

embryonic structure and inability to organise the germ plasm (Dosch et al., 2004). The polarising ability of Bucky ball is due to its disorganised protein structure, which allows Bucky ball to function as a scaffold by interacting with a high number of targets. Bucky ball is also involved in structurally organising the Balbiani body, one of the first polarised structures of the embryo, starting being localised during oocyte maturation (Houston and King, 2000). The Balbiani body is an oocyte-specific subcellular structure including mitochondria, Golgi's apparatus and germ plasm. Upon fertilisation, Buc protein is found exclusively in the germ plasm where it builds a molecular mesh with ribonucleoparticles (RNPs), supported by Tdrd6 (Roovers et al., 2018).

Interestingly, while Buc is found exclusively in zebrafish, the protein Oskar (Osk) is responsible for germ plasm organisation and defines the posterior pole during *D. melanogaster* embryogenesis (Kim-Ha et al., 1995; Nakamura et al., 2004). Similarly to Buc, Osk also controls recruitment of germ plasm components, such as Nanos and Vasa, within the germ granules and regulates germ granules assembly (Ephrussi and Lehmann, 1992; Lehmann, 2016). The fact that both Oskar (Osk) and Buc contribute to germ plasm organisation was interpreted as an example of independent functional evolution (Krishnakumar et al., 2018). In fact, these two proteins do not share any sequence conservation, despite the fact that supply similar functions. Therefore, it was suggested that the several intrinsically disordered regions found in both Osk and Buc structures allow these two proteins to accomplish similar functions despite the major differences in amino acid composition. Surprisingly, injection of Osk in *buc*$^{-/-}$ mutant embryos rescued the lack of PGCs in a zebrafish embryo, confirming that Osk is able to induce PGCs in absence of Buc (Krishnakumar et al., 2018).

1.3.4 Germ plasm dynamics and germ cell migration

The dynamics of PGC development in zebrafish were largely studied and understood in the last decades, due to the discovery of important markers and to the easy accessibility to transparent fish embryos. Precisely, the discovery of the *D. melanogaster* gene *vasa* led to the identification of a zebrafish orthologue mRNA specifically localised in the PGCs. During this study, whole-mount in-situ hybridization provided the first view of zebrafish PGC developmental dynamics from the fertilised egg until 72 hpf (Yoon et al., 1997).

Similar to the most of the organisms preforming the germ line, the germ plasm localisation in zebrafish begins in the unfertilised eggs. During oocyte activation, several RNPs and mRNAs, including *dazl* and *bruno-like*, are accumulated at the vegetal pole (Maegawa et al., 1999; Suzuki et al., 2000), whereas *nanos*, *dead-end* and *vasa* are found at the animal pole (Braat et al., 2001; Köprunner et al., 2001). As soon as fertilisation occurs, the vegetal germ plasm is pulled along the zygote and reaches the animal pole. After this event, the animal and the vegetal germ plasms co-localise along the cleavage planes in a coordinate manner, aimed to initiate the germ identity within four initial PGCs (Theusch et al., 2006). During the first embryonic cell division, the aster drives the multimerization of RNPs scattered around the zygote. The germ granules are bound by the aster microtubules through the Birc5b protein and pushed towards the edges during cell division upon microtubules extension (Nair et al., 2013). As the second cell cycle occurs, two bipolar microtubule asters nucleate at the spindle pole, pulling the germ granules along the furrow (Eno and Pelegri, 2013). The recruitment of the germ granules to the furrow continues until the embryo reaches the 32-cell stage, when the germ plasm enters the blastomeres, giving rise to four finite cytoplasmic masses equally distributed to the poles of the animal cap (Theusch et al., 2006; Yoon et al., 1997). After this point, the future germ cells asymmetrically divide, passing the germ line determinants to a single daughter cell. At these stages, the germ plasm appears as a single, bean-shaped aggregate within the blastomere

cytoplasm, made of additional small granules spreading both in the PGCs and in the adjacent somatic cells (Bontems et al., 2009). The first symmetrical division of the germ plasm-carrying cells is detected following the spread of the germ granules across the cytoplasm around dome stage. After this transition, 8-10 PGCs redistribute within the embryo, showing acquisition of migratory behaviour (Eno and Pelegri, 2016; Raz, 2003). By the 70% epiboly stage the PGCs are in the hypoblast (Yoon et al., 1997) At this stage, the germ plasm approaches to the nuclear periphery and slowly fragments in small granules, which are found in proximity of the nucleus at 10hpf (Strasser et al., 2008). This behaviour firmly continues at least until the prim-5 stage (24 hpf) (Figure 1.6). Similarly, in *X. laevis* the germ plasm was found to migrate from plasma-membrane proximity to perinuclear at the beginning of gastrulation (Taguchi et al., 2012), while in *D. melanogaster* germ granules were shown to interact with nuclear pores (Pitt et al., 2000). The intimate interaction between the germ granules and the nucleus, recurrent among organisms and coinciding with onset of PGC migration, could be indication of a crosstalk mechanism between cytoplasmic and nuclear factors, with potential influence on germ line transcriptional programme (Figure 1.6).

As soon as somitogenesis begins, the PGCs align on both sides of the first somite along the anteroposterior axis (Weidinger et al., 1999). From these two sides, the PGCs slide posteriorly and reach the primordial gonads located at the anterior level of the 8-somites, where they complete their migration into the genital ridge (Yoon et al., 1997).

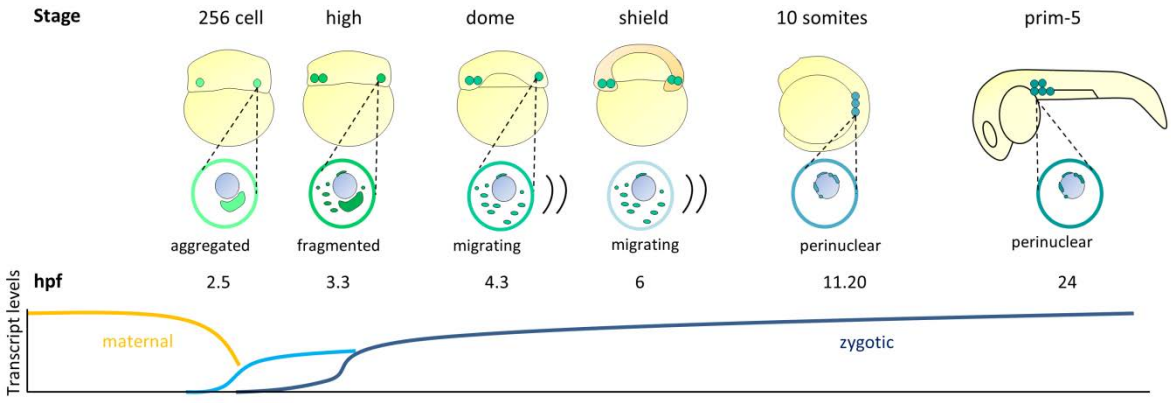

Figure 1. 6: Schematic of PGC formation during the first 24 hours of zebrafish development. Germ cells are depicted as shading from green to blue at each reported developmental stage. The schematic summarises PGC proliferation, migration and genital period. After the first wave of ZGA (light blue curve), the germ cells synchronously divide and the germ plasm spreads across the cytoplasm. During migration, the germ plasm subcellular re-localisation is observed and four steps can be identified (aggregated, fragmenting, spread and perinuclear). Hpf indicates hours post-fertilisation.

1.3.4.1 Migration of the primordial germ cells in zebrafish

Germ cell migration is a conserved mechanism among metazoans, which ensures pre-selection of 'capable' cells during germ line formation. Indeed, the PGCs arise away from the site where the genital ridge develops, therefore this could be seen as a mechanism to filter out those cells that do not express the required genes. One could imagine that this occurs because germ factors have multiple functions. In fact, germ plasm factors involved in migration (such as *dnd1* and *vasa*) perform also other functions throughout germ cell development and are often required for gonadal PGC survival (Köprunner et al., 2001; Weidinger et al., 2003), suggesting that germ cells could be 'tested' on expression of germ factors via migration. Accordingly, those PGCs that fail to reach the genital ridge undergo apoptosis or somatic differentiation (Gross-Thebing et al., 2017).

Even though the mechanisms guiding PGC migration are conserved among vertebrates and invertebrates, some PGC features are exclusively observed in zebrafish. In *D. melanogaster*, *C. elegans* and *X. laevis*, the germ cells specify at the posterior edge of the embryo whereas in zebrafish, the PGCs form in four random locations with respect to the dorsoventral axis.

Consequently, the PGCs must have the ability to migrate towards the genital ridge from any starting point (Weidinger et al., 1999). Correct migration is achieved through a multistep process, involving loss of cell adhesion and chemotaxis. Loss of cell adhesion is the first step that the PGCs need to undertake in order to migrate and depends on Dnd. Immunostaining of E-cadherin in zebrafish PGCs showed significant decrease of E-cadherin levels in migrating germ plasm-carrying cells. Interestingly, the downregulation of E-cadherin was not observed in *dnd* knock-down embryos (Blaser et al., 2005).

After loss of cell adhesion, PGC migration is chemically-driven by ligand-receptor interactions. PGCs express the C-X-C chemokine receptor type 4b (Cxcr4b) on the membrane surface that binds the ligand Cxcl12a (Doitsidou et al., 2002). Upon chemokine recognition, the PGCs form a cytoplasmic protrusion that allows directed movements via myosin contraction (Blaser et al., 2006). In zebrafish, the migration movements begin at dome stage (4 hpf) when four clusters of germ plasm-carrying cells are found at the blastoderm margin (Figure 1.7A). The early gastrulation movements push most of the PGCs towards the dorsal side from their location, while a smaller cluster could persist in the ventral side (Figure 1.7B, C). When gastrulation finishes, the dorsal PGCs align along the two sides of the trunk, while ventrally-located PGCs distribute along the lateral border of the mesoderm (Figure 1.7D, E). During somitogenesis, the ventral PGCs migrate anteriorly and unite with the dorsal cluster (Figure 1.7F). The two clusters now move towards the yolk extension (where the yolk meets the trunk) either from the anterior or posterior axis and reach the genital ridge by the end of somitogenesis (Figure 1.7G).

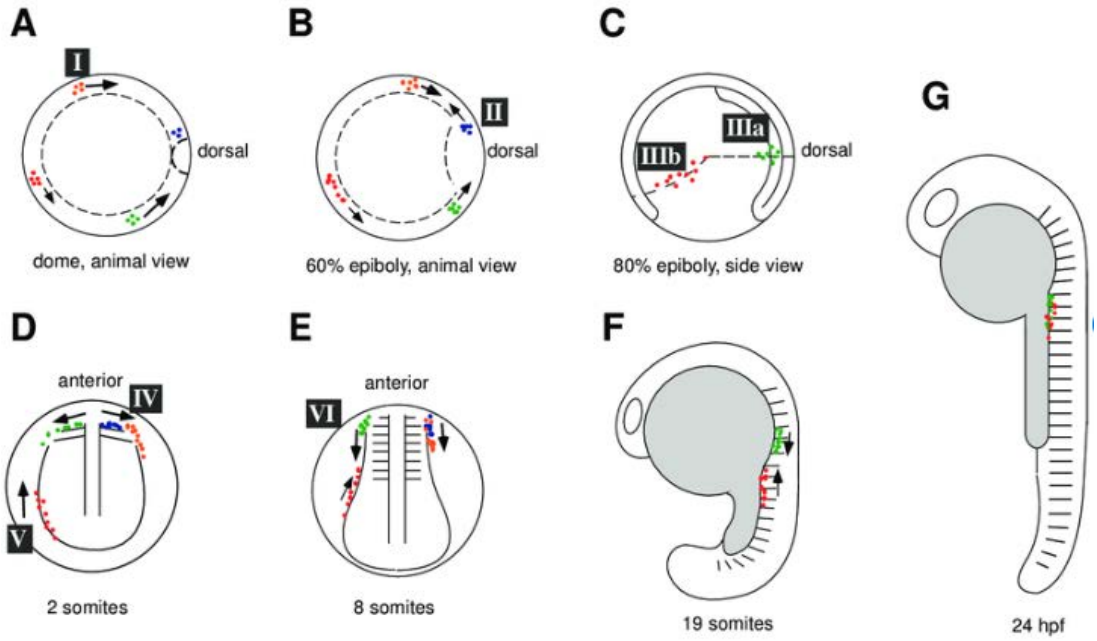

Figure 1. 7: Schematic of germ cell migration in zebrafish.
Roman numbers indicate steps of PGC migration. (A) At dome stage, four cluster of PGCs are found at the edges of the blastodisc. (B) These clusters tend to move towards the dorsal side, however, some lateral movement can be observed. (C) At the end of gastrulation, clusters of PGCs located dorsally align to the anterior border (dashed line), while ventral cluster align to the posterior border (IIIb). (D-E) The PGCs slide away from the anterior along the lateral side of the embryo and arrive where the somites are forming. (F) During the end of somitogenesis, PGCs are still moving towards the genital ridge which is located at the site where the yolk meets the trunk. (G) At 24 hpf, the PGCs are located in the genital ridge. From (Weidinger et al., 1999).

1.4 Transcriptional and post-transcriptional contribution to PGC formation

1.4.1 Zygotic genome activation

As mentioned, in cases when PGC formation is maternally-driven (via the germ plasm), the germ fate is significantly impacted by the zygotic genome activation (ZGA). The ZGA is known as the process that liberates the embryonic cells from the block of maternal transcriptional inhibitors and is observed in all metazoans (Kobayashi et al., 1988; Lamb and Laird, 1976; Zalokar, 1976). The ZGA proceeds differently in slowly-developing embryos (for instance mammals) and fast-developing embryos such as zebrafish, frog, worm and fly. In mouse and human, approximately 24 hours pass for the first embryonic cell division to occur, while in zebrafish this takes less than half an hour. In zebrafish, the pre-ZGA period formally matches the first 10 division cycles of the embryo (about 3 hours). It follows that the transcriptional silent gap is longer in the mouse but spans more cell cycles in the fish.

The ZGA is not evenly timed in the embryo and usually occurs in two waves: the early activation of a small number of genes (often required for initiating maternal RNA degradation) is defined minor wave, while the major wave follows the minor wave and is characterised by intense zygotic gene expression (Tadros and Lipshitz, 2009).

Importantly, the cell fate is tightly linked to the time and type of ZGA. Immediately after fertilisation, the zygote spends a vast amount of maternal energy in protein and DNA synthesis. Especially in extra-uterine developing organisms, DNA replication occurs very rapidly. During these rushed cell divisions, transcription does not occur, therefore protein synthesis is performed from maternal RNAs. However, few early genes skip transcriptional repression during the first minor wave of genome activation. These genes often are involved in global genome activation and clearance of maternal RNAs (Giraldez et al., 2006; Harrison et al., 2011).

The mechanisms and dynamics of ZGA were extensively studied in the past years, however the discrimination between maternal and zygotic transcripts remains a challenge. To overcome the problem of maternally-loaded RNAs masking those few early genes upregulated in the early stages of genome activation, techniques such as Nascent-RNA-seq, cap-analysis-of-gene-expression (CAGE) -seq and imaging were used (Haberle et al., 2014; Hadzhiev et al., 2019; Heyn et al., 2014). Currently, what triggers ZGA is still disputed. Three different hypotheses still stand that could explain the onset of zygotic transcription: the nuclear-cytoplasmic ratio, chromatin remodelling and translation of maternal activators. These will be further discussed in the following sections.

1.4.2 Models for mechanisms of ZGA

1.4.2.1 Nucleo-cytoplasmic ratio

According to the nucleo-cytoplasmic ratio model, zygotic gene expression does not take place until a maternally-deposited transcriptional repressor is titrated away by the increasing number of nuclei forming during the fast blastocyst divisions. In fact, the unusual, massive size of the zygote could impede nucleus-cytoplasm cross-talks. This immense difference between the nuclear and the cytoplasm sizes rapidly changes during the early cell cycles, which halve the dimension of a cell at each cell division. At the same time, the DNA mass doubles; that means that progressive cell divisions balance the DNA/cytoplasm amount.

Such dramatic changes could explain the activation of the zygotic genome. This is supported by the evidence that polyspermic *X. laevis* embryos activate zygotic transcription two cell cycles earlier than monospermic embryos (with less nuclei) (Newport and Kirschner, 1982). Results from this work also proved that there is no 'biological clock' or a required number of cleavages to trigger the ZGA. Additionally, polyploid zebrafish embryos generated by inhibition of chromosomal segregation tend to initiate the zygotic transcription earlier

compared to wild-types (Dekens et al., 2003), thus suggesting that an increased nuclear-cytoplasmic ratio can drive the ZGA.

Overall, the nucleo-cytoplasmic ratio adjustment is an observable phenomenon, however, how does it trigger the ZGA? It was suggested that the switch could be caused by dilution of a maternally-deposited transcriptional repressor. In fact, although the DNA amount exponentially increases at each cell division, the amount of a hypothetic maternal inhibitor would not; therefore, any eventual effect of an inhibitor will be mitigated upon increment of nuclei. Potential maternal inhibitors of the ZGA were identified as the histones subunits H3 and H4. When translation of H3 was blocked in pre-ZGA *X. leavis* embryos, transcriptional activation was detected earlier (Amodeo et al., 2015). Supporting this, the addition of histones in zebrafish embryos delayed the ZGA (Joseph et al., 2017). According to this model, an excess of histones loaded into the zygote could outcompete TFs binding and consequently delay transcriptional activation. This observation suggests that nucleosomes do not mask TF-binding sites directly on the DNA, but rather compete inside the nucleus with TFs for binding the DNA. In addition to histones, DNA-replication factors could play an important role in activating gene expression. During normal development of *X. laevis*, the S-phase is elongated prior to ZGA through the degradation of four eukaryotic replication factors: Cut5, RecQ4, Treslin, and Drf1. Interestingly, this has an effect on timing the ZGA. When these factors are overexpressed, extension of the rapid synchronous cell divisions beyond cell cycle 12 and delayed ZGA were observed (Collart et al., 2013).

1.4.2.2 Active chromatin regulation

A second hypothesis has suggested that reduction of cytoplasm sizes and translation of maternal RNAs could induce chromatin remodellers to access the nucleus and "reshape" the epigenome. In fact, although the fast cell divisions occurring before ZGA do not allow optimal chromatin organisation (Hug et al., 2017), the newly synthesized DNA needs to be promptly

packed with histones during the short cell cycle. Accordingly, the combination of unorganised chromatin with acquisition of transcriptional repressive histone modifications could explain the global silencing of the zygotic genome. Importantly, this hypothesis would explain how certain selected genes only can escape transcriptional inhibition and initiate the ZGA.

Experiments aimed to measure regions of accessible DNA and chromatin interactions on pre-ZGA embryos reported a high degree of DNA packaging and disorganised chromatin structure (Hug et al., 2017; Wu et al., 2011). During genome activation, the chromatin re-shapes and acquires a transcriptional-permissive landscape. In support of this, recent studies using zebrafish and flies demonstrated that cohesin and Zelda are required for the ZGA in the respective organism. Chromatin immunoprecipitation (ChIP) -seq experiments in pre- and post-ZGA zebrafish embryos for the cohesin subunit Rad21 showed that the cohesion complex increases its binding affinity before the ZGA begins and that it is found on actively transcribed promoters and putative enhancer regions enriched for H3K4me1 and H3K27ac. When Rad21 was knocked down using morpholino injections, chromatin organisation and typical RNA-Poll II accumulation in two nuclear foci was disrupted and ZGA was delayed (Meier et al., 2018).

In *D. melanogaster*, chromosome conformation capture experiments performed on pre- and post-ZGA synchronised embryos showed that enhancer-promoter interactions and topological-associated-domains (TADs) are absent before genome activation. Importantly, the TAD formation is independent of RNA-Poll II position and instead requires the TF Zelda (Hug et al., 2017). It remains unclear whether correct chromatin organisation is required to trigger genome activation or whether ZGA drives the formation of the TADs.

In mouse, studies of chromatin architecture before ZGA are complicated by the fact that embryonic transcription starts already at the 1-cell stage. It follows that a very limited amount of cells at these early stages are available for experiments, therefore next generation sequencing

(NGS) on DNA and RNA was inapplicable until recently. Recently, open chromatin profiling of 2-cell stage mouse embryos (when ZGA occurs) showed that chromatin is more accessible at transcription end sites (TESs) and highly enriched on repetitive elements and transposons. Interestingly, the overall positioning of newly-identified sites of accessible chromatin in the early embryo does not match those observed in the gametes, suggesting that upon ZGA the chromatin landscape is newly established (Li and Patel, 2016) .

In summary, although studies on chromatin regulation during ZGA remain challenging, new insights are suggesting that epigenetic landscape and epigenetic memory could play a role during onset of ZGA.

1.4.2.3 Transcriptional activators

As transcription is a highly regulated process, it requires both repressors and activators to happen. Genes are controlled in their transcription in space and time often by a combination of TF binding, chromatin accessibility and enhancer activation that act synergistically. In a pre-ZGA embryo, some or all of these listed factors may not be immediately available, resulting in transcription block. For instance, many transcription factors are maternally deposited as RNAs and translation has to take place before the proteins can reach their targets in the nucleus. In zebrafish, Nanog, SoxB1, and Pou5f3 were reported to trigger transcription of the majority of early-expressed zygotic genes. These factors have been identified as maternal transcripts highly translated before ZGA and their depletion caused arrest of development before gastrulation indicating failure of genome activation (Lee et al., 2013).

Nanog, SoxB1 and Oct4 (homolog of Pou5f3) guide transcription in the early mouse embryo and are candidates for triggering transcriptional activation at ZGA. Oct4 is enriched in the mouse oocyte, however the ZGA still occurs when Oct4 is depleted, suggesting that a combination of other factors may compensate for its loss (Foygel et al., 2008; Wu et al., 2013). Together with Oct4, other candidate TFs potentially involved in the ZGA are Sox2, Yap1 and

TIF1a (Abbassi et al., 2016; Torres-Padilla et al., 2006). Recently, the murine Dux and the human DUX4 were implicated in activation of several zygotic genes by binding in proximity of their promoters. Moreover, Dux depletion in mouse ESCs impedes the cells to transition to the state equivalent to a 2-cell mouse embryo (De Iaco et al., 2017).

1.4.2.4 The developmental importance of zygotic genome activation

Activation of zygotic transcription is crucial for embryonic survival. It has been shown that the minor wave of zygotic transcription in *D. melanogaster* is essential to reorganise the distribution of the nuclei along the embryonic antero-posterior axis before cellularization begins (Blankenship and Wieschaus, 2001). When transcription is inhibited in *D. melanogaster*, cellularization of the syncytium does not occur, however the early nuclear divisions are not affected (Edgar et al., 1986; Merrill et al., 1988). Similarly, in zebrafish and frog, block of ZGA results in developmental arrest at gastrulation without evident phenotype on the preceding stages (Kane et al., 1996; Newport and Kirschner, 1982).

The primary role of the ZGA is to address maternal transcript clearance. Genes involved in selective degradation are transcribed during the first minor wave and tend to be short and in high copy number (Giraldez et al., 2006; Heyn et al., 2014). Importantly, destabilization and degradation of maternal transcripts has been linked to 20 nucleotide (nt)-long RNA fragments known as micro-RNAs (mi-RNAs). In zebrafish, the *micro-RNA-430* (*miR-430*) was identified as responsible for the degradation of hundreds of maternal transcripts by selective binding of a recognition motif at the 3'UTR (Giraldez et al., 2006). Similarly, mi-RNAs were observed contributing to the degradation of maternal transcripts in *D. melanogaster* and *X. laevis* (Bushati et al., 2008; Lund et al., 2009).

Although zygotic transcription often initiates RNA clearance, maternal contribution is also required and oocyte-inherited molecules act in cooperation with newly-transcribed RNAs. For instance, the miRNA-mediated transcript decay is achieved through two maternal factors of

miRNA synthesis, Drosha and Dicer (Li and Patel, 2016). Drosha is a nuclear endonuclease responsible for the cleavage of long transcripts. Drosha recognises its targets after these are bound by miRNAs (Filippov et al., 2000; Lee et al., 2003). Those are exported into the cytoplasm where they bind target messenger. The resulting double-stranded RNA is recognised by the RNA-induced silencing complex (RISC) and the subunit Dicer catalyses their degradation (Macrae et al., 2006).

Alongside maternal/zygotic-coordinated processes of RNA degradation, there are maternal proteins able to induce selective RNA destabilisation on their own. These are normally deposited into the oocyte as transcripts and translated after fertilisation. An example is the fly RBP Smaug (SMG). SMG is translated by the Pan gu (PNG) Ser/Thr kinase complex in the fertilised egg and targets specific cis-acting elements on maternally-deposited transcripts. SMG triggers deadenylation of these transcripts by recruiting the CCR4/POP2/NOT-deadenylase complex, therefore leading to their destabilisation (Semotok et al., 2005).

Geno Ontology (GO) analysis was performed on maternal stable and unstable RNAs showing that transcripts involved in cell cycle and cell division are more likely to be degraded compared to those required for RNA processing (Tadros et al., 2007). This observation indicates that maternal RNA clearance is facilitated via selective transcript targeting, rather than general reduction in RNA abundance.

1.4.3 Zygotic genome activation is usually delayed in germ plasm-carrying cells

As discussed, the ZGA is crucial for correctly initiating downstream developmental events in any embryo. Interestingly, activation of zygotic transcription is temporally and spatially uneven within the embryo. In zebrafish, the first wave of zygotic activation occurs as a gradual acquisition of transcriptional ability from individual, sparse cells from as early as 64-cell stage

(Hadzhiev et al., 2019). Within a few extra cell cycles, transcription is observed in the whole embryo, resulting in the major wave of ZGA.

In several organisms, the germ plasm mediates selective regulation of the ZGA in the PGCs. In *D. melanogaster* and *C. elegans*, the somatic and germ line ZGAs are temporally distinct and germ plasm-carrying cells delay their ZGA of a few cell cycles (Figure 1.8A). Two germ plasm factors are indicated as responsible for temporarily inhibiting zygotic transcription via controlling the phosphorylation status of the C-terminal repeat domain (CTD) on polymerase II. In normal conditions, the positive transcription elongation factor b (P-TEFb) promotes transcriptional elongation upon phosphorylation of the serine 2 on the polymerase II CTD (Ahn et al., 2004). Notably, the fly polar granule component (pgc) protein, was shown to antagonize the CTD Ser 2 phosphorylation in developing germ cells, by inhibiting the recruitment of P-TEFb on gene promoters (Hanyu-Nakamura et al., 2008). Similarly, in *C. elegans*, P-TEFb is inhibited by the protein PIE-1 leading to general transcriptional repression (Batchelder et al., 1999; Mello et al., 1996; Zhang et al., 2003). Pgc and PIE-1 are found in both *D. melanogaster* and *C. elegans* and could function in cooperation to delay ZGA and initiate germ line commitment (Figure 1.8B).

It is noteworthy that in *C. elegans* the first and the second asymmetric divisions of the PGCs are transcriptionally controlled also by an alternative mechanism. RNA interference (RNAi) experiments demonstrated that depletion of OMA-1 and OMA-2 factors leads to a significant increase in the nuclear levels of the TATA-binding-protein associated factor 4 (TAF-4) (Guven-Ozkan et al., 2008). TAF-4 is required for the formation of the TFIID complex during the assembly of the pre-initiation transcription complex on the gene promoter, therefore its cytoplasmic sequestration by the OMA factors was linked to genomic silencing in the early PGCs in worm.

Importantly, transcriptional inhibition was also described in mouse PGCs, where suppression of RNA polymerase II is observed by immunofluorescence at E8.5 (Hopf et al., 2011; Saitou et al., 2002). In this instance, the ZGA has already occurred when the PGCs form from transcribing somatic cells, however, a period of transcriptional silence seems to be important to initiate germ cell development. It is unclear to what extent the genome of the murine PGCs is repressed. *Blimp-1*-positive cells were shown to actively repress expression of somatic genes, such as Hox genes, however, expression of *Stella*, *Sox2* and *Nanog* was observed (Kurimoto et al., 2008; Yamaguchi et al., 2005).

Taken together, these observations suggest that the inhibition of somatic differentiation is a conserved mechanism of germ fate acquisition and that global transcriptional repression in early PGCs may avoid expression of early genes involved in germ layer formation and gastrulation.

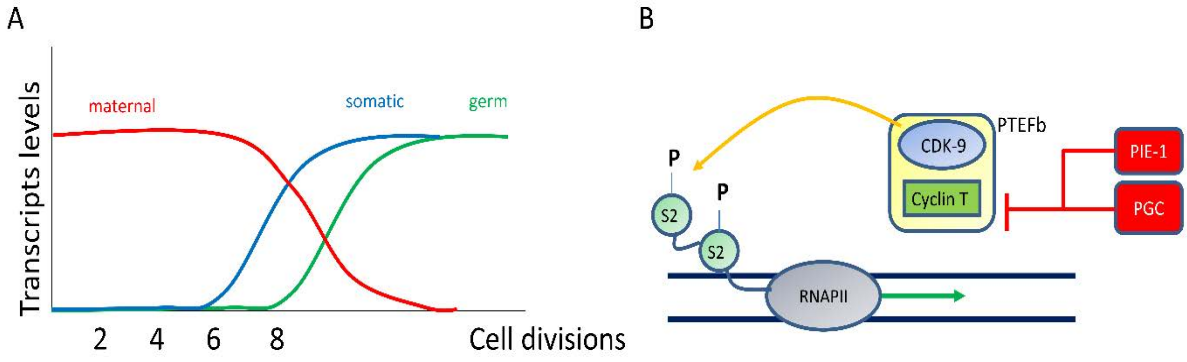

Figure 1. 8: Transcriptional inhibition in the PGCs.
(A) In *D. melanogaster*, transcriptional activity in the PGCs is detected two cell cycles later compared to the somatic cells. (B) Model of transcriptional inhibition mediated by the two germ plasm factors PIE-1 and pgc. Red lines indicate inhibition mediated by PIE-1 and PGC. Yellow arrow represents active phosphorylation of the Serines of the RNA polymerase II. Green arrow indicates transcriptional activation.

1.4.4 Alternative mechanisms of germ fate acquisition

In contrast to *C. elegans* and *D. melanogaster*, current studies report that the ZGA is not delayed in zebrafish PGCs. Immunostaining of Serine 2-phosphorylated polymerase II (active

polymerase), showed localised nuclear foci in germ plasm-carrying cells as early as 256-cell stage (Knaut et al., 2000). Although partially conclusive, detection of active polymerase II accumulations strongly suggested that transcriptional activation is not delayed in the PGCs. The evidence of active polymerase II in zebrafish PGCs during ZGA raises questions about eventual alternative mechanisms of germ fate acquisition. Due to the high amount of RNA-binding proteins populating the germ plasm, a role in translational regulation was suggested (Iguchi et al., 2006). In support, Musashi proteins, a family of translational repressors was found expressed in the germ line of *D. melanogaster*, *X. laevis* and *M. musculus* (Charlesworth et al., 2006; Siddall et al., 2006). Musashi-1 (Msi-1) was discovered to bind the 3'UTR of targeted RNA and prevent their translation by blocking the recruitment of the 80S ribosomal subunit through competition with eIF4G for the poly-A-binding protein (PABP) (Kawahara et al., 2008). A similar mechanism was identified in *D. melanogaster* and involves the interaction between the germ line proteins CUP and Bruno (Nakamura et al., 2004). It was shown that Bruno-mediated recruitment of CUP on the 3'UTR robustly inhibits translation via binding of CUP to the eIF4E factor (Figure 1.9A, B). Activity of CUP and Bruno was associated to inhibition of *nanos* and *oskar* translation (Wilhelm et al., 2003).

Translational repression is also observed in mouse. The RNA-binding-Protein DAZL, for example, inhibits the translation of several mRNA involved in pluripotency, somatic differentiation and apoptosis in the mouse PGCs (Chen et al., 2014). In zebrafish, DAZL is maternally loaded into the germ plasm although its function remains unexplored. Antibody interference experiments demonstrated that DAZL is a crucial regulator of PGC formation in medaka (Li et al., 2016), although a role in translational regulation has not been found so far.

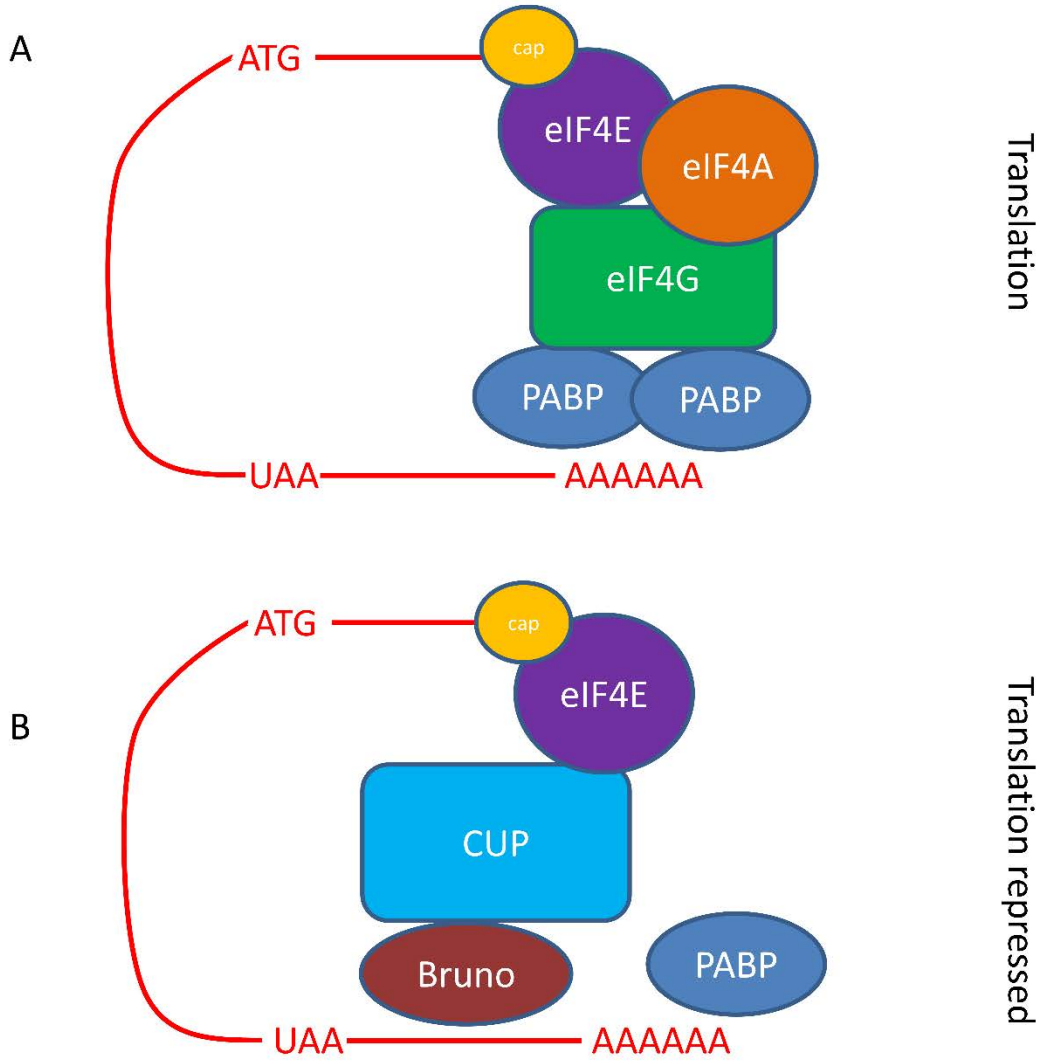

Figure 1. 9: Mechanism of translational inhibition in the PGCs.
(A) In normal condition the ribosome subunits 80S and 40S are recruited on the mRNA upon formation of the eukaryotic initiation complex. The poly-A-binding protein (PABP) recognise the poly-A tail of the messenger RNA and it is brought in close proximity of the cap marking the 5'UTR by the binding of eukaryotic initiation factors. The RNA circularisation promotes the recruitment of the ribosome at the 5'UTR. Active translation starts on the first codon for the methionine (ATG) and ends at the first termination sequence encountered by the ribosome (for example UAA). (B) The germ plasm CUP and Bruno compete with the eiF4G for the PABP, inhibiting the recruitment of the ribosome.

1.5 Gene regulation on the chromatin level

In order to pack the vast genetic information in a single nucleus, the cell has evolved strategies to wrap the DNA in dynamic hierarchical structures. The research field aiming to study DNA packaging and its effect on gene expression is known as epigenetics and involves every aspect of gene regulation that does not rely on changes of the DNA sequence (Bird, 2007).

In non-mitotic cells, the chromatin can be organised in euchromatin or heterochromatin, referring to opened and closed state, respectively. Euchromatin mainly characterises promoters and regulatory elements of transcribed genes, while heterochromatin organises non-coding-genomic region or silenced genes. According to this model, histones bind DNA and compete with transcription factors interactions and accessibility to the genes, controlling transcriptional activity within specific genomic regions (Margueron and Reinberg, 2010).

Chromatin accessibility plays a critical role in gene expression by affecting binding of transcription factors (TFs) and polymerases (Tsompana and Buck, 2014). However, the function of the chromatin is more complex than the mere switching between opened and closed conformation. Indeed, non-coding RNAs, enhancer looping and chemical modifications on DNA and histones all contribute to such transcriptional events (Kouzarides, 2007).

1.5.1 Post-translational histone modifications and gene expression regulation

Most of the somatic cells in the body share the same genetic code, however different cell types range considerably in shape, behaviour and gene expression in general. In order to allow cell type-specific differentiation and behaviour from the identical DNA sequences, the cell has evolved mechanisms of epigenetically control gene expression.

The chromatin is a complex of DNA, RNA and proteins whose main role is to ensure the genome undergoing hierarchical levels of compaction within the nucleus (Margueron and Reinberg, 2010). The basic chromatin unit is the nucleosome, consisting in a histone core, a

histone tail and a 147 bp DNA strand. The core has four histones (H2A, H2B, H3 and H4) in duplicate, which assemble in a single globular domain and eight histone tails. The H1 histone behaves as a linker and it is found between two nucleosomes, while the 147 bp DNA strand is wrapped around the core (Wolffe and Guschin, 2000).

Histone post-translational modifications usually occur on the histone tail and their function can have an impact on how proteins interact with DNA. Acetylation of lysines on histone tails, for instance, neutralizes the positive charge of the lysine, which usually strongly binds the negative-charged-DNA. When this interaction is contrasted, opening of the chromatin results in an increased likelihood of transcription. Histone acetylation is controlled by the activity of histone acetyl-transferases (HATs), enzymes catalysing the addition of acetyl groups on histone lysines, and histone deacetylases (HDACs), which, instead, remove the acetyl group (Lee and Workman, 2007; Leipe and Landsman, 1997). Equally, acetyl groups on histone tails flag regions of regulated chromatin, inducing subsequent events by the recruitment of protein effectors.

As well as acetylation, histone methylation is greatly involved in chromatin packaging. Like histone acetylation, methylation can also act directly or indirectly in regulating DNA compaction and transcription. Histone methylation can identify both eu- and heterochromatin. For example, H3K4me3 is often enriched at promoters of highly transcribed genes, while H3K36me3 is found on actively transcribed gene bodies (Liang et al., 2004; Ng et al., 2003). Regarding transcriptional inactivity of histone methylations, H3K9me3 is associated with long term repressed genes, while short term repression usually involves H3K27me3 recruitment (Barski et al., 2007).

Ubiquitination and phosphorylation modifications can also be found on histone tails. However, their role will not be discussed in this thesis.

As a consequence of the important role played by histone modifications in gene regulation, their genomic position can provide important information about the function of a certain genomic sequence. For instance, acetylated histones are enriched at cis-regulatory elements of actively transcribed genes. Specifically, acetylations on lysines 9 and 27 establish and mark euchromatin (Eberharter and Becker, 2002). Accordingly, genomic elements can be identified by specific histone marks. Active cis-regulatory elements, for example, are often marked by H3K4me3, H3K4me2, H3K4me1 and H3K27ac on flanking histones that keep the chromatin accessible and allow these elements to interact with activating complexes (Shlyueva et al., 2014). On the other hand, silenced genes show repressive histone marks on their promoter regions, mostly H3K9me3 and H3K27me3 (Margueron et al., 2005).

1.5.2 Chromatin features of cis-regulatory elements

The advent of NGS technology has immensely boosted genome-wide studies in the epigenetics field. High throughput sequencing can be applied to samples that are specifically selected for certain genomic features, such as accessible chromatin or histone-bound DNA. Integration of NGS techniques can be used to assign a function to a certain genomic location in a determined time point or cell type. For example, active promoters could be defined as having open chromatin and histone marks associated with permissive transcription (e.g. H3K4me3), while inactive promoters would be found in heterochromatin regions marked by H3K27me3. Promoters and enhancers are examples of cis-regulatory elements, genomic non-coding regions required to trigger gene transcription. The textbook classification of enhancers and promoters was challenged over the past years by several lines of evidence reporting many examples of interchangeable functions (Melamed et al., 2016). Ultimately, enhancers seem to be nothing else than less regulated promoters, able to initiate transcription, which results in abortive products in most of the cases due to lack of regulation.

1.5.2.1 Open chromatin can define genomic regulatory elements

The compaction of the chromatin is the first barrier against transcriptional activation and open chromatin regions are most of the times associated with active promoters or enhancers and provide landing sites for TFs and polymerases (Luger et al., 1997). On highly-regulated regions, such as transcription start sites (TSSs), the nucleosome occupancy is conserved and follows determined rules. Immediately downstream the TSS of transcribed genes, a nucleosome-free region flanked by a downstream (+1) and upstream (-1) nucleosomes is present. The position and chemical modifications of the +1 nucleosome strongly control transcriptional outputs. The gene body is also bound by histones, which positioning is less conserved and highly variable (Mavrich et al., 2008; Yuan et al., 2005). In general, nucleosomes preferentially contact G/C-rich DNA regions via the nucleosome centre, while A/T regions are bound by the edges (Kaplan et al., 2010).

Currently, there are two methods routinely used to map open chromatin regions at a genome-wide level, both relying on hyperactive nucleases to preferentially digest unprotected DNA. The DNase Hypersensitivity Assay (DHA) exploits the cleaving activity of the DNaseI endonuclease which has shown a preference to open chromatin sites named DNase Hypersensitivity Sites (DHSs). Compact chromatin or nucleosome-bound DNA are protected by the cleavage. When DNaseI digestion is combined with NGS (DNase-seq), an extensive map of open chromatin can be obtained for a certain sample. Consequently, DNase-seq has been used to predict TF-binding sites and cis-regulatory elements both *in-vivo* and *in-vitro* (Crawford et al., 2006; He et al., 2012). However, DNase-seq has been outcompeted by a robust method that provides additional information about nucleosome occupancy. The Assay for Transposase-Accessible Chromatin (ATAC-seq) was recently developed around the activity of the transposase Tn5 (Adey et al., 2010). ATAC-seq is a highly efficient DNA library preparation technique, applicable to low cell numbers that allows deep sequencing of accessible

chromatin genome-wide (Buenrostro et al., 2013). The use of a transposase over an endonuclease carries several advantages. While cutting, Tn5 can add at the same time sequencing adaptors. This feature was particularly important in determining histone positioning. In fact, when the histone is released, both ends of the DNA fragments are ligated with adaptors, therefore the histone-bound DNA can be tracked.

1.5.2.2 Chromatin organisation

The advent of chromosome conformation capture (3C) technologies has evidenced how the whole genome forms hierarchical and repetitive structures when folding on itself. When global interactions between genomic locations are taken into account, the genome appears distributed in two main nuclear regions: the A and the B compartments (Lamond and Earnshaw, 1998; Lieberman-Aiden et al., 2009). On a smaller scale, robust contact between promoter and enhancers, delimited by CTCF-Binding Sites (CBPs) make up the so called Topological Associated Domains (TADs) (Dixon et al., 2012).

1.5.2.3 Promoters

The promoter is the site on the DNA where the transcriptional machinery assembles and transcription starts. The canonical promoter is made of two elements: the core promoter and the proximal promoter. The core promoter is defined as the DNA region surrounding the TSS, generally including the 40 bases up and downstream this site. A promoter is generally not composed of identical elements or recurrent sequences, however several conserved elements can be classified.

The TATA-box is one of the most known elements of the core promoter. It is found throughout the animal kingdom and it is the landing site of the TATA-box binding proteins (TBPs). In the vertebrates, the TATA-box is located 30 bases upstream the TSS, while in many invertebrates is found 10 bases upstream (Lifton et al., 1978). TATA-box promoters are less represented

compared to TATA-box-lacking promoters and can be assigned to very specific classes of tissue-specific genes (Schug et al., 2005).

The element immediately downstream the TSS is the initiator (Inr). Further downstream is found the downstream of promoter element (DPE), which is the DNA region bound by TBP-associated factors, such as TAF6 and TAF9, and is located a few bases downstream from the TSS (generally 7 bases) (Burke and Kadonaga, 1996, 1997). The proximal promoter usually extends 200-250 bases upstream from the TSS, and consists of many landing sites for activators. It is often marked by open chromatin and could be involved in the interaction with distal enhancers (Figure 1.10).

The process of transcription commences when general transcription factors bind the core promoter and form the pre-initiation complex (PIC), which will ultimately recruit the RNA polymerase II. The PIC is made of six main TFs (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH), which recognise the core promoter and assemble subsequently. Normally, the TBP directly recruits TFIIA and TFIIB. Polymerase II is then brought to the DNA following complete assembly of the PIC (Thomas and Chiang, 2006).

1.5.2.4 Enhancers

The definition of enhancers comes from their intrinsic ability to enhance transcription of target gene in a distance- and orientation-independent manner. The importance of the enhancers is highlighted during embryonic development, as their activity allows a controlled gene expression in time and space (Levine, 2010). It is nowadays accepted that differential transcription in various tissues or developmental stages is mainly due to enhancers regulating transcriptional outputs.

Although the only evidence for an enhancer to be such comes from experimental validation, chromatin profiles allow prediction of candidate enhancers with fair accuracy (Tsompana and Buck, 2014). This is due to the fact that enhancer elements commonly contain unique features

such as regions marked by open chromatin that do not overlap promoters. Enhancers are found in intergenic regions as well as introns and exons. In contrast to promoters, which always tend to have nucleosome-free regions regardless the state of transcription, enhancers show accessible chromatin only when active (Heintzman et al., 2009). In addition to this, enhancers carry selective histone marks depending on whether they are active or inactive. Active enhancers are usually marked by H3K4me1 and H3K27ac, two well-known histone modifications associated with opened chromatin (Creyghton et al., 2010; Heintzman et al., 2009). They are distinguishable from promoters because enhancers lack H3K4me3. Similar to promoters, inactive enhancers organise in heterochromatin regions and overlap with H3K27me3 (Calo and Wysocka, 2013).

The various mechanisms of enhancer functions are still not completely uncovered. A generally accepted model is the enhancer-promoter interaction via looping, which brings the enhancer and the target promoter in close proximity upon 'bending' of the chromatin. In this way TFs recruited by the enhancer are shared with the promoter and boost transcription (Rippe et al., 1995).

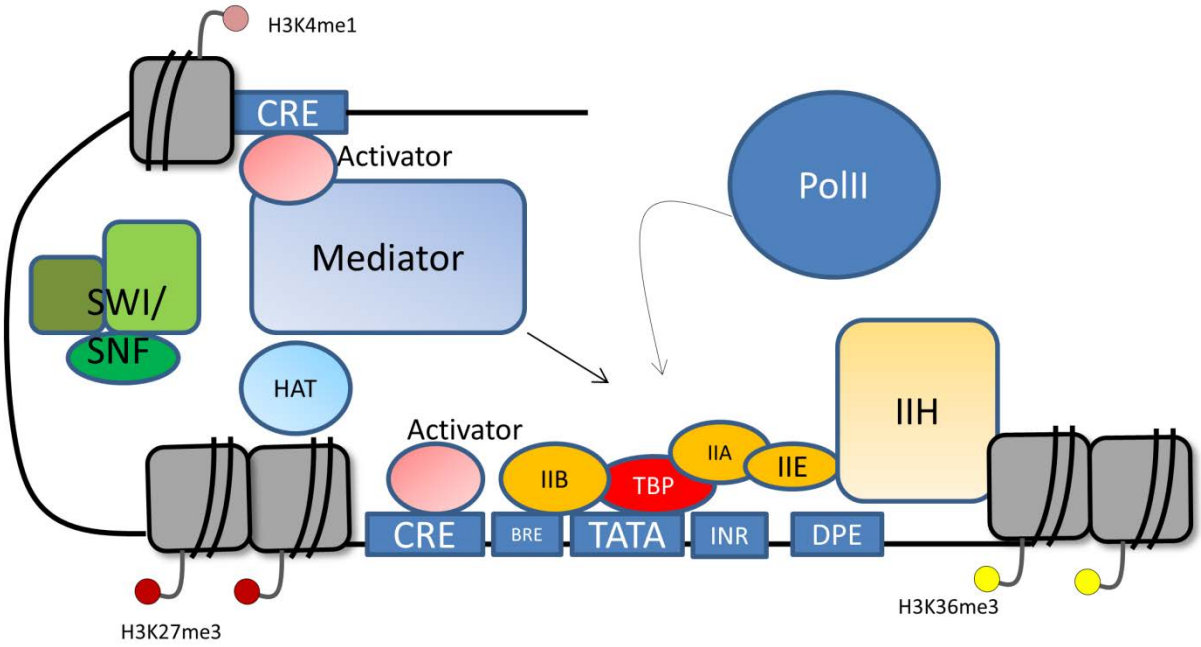

Figure 1. 10: Model of transcriptional activation through enhancer looping.
The canonical promoter sequence is made of conserved elements (blue rectangles), which provide target sites for the PIC assembly and polymerase II recruitment. The surrounding chromatin state and cis-acting elements provide further control of transcriptional regulation. Promoters are shaped by surrounding nucleosomes (in grey) presenting determined histone modifications. Inactive promoters are often marked by H3K27me3, while H3K36me3 are found along the gene body. Gene activation requires enhancer (or cis-regulatory element in general) contribution through physical contact between promoter and enhancer elements via activators and mediators. Enzymes, such HAT and protein complexes (SWI/SNF) control the state of the chromatin on various elements. This regulation allows the recruitment of the PIC, which occurs in a stepwise manner. The initial binding of the TBP to the promoter element inducing the formation of the PIC constituted of various TF (in orange). Ultimately, the RNA polymerase II is brought onto the TSS by the PIC.

1.5.2.5 Insulators and silencers

Inverse to enhancers, silencers can inhibit transcription of genes. They are found upstream of the TSS and have the ability to bind repressor proteins thus sequestering the PIC (Ogbourne and Antalis, 1998). There are two main types of silencers identified in eukaryotes: the silencer elements and the negative regulatory elements (NREs). Of the first type, the fly *zen* gene is transcriptionally inactivated in the ventral embryonic region by a silencer upstream its TSS. In certain conditions, the repressor DSP1 binds the silencer sequence in order to sequester TBP thus interfering with the PIC assembly and repressing transcription (Doyle et al., 1989).

Following this, NRE is a silencer that acts passively and in a position-dependent manner. These are identified as regions that interfere with transcription upon binding of TFs or heterochromatinization. For example, a silencer was identified upstream the human oncogene LMO2. In normal conditions, the silencer prevents transcription of LMO2 by inducing heteochromatin formation, however, the Acute Mieloid Leukemia (AML)-inducing translocation liberates LMO2 from its silencer, promoting transcription of the oncogene (Hammond et al., 2005).

Insulators are short DNA elements that were observed inhibiting physical interactions between two or more loci. Most of the insulators act through sterically inhibiting contacts mainly by recruitment of protein complexes or direct chemical modifications of the DNA. Several retro-elements act as insulators due to their tendency to DNA methylation. In fact, it was shown that 5mC-rich regions prevent interactions between flanking elements. A well-studied example is the insulator *gypsy* in *D. melanogaster*, a Long-Terminal-Repeat (LTR) retrotransposon that prevents promoter/enhancer interactions by recruitment of Hairy wing [su(Hw)] and modifier of mdg4 [mod(mdg4)] proteins (Gdula et al., 1996). Retroelements function as insulators also in the human genome.

1.5.3 DNA methylation regulates gene expression and affects chromatin packaging

DNA methylation occurs at the carbon 5 position of cytosines almost exclusively belonging to CpG islands, promoting the formation of 5-methyl-cytosine (5mC) (Kafri et al., 1992). Such modification is carried out by DNA methyltransferases (DNMT), whose activity is finely controlled in different tissues and developmental stages (Gowher and Jeltsch, 2018) in order to regulate gene expression. Currently there are no identified enzymes that de-methylate cytosines. It is believed that DNA de-methylation is achieved either through passive dilution upon replication of the DNA or direct chemical modification of the 5mC. In fact, the Ten eleven

translocation (Tet) proteins were shown to catalyse the selective oxidation of 5mC groups to 5-hydroxymethylcytosine (5hmC) (Figure 1.11).

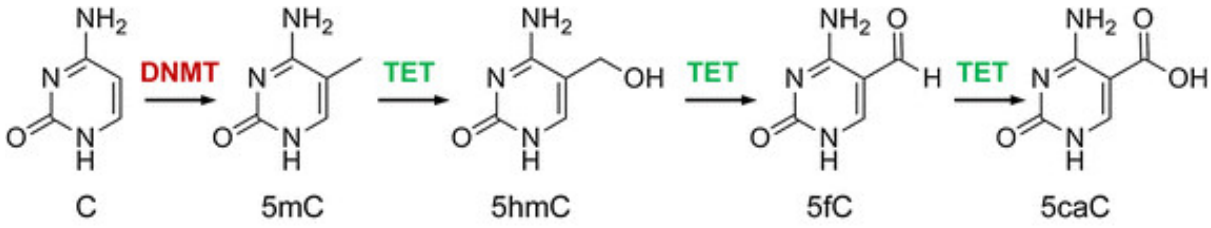

Figure 1. 11: Methylation reactions on cytosines.
DNMTs methylate cytosines (C) to 5-methylcytosine (5mC). 5-mC can be oxidised to 5-hydroxymethylacytosine by the TET enzymes. TET can also promote formation of 5-formylcytosine (5fC) or 5-carboxilcytosine (5caC). Adapted from (Breiling and Lyko, 2015).

The exact mechanism by which DNA methylation regulates gene expression is currently unclear. Several hypotheses and models propose that methylation occurring at the promoter inhibits the recruitment of the pre-initiation complex (Razin and Cedar, 1991). DNA methylation was also observed to inhibit enhancer promoter interactions by sterically avoiding their contact (Hashimshony et al., 2003). It is unclear to what extent DNA methylation affects chromatin packaging. Several lines of evidence suggest that DNA methylation indirectly regulate gene expression and DNA organisation by recruiting or interfering with protein complexes. This regulation occurs through a crosstalk between DNA methylation on CpG islands toward surrounding histone modifications and thus TF binding (Rose and Klose, 2014), which is fundamental during developmental processes and cell commitment. In fact, it was shown that some transcription-activating histone modifications (such as H3K4me3) repel DNA methyltransferases, avoiding DNA methylation on CpG islands (Cheng, 2014). Accordingly, sites occupied by the histone modifications H3K4me3 and H3K27me3 were shown to be complimentary to DNA methylation (Voigt et al., 2013). Therefore, methylation on CpG island was associated with long-term transcriptional inhibition, which compensates the short-term transcriptional control provided by histone modifications (Meissner et al., 2008). This mechanism appeared fundamental during germ line development, since it is required to keep

certain genes either actively transcribed or shortly silenced (Lesch and Page, 2014). Nevertheless, the plasticity of the germ line epigenome favours short-term epigenetic control on dynamically regulated genes, while DNA methylation is found on terminally inactivated regions (Voigt et al., 2013).

1.6 Epigenome of PGCs in mammals and epigenetic reprogramming

Like every cell type, the PGCs also undergo lineage determination during embryogenesis through specialised epigenetic reprogramming. Several recent works carried out mainly on human, mouse and *C. elegans* showed that PGCs acquire a specific chromatin landscape coherent with expression of pluripotency-associated genes and repression of somatic fate (Bender et al., 2004; Gkountela et al., 2015; Guo et al., 2015). It is still not known whether the epigenetic reprogramming of the germ line during embryogenesis is conserved among the animal kingdom. In mammals, the most dramatic epigenetic reprogramming is reported during PGC migration, indicating that extensive epigenetic regulation is necessary during the early phases of PGC formation. Interestingly, only a few genes are differentially regulated between pre-migratory PGCs and somatic cells in mouse and *D. melanogaster*, suggesting that the onset of migration is post-transcriptionally regulated (Saitou et al., 2002; Siddiqui et al., 2012).

1.6.1 Global DNA demethylation occurs in mammalian PGCs

Analyses of mammalian DNA methylomes have reported that a general demethylation occurs in PGCs during early embryogenesis in both mouse and human (Gkountela et al., 2015b; Guo et al., 2015). The causes of the general DNA demethylation observed in human and mouse embryos are still unclear.

Interestingly, a global erasure of parental epigenetic memory takes place in the early embryo, which subsequently re-builds the methylome via a zygotic programme. As establishment of DNA methylation profile is important during normal mammalian development (Bird, 2002; Smith and Meissner, 2013), complete erasure of 5mC will avoid to transmit deleterious methylation patterns in the next generation. Importantly, this argues against the tendency of passing epigenetic information from the germ line to the next generation, which is instead preferentially ‘rewritten’. However, a small number of single copy genes and certain

transposable elements showed retention of parental DNA methylation patterns, indicating these sites as potential evidence of epigenetic inheritance (von Meyenn and Reik, 2015).

Interestingly, in human and mouse the germ line undergoes a second wave of DNA demethylation later during development, which brings the total DNA methylation levels below 10% of the CpG islands. This event is observed between the 4th and the 7th week of gestation in human and around E10.5 in mouse, when PGCs have just formed (Gkountela et al., 2015). At this stage, DNA demethylation might be necessary to induce a diverse developmental fate from the somatic line or simply to wipe out any parental germ line memory in order to protect the newly formed germ cells. One possible explanation to this hypothesis would be that such high energy-demanding process cannot be supported throughout the embryo, therefore it evolved in the germ line only. Concomitant transcriptomic and methylome analyses underlined that human PGCs undergo general demethylation at the most of TSSs, gene promoters and enhancers (Gkountela et al., 2015; Guo et al., 2015), however this does not correlate with changes in gene expression. These observations are in accord with the hypothesis that DNA demethylation may be required for pure erasure of epigenetic memory and for generating a new methylome (von Meyenn and Reik, 2015). It follows that alternative mechanisms of gene expression may exist in order to maintain a regulated epigenome in PGCs whilst undergoing intense DNA demethylation.

The mechanisms of DNA demethylation in PGCs are currently under investigation. When global levels of 5mc and 5hmC were quantified in the mouse PGCs, a significant decrease in both was observed, indicating that 5mC lost is not caused by 5mC to 5hmC conversion. Moreover, depletion of TET1 in mouse PGCs did not affect the global loss of 5mC, further supporting the observation that DNA demethylation requires cytosine oxidation. Interestingly, it was observed that targeted 5hmC on a subset of germ line-specific genes is required during the transition from PGCs to gonocytes and to support gametogenesis (Hill et al., 2018).

1.6.2 Chromatin reprogramming of PGCs

Because DNA methylation does not correlate with transcriptional activation in PGCs, it is possible that alternative epigenetic mechanisms and regulatory events could be involved in transcriptional repression (Gkountela et al., 2015). The most obvious candidates for regulating gene transcription upon DNA demethylation are histone modifications. Murine gonadal PGCs acquire specific histone modifications alongside DNA demethylation. Between E10.5 and E12.5, PGCs exhibit low levels of histone H1 and the chromatin de-condensates. Relaxing of the chromatin is also evidenced by loss of H3K27me3 and H3K9me3, although a reduction in the histone mark H3K9ac was observed alongside (Hajkova et al., 2008). Interestingly, mouse PGCs are enriched in H3K27me3 on somatic genes and depleted in H3K9me2 (Ng et al., 2013; Seki et al., 2005). Therefore, PGCs may skip somatic development by accumulating repressive histone modifications on somatic genes. On the other hand, in human, PGCs have low levels of both H3K27me3 and H3K9me2, suggesting the presence of a different repressive histone mark, identified as H3K9me3 (von Meyenn and Reik, 2015).

Recently, the Polycomb Repressive Complex (PRC1) was associated with gonadal de-repression of germ line-specific genes in cooperation with TET1 (Hill et al., 2018). The PRC1 catalyses the ubiquitylation of histone H2A on lysine 119 (H2AK119U), promoting chromatin compaction and gene downregulation (Francis et al., 2004). When PRC1 is pharmacologically inhibited in mouse ESCs, a significant increase in germ line-specific genes expression is detected, indicating that PRC1 is involved in repression of germ line genes. Moreover, loss of TET1-mediated DNA demethylation led to significant decrease in the levels of the same gene subset. Taken together, these results point towards a model where TET1-mediated oxidation of 5mC and chromatin relaxation promote timed expression of germ line genes during murine PGC development (Hill et al., 2018).

1.7 Project aims

1.7.1 General motivations

Currently, most of the work carried out on zebrafish PGC development relies on imaging. In fact, the discovery of several tagging methods to track the PGCs, in combination with the transparency and extrauterine development of a zebrafish embryo, have allowed to study movements and dynamics of the embryonic germ line. However, the limited number of PGCs have impeded so far in-depth investigation of molecular mechanisms required to drive PGC formation. Moreover, little is known about the epigenetic contribution to PGC development in zebrafish, although its understanding may provide insights into the combinatorial activity of germ granules and epigenetic reprogramming. Along this line, recent works highlighted how the mammalian PGCs establish a germ cell-specific epigenetic profile via global DNA demethylation (Gkountela et al., 2015; Guo et al., 2015). Although, it is known that timed expression of germ cell genes requires DNA demethylation, this phenomenon is not restricted to germ cell genes only (Hajkova et al., 2008; Hill et al., 2018). Therefore, alternative hypotheses have suggested that the global DNA demethylation may reset the epigenome of the mammalian PGCs (Guo et al., 2015; Kobayashi et al., 2017), facilitating the branching of the germ line from the somatic tissues and preserving the pluripotent state.

Interestingly, zebrafish PGCs do not undergo DNA demethylation during the migratory period (Skvortsova et al., 2019), hinting for alternative scenarios, including lack of a PGC-specific early epigenetic reprogramming, which may be replaced by intrinsic germ plasm functions. However, while the role of the germ plasm was mainly indicated as cytoplasmic and involved in post-transcriptional regulation, the existence of germ factors-mediated chromatin regulation is not excluded. Moreover, it is well known that loss of individual germ plasm factors leads to PGC death, migration defects or somatic differentiation (Gross-Thebing et al., 2017; Köprunner et al., 2001a; Weidinger et al., 2003). While the mechanisms of PGC degeneration

are not known, this suggests that the germ plasm is dispensable for germ fate gain. Therefore, it is likely that germ factors are required for activating downstream processes including acquisition of epigenetic identity.

My principal goal was to investigate germ plasm-driven PGC formation on a transcriptional and epigenetic level. For this purpose, I performed NGS- and imaging-based characterisation of PGC development in zebrafish during the first day of zebrafish embryogenesis, when the PGCs show the most dynamic behaviour. Recent advances in sample handling and genomic library preparation, as well as increasing robustness of high throughput sequencing have allowed the generation of high-quality genomic data from small cell population. For these reasons, we have selected the most advanced techniques to efficiently isolate and sequence small amount of nucleic acids in combination with optimisation of the isolation of GFP-tagged PGCs.

1.7.2 Specific Objectives

The global aim of this project was to investigate the molecular events required to specify the embryonic germ line in zebrafish. Primarily, this thesis focuses on understanding the link between maternally-provided cytoplasmic molecules and embryonic transcriptional regulation during the early phases of PGC formation. Moreover, we aimed to study the functional role of the germ plasm during various phases of germ cell development.

The objectives of this work were:

- To determine when and how the germ fate is acquired during the early stages of zebrafish embryogenesis and what is the role of the germ plasm. Lines of evidence in other organisms suggest that germ plasm factors affect the zygotic genome activation

in the PGCs (Batchelder et al., 1999; Hanyu-Nakamura et al., 2008b; Mello et al., 1996), therefore I focused on the very beginning of zebrafish development.

For this purpose, I exploited single-cell resolution microscopy and NGS assays to compare the somatic and germ transcriptomes and epigenomes. I investigated the various steps needed for PGC formation by focusing on timepoints selected based on germ plasm features, indicative of global molecular events. By this, I aimed to identify key factors and mechanisms involved in germ cell specification.

- To verify if the germ line of a germ plasm-dependent organism undergoes epigenetic reprogramming during PGC migration. This phenomenon was linked to global DNA demethylation in mammals, where no germ plasm is detected (Gkountela et al., 2015; Guo et al., 2015; Kobayashi et al., 2017). Moreover, DNA demethylation seems not essential for PGC specification in zebrafish (Skvortsova et al., 2019), raising questions about cooperation or complementarity between germ plasm and epigenetic control.

  Accordingly, the global epigenome of zebrafish PGCs was profiled at various stages of development by looking at the open chromatin landscape. I explored potential roles of cis-acting elements such as promoters and enhancers. Alongside, transcriptomic analysis was performed in order to functionally explain epigenetic differences between PGCs and somatic cells.

- To link the dynamic subcellular localisation of the germ plasm with epigenetic and transcriptional phenotypes. In fact, although remarkable germ plasm movements have been reported during the early embryogenesis of several organisms (Knaut et al., 2000; Köprunner et al., 2001; Strasser et al., 2008; Yoon et al., 1997), it remains unclear how germ plasm localisation affects PGC development. Pioneer studies have found that germ granules are able to extend the nuclear pore for RNA processing (Pitt et al., 2000), however the effect on transcriptional regulation has not been studied yet. In order to

better understand the role of the germ plasm at various developmental stages, time series analysis of the PGCs transcription and open chromatin profile were performed combining imaging and NGS tools.

In summary, the specific aim of this project was to globally characterise the development of the PGCs in a germ plasm-dependent vertebrate, representing a biological system in where transcriptional and post-transcriptional regulation may play a cooperative role. The presented results support novel scenarios proving direct involvement of the germ plasm in epigenetic reprogramming of the germ line.

2 Materials and Methods

2.1 Materials

2.1.1 Fish strains

- Tg(Buc-GFP) fish line
- Wild type (AB) fish line

2.1.2 Antibiotics

Name	Catalogue number	Source
Ampicillin sodium salt	A9518	Sigma-Aldrich, UK

2.1.3 Antibodies

Name	Catalogue number	Source
H3K4me3	C15410030	Diagenode, Belgium
H3K27me3	C15410195	Diagenode, Belgium

2.1.4 Bacterial strains

Name	Catalogue number	Source
Alpha-selected competent cells	Bio-85026	Bioline, UK

2.1.5 Buffers and solutions

Luria-Bertani (LB) broth: 20g LB broth powder (Sigma-Aldrich, L3022, UK)/1 litre of H2O.

4% Paraformaldehyde (PFA): PFA (Alfa Aesar, 43368, USA) was dissolved in sterile PBS at 60°C with constant stirring. pH was adjusted to 7.4.

FACS Dissociation Buffer: 0.25% BSA (Sigma-Aldrich, A2058, UK), 100 mM Hepes (Sigma-Aldrich, H3375, UK) in Hank's Balanced Salt Solution (HBSS, Sigma-Aldrich, 14025092, UK), pH 7.4.

E3 Medium: 0.5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl2, 0.33 mM MgSO4, pH 7.2.

PBST 1X: PBS, 0.1% Tween-20 (Sigma-Aldrich, P1379, UK), pH 7.4.

FISH 20X saline sodium citrate (SSC) Buffer: NaCl (3M); Sodium citrate (0.3M).

FISH Blocking Solution: 1x PBST, 2% sheep serum (vol/vol) (Sigma-Aldrich, S3772, UK), 2 mg/ml BSA.

FISH Hybridization Buffer (+): 50% deionized formamide, 5X SSC, 0.1% Tween-20, 500 µg/ml of extracted RNase-free tRNA (Thermo Scientific, AM719, UK) adjusted to pH 6.0.

FISH Hybridization Buffer (-): 50% deionized formamide, 5X SSC, 0.1% Tween-20.

FISH Staining Buffer: 1 M Tris-HCl, adjusted to pH9.5, 50 mM MgCl2, 100 mM NaCl, 0.1% Tween-20.

2.1.6 Chemical reagents

All chemical used during this work were quality grade and aimed for research purposes only. Routinely used chemicals are not listed.

Name	**Catalogue number**	**Source**
Bovine Serum Albumin	A3059	Sigma- Aldrich, UK
dNTPs	R0192	Thermo Scientific, UK
Nuclease-free Water	AM9937	Ambion, UK
Paraformaldehyde	43368	Alfa Aesar, USA
Phenol Red	P0290	Sigma-Aldrich, UK
Glycogen	10901393	Roche, UK

Triptolide	T3652	Sigma-Aldrich, UK
Hepes	H3375	Sigma-Aldrich, UK

2.1.7 Enzymes

All enzymes were purchased from New England Biolabs (NEB) unless stated otherwise.

2.1.8 Kits

Name	**Catalogue number**	**Source**
Superscript IV	18090050	Thermo Scientific, UK
RNeasy micro Kit	74044	Qiagen, UK
Plasmid Spin MiniPrep	27104	Qiagen, UK
mMESSAGE mMACHINE	AM1340	Thermo Scientific, UK
QIAquick PCR purification	28104	Qiagen, UK
SMART-Seq® v4	634889	Takara Bio Europe, France
Nextera library preparation kit	15028212	Illumina, UK
MinElute PCR purification	28004	Qiagen, UK
DIG wash and block buffer set	11585762001	Roche, UK
TSA Plus Cyanine 3 System	NEL744001KT	Perkin Elmer, UK

2.1.9 Oligonucleotides

All the oligonucleotides were designed using the online resource Primer3 and tested with the In-Silico PCR tool provided by the UCSC genome browser. All oligonucleotides were purchased from Sigma-Aldrich in lyophilised form and resuspended in ddH2O on their arrival. The underlined sequence in Ad2 reverse provides the barcode and was replaced with alternative octamers for each sample in order to allow sample pooling.

Name	Forward sequence	Reverse sequence
Dnd1	TTCACTCTTCATGGCTCGTG	GTCAACAGACTCGGCTCTCC
Nanos3	AGACTGAGGCCGTGTACACCTCTCACTACT	GAGCAGTAGTTCTTGTCCACCATCG
Myod1	TCCGTCTTCTCGTCTGACAC	AGTCCGAGATCCAACTGCTC
Nanog	TCACAGCGGGCTACTTTACC	TGGTCTGCTCTGCATCTTTG
Ddx4	AGGATCCTTCAAGAGCGATGA	GGTATTGAAGAAGCTCGCACA
Myl12.1	CCAAGGTAAAGCTGCACTGT	CCACGAGAGCCCTGAACTTA
Nanos3	AGACTGAGGCCGTGTACACCTCTCACTACT	GAGCAGTAGTTCTTGTCCACCATCG
Tdrd7	TCTACCCAGCGGAAGCTTTA	CTGGTGTCCCACTGGTCTTT
Tdrd9	GGTCTCCGATCCGTAATCAG	AGCCTCCATCTCATCAAAGC
Dnd1	TTCACTCTTCATGGCTCGTG	GTCAACAGACTCGGCTCTCC
Bact	TCTCTCTGTTGGCTTTGGGA	CCTGACCCTCAAATACCCCA
ATAC adaptor1	AATGATACGGCGACCACCGAGATCTA CACGCCTCCCTCGCGCCATCAG	-
ATAC adaptor2	CAAGCAGAAGACGGCATACGAGATCG GTCTGCCTTGCCAGCCCGCTCAG	-
Ad1 Forward	AATGATACGGCGACCACCGAGATCTA CACTCGTCGGCAGCGTCAGATGTG	-
Ad2 Reverse	CAAGCAGAAGACGGCATACGAGATTC GCCTTAGTCTCGTGGGCTCGGAGATGT	-

2.1.10 Morpholinos

All morpholinos were purchased from GeneTools, LLC, USA. The underlined nucleotides show sites of mismatch with the target RNA.

Name	Target Sequence	5-Mismatches target
miR-430 MO	CACACGCATCTTGTTGTCTGCTGTT	CCCACTCATATTGTTGTATGCTTTT
Tdrd7 MO	AACCAACTCCACGTCACTCATCCTG	ACCCAACTGCACGCCACTAATACTG

2.1.11 Equipment

Name	Catalogue number	Source
Stereoscopic Zoom microscope	SMZ1500	Nikon, UK
Fluorescent Zoom microscope	Axio Zoom.V16	Zeiss, Germany
Confocal	LMS780	Zeiss, Germany
Lightsheet microscope	Z.1	Zeiss, Germany
ScanR Screening Station	Scan^{r}	Olympus, Germnay
NanoDrop ND-Spectophotometer	ND-1000	NanoDrop Technologies, UK
QuBit	3.0 Fluorometer	Thermo Fisher Scientific, UK
Bioanalyzer	2100	Agilent technologies, UK
Tapestation	4200	Agilent technologies, UK
Tabletop microcentrifuge	5415D	Eppendorf, Germany
Refrigerating centrifuge	1-14K	Sigma-Aldrich, UK
Thermocycler	SensoQuest	GeneFlow, UK
Thermocycler	PRISM 7900HT	Applied Biosystems. UK
Transilluminator	Gene Flash	Syngene, UK
Sequencing Platform	Nextseq550	Illumina, UK

2.1.12 Plasmids

Name	Catalogue number	Source
PGEM-Teasy	A1360	Promega, UK

2.2 Methods

2.2.1 Molecular biology methods

2.2.1.1 Phenol-chloroform DNA/RNA purification

Nucleic acids were phase-separated from cellular extract in phenol-chloroform gradients. An equal volume (1:1) of basic phenol-chloroform (pH 8) for DNA or acidic phenol-chloroform (pH 6.5) for RNA was added to the cellular extract. After mixing, the solution was spun down at 12000 x g for 6 minutes or until the phase separation was achieved. The aqueous (top) phase was transferred to a new tube and an equal volume of chloroform was added. After mixing and spinning down, 2 volumes of isopropanol and 1/10 of the volume of Ammonium Acetate 7.5 M (Sigma-Aldrich, A2706, UK) were added to the aqueous phase. DNA/RNA were then chilled over night at -20°C and precipitated in a cold centrifuge for 30 minutes at maximum speed. The pellet was washed once with 70% ethanol and resuspended in the desired amount of water.

2.2.1.2 Restriction digestion

Restriction digestion was used prior to ligation on both vectors and inserts. Restriction digestion was performed by restriction enzymes purchased from New England Biolabs (NEB) usually in a 50 µl volume reaction with appropriate buffer and enzyme units (see below), according to manufacturer's instructions.

2.2.1.3 DNA ligation

Ligation of DNA fragments was achieved by mixing of vectors and inserts in a proportion of 1:3 in presence of T4-ligase (NEB, M0202T, UK). The reaction was incubated either at room temperature for 1 hour or at 16°C overnight for maximum efficiency.

2.2.1.4 Bacterial transformation

A ratio of 450 ng of plasmid (pGEM-Teasy, A1360, Promega) per 200 µl of *E. coli* competent cells (DH5-Alpha cells) was used for transformation. Cells were left in the presence of the plasmid for 30 minutes on ice and intake of plasmid was helped by heat shock, occurred in a 42°C water bath for 40 seconds. After addition of 1 ml of LB medium, the cells were grown at 37°C for 40-60 minutes with vigorous shaking.

Cells were harvested by gentle centrifugation for 5 minutes (1000 x g) and 800 µl of supernatant was discarded. Cells were resuspended in 200 µl of LB medium, streaked on Ampicillin agar plates and grown overnight at 37°C.

2.2.1.5 Generation of bacterial liquid cultures

Single colonies were selected and manually picked by a tip and placed in 4 ml of LB medium supplemented with 100 µg/µl Ampicillin. Liquid colonies were grown overnight at 37°C in a non-sealed 15 mL Falcon tubes with agitation.

2.2.1.6 Plasmid Midiprep

Single colonies were grown in 200 ml of LB supplemented with 100 µg /ml ampicillin over night at 37°C and gentle shaking (200 rounds per minute, rpm). Cells were harvested by centrifugation at 6000 x g for 10 minutes at 4°C and the supernatant was completely discarded. Pelleted cells were resuspended in 250 µl of Buffer P1 and 250 µl of Buffer P2 were added (Qiagen, 27104, UK). After 5 minutes of incubation on ice, cells were lysed by addition of 250 µl of Buffer P3 (Qiagen, 27104, UK) and incubated on ice for 15 minutes at 17900 x g. Proteins and genomic DNA were precipitated by centrifugation at 4°C for 15 minutes. Plasmid DNA was bound to the membrane-fitted column (Qiagen, 27104, UK) while the supernatant went through the membrane by gravity. The column was washed two times in Buffer QRT and

plasmid eluted in 5 ml of Buffer QF (Qiagen, 27104, UK). The DNA was subsequently precipitated in isopropanol as described in 2.2.1.1 and resuspended in 200 µl of water.

2.2.1.7 Plasmid linearization for in-vitro transcription

Plasmid linearization occurred in the following digestion reaction:

Component	µl in a 200 µl reaction	Final Concentration
NotI enzyme (10u/µl)	4	0.2 units
10X Reaction Buffer	20	1X
Plasmid (500ng/ µl)	20	50 ng/µl
Nuclease-Free Water	To 200	-

Linearized plasmid was purified using phenol-chloroform extraction, precipitated with 10% of 7.5M of Sodium acetate and ethanol and re-suspended in nuclease-free water as in 2.2.1.1.

2.2.1.8 Gel electrophoresis

Gel electrophoresis was used in order to determine nucleic acid quality and size. 1 to 2 % agarose gels were prepared by melting agarose powder in 1X TAE Buffer (40 mM Tris acetate, 1 mM EDTA, pH 8.3). Liquid agarose was supplemented with 0.5 µg/mL of Ethidium Bromide (Sigma-Aldrich, E1510, UK) and cooled down in a suitable tray. Electrophoretic migration of DNA occurred in 1X TAE Buffer by applying 90-150 Volt current until proper band separation was achieved.

2.2.1.9 Synthesis of capped mRNA by in-vitro transcription

In-vitro synthesis of capped mRNA was achieved by usage of mMESSAGE mMACHINE® Kit. A linearized plasmid with either a SP6 or T7 promoter was used as template.

The transcription reaction was assembled according to the protocol as follows:

Component	µl in a 20 µl reaction	Final Concentration
2X NTP/CAP	10	1X
10X Reaction Buffer	2	1X
Linearized plasmid (200ng/ µl)	5	50 ng/µl
Nuclease-Free Water	To 20	-

The transcription reaction was incubated 2 hours at 37°C, followed by digestion of the DNA template through addition of 1 µl of Turbo DNase and incubation for 15 minutes at 37°C. Synthetized mRNA was purified by phenol-chloroform extraction (described in 2.2.1.10) and resuspended in 20 µl of nucleases-free-water.

RNA concentration was measured by spectrophotometry and RNA size/purity by gel electrophoresis.

2.2.1.10 RNA purification

In-vitro synthetized mRNA was retrieved by Phenol-Chloroform extraction. 130 µl of Stopping solution (15 µl of Ammonium Acetate + 115 µl of water) were added to the transcription mix. One volume of Acidic Phenol Chloroform was added and the sample was mixed thoroughly for several seconds. The two phases were separated by 7 minutes of centrifugation at 11000 x g and the aqueous phase was taken and transferred into a new tube. One volume of Chloroform was added, mixed and centrifuged as in the previous step. Aqueous phase was retrieved and RNA was precipitated in one volume of isopropanol at – 20°C overnight. Precipitated RNA was pelleted by centrifugation at 15000 x g and resuspended in the required volume of nuclease-free-water.

2.2.1.11 Genomic DNA extraction from fish embryos

Pre-5dpf embryos were collected in Eppendorf tubes and dissolved into 50 mM NaOH by 10 minutes incubation at 95°C. The basic solution was neutralised by adding 1:10 of Tris-HCl (1M) to the sample and debris were removed by centrifugation. The supernatant was used as source of genomic DNA and kept for up to three months at 4°C.

2.2.1.12 Polymerase Chain Reaction (PCR)

DNA amplification was performed with GoTaq Polymerase (Promega, 9PIM300, UK). The following reagents were pipetted into a 0.25 ml PCR reaction tube (Thermo Scientific, AB0451, UK).

Components	25 µl reaction	50 µl reaction	Final concentration
5X GoTaq Reaction Buffer	5	10	1X
dNTPs Mix (10 mM)	0.5	1	0.2 each
10 µM Fwd Primer	1	2	0.4 µM
10 µM Rev Primer	1	2	0.4 µM
GoTaq DNA Polymerase (5u/µl)	0.25	0.5	0.5 U/µl
Template DNA	-	-	-
Nuclease-free water	To 25 µl	To 50 µl	-

Template DNA	Amount
Genomic	50 ng-250 ng
Plasmid	1 pg-10 ng

Thermocycling conditions for a routine PCR:

STEP	TEMPERATURE (°C)	TIME	CYCLES
Initial denaturation	95-98	2 minutes	1
Denaturation	95-98	30 seconds	30-35
Annealing	60-65	30-45 seconds	30-35
Extension	72	30-45 seconds	30-35
Final extension	72	5 minutes	1

Extension time was calculated based on amplicon length as 45 seconds per 1kb.

PCR products were checked for amplification and size on 1% Agarose gel and the rest of product was purified using QIAGEN Gel Extraction kit (Qiagen, 28104, UK) and/or QIAGEN MinElute kit (Qiagen, 28004, UK) according to the manufacturer's instruction.

2.2.1.13 Real Time PCR

DNA amplification was performed with the PowerUp SYBER Mix (Thermo Scientific, 4368577, UK) and monitored in a real-time PCR machine as follows.

The following reagents were pipetted into a 384-well-plate:

Component	µl in a 10 µl reaction	Final Concentration
SYBER Green Enzyme Mix	5	1X
Forward PCR Primer (10mM)	0.5	500 µM
Reverse PCR Primer (10mM)	0.5	500 µM
DNA template	-	1 ng/µl
Nuclease-Free Water	To 10	-

Thermocycling conditions for a routine real time-PCR:

STEP	TEMPERATURE (°C)	TIME	CYCLES
Initial denaturation	95-98	10 minutes	1
denaturation	95-98	20 seconds	20-40
Annealing	60-65	15-25 seconds	20-40
Extension	72	15-25 seconds	20-40

For data analysis, cycle threshold (ct) values were used. In standard real time PCR analysis, the cycle number at which the DNA intensity reaches a pre-set threshold (ct value) is proportional to the initial amount of DNA and allows quantification of input DNA relative to a reference. Relative quantification was achieved via an extensively used double-normalisation method, which calculates the delta-delta ct as summarised:

Ct gene of interest control (CtGNcontrol)

Ct housekeeping gene control (CtHKcontrol)

Ct gene of interest treatment (CtGNtreatment)

Ct housekeeping gene treatment (CtHKtreatment)

Delta-delta Ct = [(CtHKcontrol - CtGNcontrol) – (CtHKtreatment - CtGNtreatment)]

2.2.1.14 DNA purification and size-selection

The required volume of AMpure beads (1.2X) (Beckman Coulter, A63880, USA) was added to the DNA to be purified and mixed by pipetting. The solution was incubated 10 minutes at room temperature and then beads were captured on a magnetic stand for 5 minutes. The supernatant was discarded and the beads with DNA bound were washed twice in freshly made 80% ethanol. In order to let all the ethanol dry, tubes were left 10 minutes with the lid open at room temperature. DNA was eluted in the desired volume of EB (Qiagen, 28104, UK) for 15 minutes at room temperature.

2.2.1.15 Sequencing

Plasmids or PCR fragments were sequenced using an external service when high sequence fidelity was needed. Either 1 µg of plasmid or 200 ng of PCR reaction were sent alongside 10 µl of the sequencing primer (5 mM). All Sanger sequencing was performed by SourceBio, UK.

2.2.1.16 Quantification of nucleic acids

DNA or RNA were quantified using a valid light absorbance-based approach. For routing quantification and purity validation, a Nanodrop ND 1000 spectrophotometer (Thermo Scientific, UK) was used which determines the concentration of a certain nucleic acid with the Lambert-Beer's Law (A = e * L*c), where A is the absorbance, e is the molar extinction coefficient, L is the path length that the light crosses and c is the concentration.

The Nanodrop read also provides measurement of the nucleic acid purity/contamination. This is given by the ratio 260/280λ and 260/230λ. A 260/280λ ratio of 1.8 and 2.0 were considered acceptable for DNA and RNA respectively, while a 260/230λ ratio ranging between -2.2 and -1.8 was taken as purity validation.

2.2.2 PGCs isolation via FACS

2.2.2.1 Preparation of single cell suspension from transgenic embryos

PGCs were isolated at different stages via FACS. Transgenic (Tg) Tg(Buc-GFP) heterozygous embryos were grown at 28.5°C by incubation in E3 medium supplemented with 1mg/ml gentamicin. Embryos were washed three times in sterile water before collection and about two hundred of them were pooled in single eppendorf tubes. 500 µl of Hank's balanced salt solution (HBSS) supplemented with 0.25% BSA and 10mM Hepes (Sigma-Aldrich, H3375, UK) were added and dissociation occurred by pipetting for 2 minutes with a glass pasteur pipette. Excess of yolk was removed by two rounds of 3 minutes centrifugation at 250 x g at 4°C, while pelleted

cells were resuspended in 1ml HBSS supplemented with 0.25% BSA and 10mM Hepes. The cell suspension was kept on ice for the entire isolation procedure.

Post-blastula PGCs were collected from the Tg(Buc-GFP) line. Embryos were washed as described before with extra pipetting. Cells were pelleted by 4 minutes centrifugation at 250 x g, resuspended in 1 ml of HBSS supplemented with 0.25% BSA and 10mM Hepes and filtered through a 50 μm mesh before being taken to the sorter.

2.2.2.2 FACS procedure

The cells were sorted on a BD FACS Aria IITM flow cytometer, based on size, granularity and fluorescence. In the flow cytometer, cells were passed through a nozzle which controlled pressure and sorting speed. For early embryos, a 100 microns nozzle was used, while for post-gastrulation embryos the nozzle diameter was 70 microns. Nozzle vibrations are controlled by frequency and amplitude parameters (Table 2.3).

Cell shape and size were used to discriminate alive cells from debris or damaged cells and were shown as forward scatter (FSC) and side scatter (SSC) respectively. FSC is the measure of the light diffracted by the specimen in the forward direction, while the SSC is detecting light diffracted at 90° from the specimen.

As a cell is hit by the laser beam, the light is converted to electrical signal and a 'pulse' of emission is detected and measured over time. The pulse is a prompt increase of emission detected over a baseline state. The level of the baseline state is controlled by the voltage (Table 2.3).

Setting	Nozzle (70 microns)	Nozzle (100 microns)
Pressure (psi)	70	20
Amplitude	60	12
Frequency	90	30
Voltage	120	90

In order to minimize the carry-over of GFP-negative cells attached to GFP-positive cells, existence of doublets was checked with standard procedures. In brief, the area versus the height of the pulse generated by the FSC were plotted. With these settings, events distributing on the diagonal represent single cells, while doublets normally appear on the top of the graph as the pulse height is higher in proportion to the pulse area.

After the described filtering, a diode laser was used to excite the specimen in order to identify and separate GFP-positive cells (PGCs). The use of two lasers allowed for better discrimination between true GFP-positive and auto-fluorescent cells. The detected emission light was for the PE and FITC fluorophores.

2.2.2.3 Purity check of sorted cells

Single cell suspension was prepared as described from Tg(Buc-GFP)$^{+/-}$ embryos. GFP-positive cells were collected by FACS in 20 µl of PBS supplemented with 0.25% BSA and 100mM Hepes and immediately placed on a glass slide with coverslip for imaging.

2.2.3 DNA library preparation for NGS

2.2.3.1 Assay for Transposase Accessible Chromatin (ATAC)

Open chromatin regions were detected genome-wide by ATAC-seq as described in Buenrostro et al., 2013.

About 5000 cells were collected in supplemented HBSS. Nuclei were extracted by pipetting cells up and down in lysis buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630, pH 7.4) and pelleted in a cold centrifuge for 10 minutes at 500 x g.

The transposition reaction was assembled as follows:

Component	µl in a 10 µl reaction	Final Concentration
TD Buffer (2X)	25	1 X
Tn5 Transposase	5	1 U/µl
Nuclease-Free Water	To 50	-

The Tn5 transposase and the 2x TD Buffer were part of the Nextera™ DNA Sample Prep Kit (Illumina, FC-121-1030, UK).

The reaction was incubated at 37°C for 30 minutes and the DNA was extracted by a QIAGEN MinElute Clean up Kit. Transposed DNA underwent four cycles of PCR amplification using the NEBNext High-Fidelity 2X PCR Master Mix (NEB, M0541S, UK) by Customized Nextera PCR Primer forward and reverse (Table 2.1) (Buenrostro et al., 2013). In order to calculate the exact number of cycles needed to reduce CpG bias, amplification reaction was monitored by real-time PCR and then completed with required standard PCR cycles.

Primer dimers and fragments shorter than 150 bp were removed by 1.2X AMPure beads purification as described above. Fragments size and molarity were detected by Agilent 2100 Bionalyzer. Indexed fragments were checked in concentration by qPCR, equimolarly pooled and sequenced on an Illumina Next-Seq 550.

2.2.3.2 RNA-Seq

cDNA was prepared according to the manufacturer's (Takara Bio Europe, 634889, France) instructions as follows. Exactly 200 cells were sorted as described in 2.2.2.2 in 8.5 µl of water (0.2U/µl RNase inhibitor). Immediately after collection, 1µl of SMART-Seq v4 lysis buffer

(0.2U/µl RNase inhibitor) and 1 µl of ERCC Mix1-2 (final dilution $1x10^{-6}$) were added to the cells prior to freezing (Thermo Scientific, 4456739, UK).

Frozen cells were thawed on ice for five minutes and 2 µl of 3' SMART-Seq CDS Primer II A were added to each sample. The reaction was pre-heated at 72°C for 3 minutes and the following mix was added:

Component	µl per reaction	Final Concentration
5X Ultra Low First-Strand Buffer	4	1.875 X
SMART-Seq v4 Oligonucleotide (48 µM)	1	3.2 µM
RNase Inhibitor (40 U/µl)	0.5	2.6 U/µl
S.Scribe Reverse Transcriptase	2	NA
Total	7.5	-

Reverse transcription occurred in a thermocycler set as follows:

STEP	TEMPERATURE (°C)	TIME	CYCLES
Reverse transcription	42	90 minutes	1
Inactivation	70	10 minutes	1
Hold	4	forever	NA

After reverse transcription, cDNA was amplified by addition of the following reaction:

Component	µl per reaction	Final Concentration
2X SeqAmp PCR Buffer	25	1.2 X
PCR Primer II A (12 µM)	1	0.4 µM
SeqAmp DNA Polymerase	1	NA
Nuclease-Free water	3	NA

Total	30	-

The cDNA was amplified in a thermocycler as manufacturer's recommendations:

STEP	TEMPERATURE (°C)	TIME	CYCLES
Initial denaturation	95	1 minutes	1
denaturation	98	10 seconds	12
Annealing	65	30 seconds	12
Extension	68	3 minutes	12
Final extension	72	10 minutes	1
Hold	4	forever	-

2.2.4 Imaging methods

2.2.4.1 Immunostaining

Pre-dechorionated zebrafish embryos were collected in Eppendorf tubes at the desired developmental stage and set on ice. All the liquid was removed and fixation occurred in 3.5% PFA (in PBS) + 1.25 mM Hepes buffer at 4°C for 40 minutes (if embryos were younger than 6hpf) or overnight. The following steps were performed at room temperature, unless specified otherwise. Embryos were permeabilised in 0.5 % Triton-X 100 for 30 minutes with gentle shaking. Residual solution was removed by three washes in PBS followed by 2 hours blocking in 10% Goat Serum in PBS. Embryos were incubated in 400 µl of primary antibody solution (10% goat serum in PBS + 10ng/µl of antibody) overnight at 4°C.

The day after, samples were washed five times in 0.5% Tween-20 PBS, replacing the solution every 30 minutes. They were then incubated overnight at 4°C in the secondary antibody

solution. Secondary antibody was washed in 0.5% Tween-20 in PBS and incubated 30 minutes in mounting solution with DAPI (VectaShield, Vector laboratories, H-1000, UK).

Antibodies used: H3K4me3 (Diagenode, C15410030, Belgium), H3K27me3 (Diagenode, C15410195, Belgium).

2.2.4.2 Fluorescent In-Situ Hybridization (FISH)

Dechorionated embryos were collected at the desired stage, washed in cold PBS and fixed in 4% PFA at 4°C for 1 hour. Fixed embryos were then dehydrated in increasing dilutions of methanol (25, 50, 75, 100%) and left from overnight to one month at -20°C.

The embryos were rehydrated in decreasing dilutions of methanol (75, 50, 25%) and washed 5 times in PBST (0.1%) for 5 minutes with gentle agitation. Embryos older than 24 hpf were further permeabilised with 10 minutes incubation at room temperature of Proteinase K (10µg/ml). In order to 'acclimatise' the sample to the high temperature and the hybridization conditions, the embryos were incubated for 2 hours at 70°C in 200 µl of FISH Hybridization Buffer (FHB) (described in 2.1.5). From 50 to 100 ng of digoxigenin (DIG)-labelled RNA probes were added to the FHB and incubated overnight at 70°C. The next day, the probes were removed by four washes in increasing dilutions of 2 X SSC at 70°C (25, 50, 75, 100%). The sample was then washed twice in 0.2% of SSC at 70°C 30 minutes each. The 0.2X SSC was replaced by four serial dilutions in PBST 0.1 % in order to remove any left-over probe. Washes were performed at room temperature with gentle agitation. The embryos were at this point blocked in Blocking Buffer and Maleic Acid (Roche, 11585762001, UK) for at least 3 hours before the anti-DIG antibody was added in a concentration of 1:5000 and incubated overnight at 4°C. The next day the antibody was washed five times in 0.1 % Tween in PBS with gentle agitation at room temperature (30 minutes each) and fluorescently-labelled by horseradish peroxidase-catalysed signal amplification (Thermo Scientific, T20913, UK).

2.2.4.3 Imaging acquisition

Generally, live embryos older than 15 hours were anesthetized prior to imaging by addition of 0.03% Ethyl 3-aminobenzoate methanesulfonate (MESAB) (Sigma-Aldrich, 886862, UK). Embryos were placed in an agarose-coated petri dish and eventually embedded in agarose and imaged with a Zeiss Axio Zoom.V16 or Leica stereo microscope. Images were taken either with the Zeiss ZEN pro 2.0 or the ScanR acquisition software with standard settings.

Fixed samples were mounted in glycerol-based VectaShield (Vector laboratories, H-1000, UK) on a slide and covered with a glass slip. When preservation of the body shape was required, imaging dishes with glass bottom were used to avoid disintegration of the embryos.

2.2.5 Bioinformatic methods

2.2.5.1 ATAC-Seq analysis

2.2.5.1.1 Read quality check

Sequencing quality and statistics were checked by FastQC software (version 0.11.6).

In brief, fastq files were automatically generated by the Illumina bcl2fastq2 software (version v2.20), downloaded and merged if they were sequenced on different sequencing lanes. The merged fastq files were sampled by FastQC for standard read quality checks. No trimming or further filtering were applied to the reads.

2.2.5.1.2 Alignment to the reference genome (Zv9)

Merged fastq files from different lanes were used as input for bowtie2 (version 2.3.3.1). The reference genome (Zv9 or Zv10) was downloaded in fasta format and an index was built according to bowtie2 instructions. Alignment of the reads to the genome occurred with the following command with standard settings:

```
bowtie2 --threads 8 -x genome-index/Zv10 -q -1
ATAC_paired/combinedFile1.fastq.gz -2
ATAC_paired/combinedFile.fastq.gz -S bwt_out/sampleFile.sam
```

2.2.5.1.3 Sorting and indexing the aligned reads

After alignment, sam file were converted into bam format. In order to perform peak calling, the bam files were sorted, indexed and reads outside chromosomes discarded:

```
samtools view -bS eg2.sam > eg2.bam
samtools sort -o file.sorted.bam file.bam # sorting the bam file by
genomic coordinate
samtools index file.bam # generating an index file to navigate the
bam file
samtools view -b -@ 4 -o output.goodChr.bam file.sorted.bam chr1
chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14
chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22 chr23 chr24 chr25 #
filtering only good chromosomes
```

2.2.5.1.4 Peak calling

Peak calling was performed by macs2 (version 2.2.1) from bam files. Briefly, filtered bam files were imported into R and converted into RDS files with the functions *readGAlignmentPairs* (version 1.22.0) and *granges* (version 1.38.0). A new grange object was then generated to retrieve Tn5 cut sites with the following command:

```
cut0 = GRanges(seqnames(atacR), IRanges(start(atacR)+5-25,
start(atacR)+5+25))
```

were atacR is a variable containing the RDS files. The generated objects were exported as bed files and peaks were called in the command line with macs2.

Finally, the generated peaks were loaded into R and those with an enrichment score lower than 4 were discarded:

```
for file in *.bed
macs2 callpeak --treatment $file --format BED --gsize 1.42e9 --nomodel
--keep-dup all --outdir MACSresults/ --name $file'macsNomodelKeepdup'
peaks = subset(peaks, peaks$enrichment >= 4)
```

2.2.5.1.5 Differential openness analysis

For determining differentially opened regions between two conditions, the R package DiffBind (Ross-Innes et al., 2012) was used. DiffBinds requires as input a bam file from which it performs read count and a bed file, used to find coordinates of called peaks. The output is a Granges object which contains regions associated with a log_2FC score and an FDR value. The Granges object was subset as follow:

```
Tdrd7.DB = dba.report(Tdrd7) #subset the differential bound samples
in a grange object
totalSig = subset(Tdrd7.DB, Tdrd7.DB$FDR < 0.01) #subset of the
regions which significantly change
Downreg = subset(totalSig, totalSig$Fold < 0) #downreg in PGCs vs
Soma (or group1 vs group2)
Upreg = subset(totalSig, totalSig$Fold > 0) #upreg in PGCs vs Soma
(or group1 vs group2)
```

2.2.5.1.6 Distinguish promoters and distal elements

Promoter regions were determined by the R package GenomicFeatures (version 1.36.4) with the following command:

```
txdb <- makeTxDbFromUCSC('danRer7', 'ensGene')
promoter = getPromoters(TxDb=txdb, upstream=200, downstream=200)
## find somatic peaks overlapping promoters
distalelementsSoma = subsetByOverlaps(Downreg, promoter, invert = T,
ignore.strand = T)
```

2.2.5.1.7 Find enhancer candidates

Putative enhancers sites were defined as those regions showing open chromatin and being enriched for H3K4me1 but not H3K4me3. These regions were retrieved from previously published datasets (Bogdanovic et al., 2012) and bed files were downloaded from the NCBI Gene Expression Omnibus (GEO) (accession GSE32483). Regions with putative enhancers were loaded into R and overlapped with ATAC peaks:

```
active_enhancers =
read.csv('allstage_activeEnhancer_merged_sorted_Zv9.bed', header
= F, sep = '\t')
colnames(active_enhancers) = c('chromosome', 'start', 'end')
active_enhancersGR = toGRanges(active_enhancers)
SomaEnhancers = subsetByOverlaps(distalelementsSoma,
active_enhancers, ignore.strand = T)
```

Each putative enhancer was then associated with an ensembl ID based on the closest gene, usually within a 50000 bp range. The script assigns an ENSEMBL ID to each ATAC peak based on the first promoter found either up or downstream the peak:

```
SomaEnhancersID <- annotatePeakInBatch(SomaEnhancers,
                   AnnotationData=annoData,
                   output="nearestBiDirectionalPromoters",
                   bindingRegion=c(-50000, 50000))
```

2.2.5.1.8 Correlation between chromatin accessibility and gene expression

In order to correlate gene expression and ATAC signal, four object bins were generated according to RNA-seq scores. Each gene was assigned to a bin based on its expression value measured in tpm (see 2.2.5.2). The bins were determined as tpm < 1, 1 < tpm < 10, 10 < tpm < 100 and tpm > 100. For each bin, the ATAC enrichment over background was plotted.

2.2.5.2 RNA-seq analysis

2.2.5.2.1 Read quality check

Sequencing quality and statistics were checked by FastQC software (version 0.11.6).

In brief, fastq files were automatically generated by the Illumina bcl2fastq2 software (version v2.20), downloaded and merged if they were sequenced on different sequencing lanes. The merged fastq files were sampled by FastQC for standard read quality checks. No trimming or further filtering were applied to the reads.

2.2.5.2.2 Alignment to Zv10

Fastq files were aligned to the reference genome (Zv10) by STAR-2.6a software with standard parameters. Multi-mapped reads were allowed and outputted in a separated file, no mismatch filter was applied. Minimum and maximum intron length were set as default for a large genome. The parameter *alignSJoverhangMin* and *alignSJDBoverhangMin* refer to the minimum allowed overhang overlap on adjacent splice junctions.

```
for file in *_1P.fastq.gz
do
 basename="${file%_*}"
 STAR --runThreadN 8 --genomeDir ~/ --quantMode GeneCounts --
readFilesIn $file $basename"_2P.fastq.gz" --readFilesCommand zcat
--outFilterMultimapNmax 1000 --outSAMmultNmax 1000 --
outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.06 --
alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterType
BySJout --alignIntronMin 20 --alignIntronMax 500000 --
alignMatesGapMax 500000
--outFileNamePrefix --outSAMtype BAM SortedByCoordinate
done
```

With these settings, STAR generates count files that can be directly used as input for differential expression analysis in R (version 3.5).

2.2.5.2.3 Differential expression analysis

Differential expression and sample correlation were obtained with the R package DESeq2 (Love et al., 2014) as vignette suggests. In brief, the following standard pipeline was applied:

```
totalReadCount = read.csv('read.count.file.out', sep = '\t')
stage = rep(c('s256', 'high', 'dome', 'somites10', 'prim5',
'prim5', 'prim5'), each = 4)
type = rep(c('PGC', 'Soma'), times = 14)
## deseq requires a reference dataframe named coldata
run.deseq.genes <- function(x){
  coldata = data.frame(stage = stage, condition = type)
  row.names(coldata) = colnames(totalReadsCount)
  dds = DESeqDataSetFromMatrix(countData = x, colData = coldata,
design = ~ stage + condition)
  dds = DESeq(dds)
  res = results(dds)
}
Diff.genes = run.deseq.genes(TotalReadCount)
Significantly differentially expressed genes were retrieved as
follows:
Sign.genes = subset(diff.genes, diff.genes$padj < 0.1) # or 0.05
Sign.up = subset(diff.genes, diff.genes$log2FoldChange < 0) #
significantly upregulated in PGCs
```

2.2.5.2.4 RNA-seq reads normalisation

For some downstream analysis, RNA-seq reads were normalised by library size and gene length. Currently, two are the most widely used method for normalising RNA-seq reads:

- Reads per kilobase per million:

  (Single read/(sum of reads/1000000))/length of gene in kb

- Transcripts per million

  (Single read/length of the gene in kb)/(sum of reads/1000000)

Both rpkm and tpm normalisations were used for the analysis based on the pipeline requirements. For example, rpkm normalisation was used for determining RNA concentration as advised by the reference manual (see 2.2.5.2.6).

2.2.5.2.5 Retrieve differentially regulated genes upon chromatin reprogramming

In order to determine the effect on transcription of nearby enhancers, a subset of genes differentially regulated between PGCs and somatic cells at prim-5 and upregulated from previous developmental stages were selected. The rational was to individuate those genes affected by chromatin reprogramming, filtering out those differentially regulated due to earlier mechanisms.

To do this, significantly upregulated genes in the somatic cells at prim-5 versus PGCs were overlapped with significantly upregulated genes at in the somatic cells at prim-5 from dome stage with the following command:

```
Soma_prim5_ids = as.vector(rownames(Up.prim5.somatic))
overlap_Soma_genes = Up.prim5[Soma_prim5_ids,]
overlap_Soma_genes.prim5 = na.omit(overlap_Soma_genes)
#overlap upregulated genes in soma vs PGCs and from previous dome
Soma_prim5_dome_ids = as.vector(rownames(Up.soma.dome.prim5)) # get
ENSEMBL IDs
overlap_Soma_genes.dome= Up.prim5[Soma_prim5_dome_ids,]
```

2.2.5.2.6 Determine RNA concentration based on rpkm

A calibration curve was generated based on ERCC rpkm values and the respective known concentration. For each sample, a standard equation was calculated by the following command:

```
line <- lm(log(ercc.rpkm) ~ log(ercc.standard.concentration))
#fits a regression line for the dataset
#An intercept and slope were calculated for each regression line
as follows:
intercept = summary(line)$coefficients[1]
slope = summary(line)$coefficients[2]
#The two values were used to transform rpkm in each sample into
attomoles/ng of total RNA with the following formula:
unknown.conc = slope(sample.rpkm) + intercept
```

2.2.6 Animal procedures

2.2.6.1 Animal procedures

All animal work was performed under the Project Licence # Ib6b8b391 and in accordance with the Home Office Regulation. Up to eight pairs of adult zebrafish were kept in 3.5 litre polycarbonate tanks in a ZebTEC recirculating housing system (Techniplast, UK) with water temperature at 26°C. Adults were fed at least three times a day with a combination of brine shrimps cysts and ZM Medium Premium Granular dry food (ZMSystems, UK) and kept on a light-dark cycle of 14 and 8 hours.

2.2.6.2 Zebrafish strains

All wild-type lines came from AB and AB* strains. For PGCs detection, Dr. Roland Dosch (Georg August University, Göttingen) kindly donated Tg(Buc-GFP)$^{+/+}$ embryos. Adult Tg(Buc-GFP)$^{+/+}$ were outcrossed with ABs and the Tg(Buc-GFP)$^{+/-}$ offspring used for most of

the described experiments. This was to maximize egg yield, as we identified heterozygous adults as better breeder than homozygous.

2.2.6.3 Egg collection

AB and transgenic crosses were performed by coupling equal number of males and females in the evening. For transgenic outcrosses, AB females were crossed with Tg(Buc-GFP)$^{+/+}$ males in a ratio of 2:1 by keeping them over night separated by a transparent gate. All crosses were set in 1 or 3 litre/s breeding cages provided with a bottom mesh. The next morning, the gate was removed, fish eggs were naturally laid and fertilised and collected at intervals of 5 minutes to ensure synchrony.

Embryos were immediately collected and then incubated into E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl2, 0.33 mM MgSO4) at 28.5°C up to five days. Long incubations were supplemented with Methylene blue (Sigma-Aldrich, M9140, UK) in order to prevent algae/bacteria growth.

2.2.6.4 Microinjections

In-vitro transcribed mRNA or morpholinos were injected into fertilised single-cell-embryos by a pressure-controlled microinjector from Tritech Research, US. Chorionated embryos were collected immediately after fertilisation in Petri dishes and all the liquid was removed. About 2-3 µl of injection solution were loaded into pre-pulled glass needles and injected into the embryo. About 1 nL of solution was injected per embryo. Successful injections were detected by phenol red (Sigma-Aldrich, P0290, UK) and/or rhodamine (Sigma-Aldrich, R6626, UK). Once injected, embryos were transferred into E3 buffer and incubated at 28.5°C for further analyses. Injection solution was prepared as described below:

Component	µl in a 10 µl reaction	Final Concentration
In-vitro transcribed mRNA	-	25 ng/µl
Phenol Red solution (stock)	2	0.2 X
Nuclease-Free Water	To 10	-

2.2.6.5 Embryos dechorionation

Dechorionation occurred either by manual removal of the chorion by sharp forceps or by pronase treatment. About 1 ml of pronase solution (Sigma-Aldrich, 10165921001, UK) (10 mg/ml of powdered pronase in ddH2O) was added to a single petri dish (about 25 ml E3) and left until about 10% of the embryos came out of the chorion. The rest of the embryos were dechorionated by washing them four times with E3 medium and transferred to petri dishes with fresh E3 medium.

2.2.6.6 Embryo sacrifice

Embryos were sacrificed according to the Home Office regulations. When necessary, embryos were treated with excess of the anaesthetic 3-amino benzoic acidethylester (Sigma-Aldrich, 127671, UK) diluted in E3 medium and death was confirmed by maceration.

2.2.7 Online resources and tools

NCBI: http://www.ncbi.nlm.nih.gov

NCBI GEO: http://www.ncbi.nlm.nih.gov/geo/

PubMed: http://www.ncbi.nlm.nih.gov/pubmed

UCSC genome browser: https://genome.ucsc.edu/

Sequencing tracks were uploaded and visualised in the UCSC genome browser.

ENSEMBL: https://www.ensembl.org

ENSEMBL IDs where obtained using Biomart. Genes involved in germ line development were retrieved under the accession GO:0007281, while GO:0032502 was used for developmental genes. Pluripotency-associated genes were identified by the accession GO:0019827, while housekeeping genes were retrieved under the accessions GO:0030964 and GO:0044391.

STRING: https://string-db.org/

Primer-3: http://primer3.ut.ee/

GeneTools: https://www.gene-tools.com/

2.2.8 Statistics

Shapiro test was used to evaluate the statistical distribution of datasets. For normally distributed datasets, significance was estimated by t-test, while for non-normally distributed datasets the Wilcoxon test was used.

3 Isolation and characterisation of zebrafish PGCs during the first 24 hours of development

3.1 Introduction

This chapter aims to characterise the zebrafish PGCs mainly from the point of view of the germ plasm localisation. The principal goal was to observe and describe key stages during zebrafish germ line development in order to select meaningful time points for downstream analyses. In fact, the overall purpose of this thesis is to study the early phases of PGC development, identify crucial steps for germ fate acquisition and better understand the involvement of the germ plasm. Globally, this study was motivated by the hypothesis that germ cell and germ plasm behaviours are linked, in particular during the first day of zebrafish development. Indeed, according to literature, the major changes in germ plasm subcellular localisation correspond to important phases of PGC development (synchronous cell division, migration, gonadal period) (Knaut et al., 2000; Köprunner et al., 2001; Strasser et al., 2008). Therefore, because the most obvious changes in germ plasm phenotypes occur during early embryogenesis, I focused on the first 24 hours of development. During this time window, the PGCs undergo several phases of germ plasm re-localisation, acquire unique migratory ability and move from original edges of the embryo to the genital ridge (Theusch et al., 2006; Yoon et al., 1997).

In order to identify potentially informative time points associated with PGC or germ plasm behaviour, I tracked early PGC development in Buc-GFP-positive embryos. This transgenic fish line accumulated Buc-GFP protein in the germ plasm, making it easily detectable with standard fluorescent microscopy (Bontems et al., 2009; Riemer et al., 2015).

After having observed and characterised the initial stages of PGC development, this chapter describes optimised isolation of GFP-tagged PGCs from zebrafish embryos at different developmental stages. Indeed, although FACS is a standard procedure utilised nowadays to

physically separate fluorescent-emitting cells (Manoli and Driever, 2012), isolation of PGCs from transgenic zebrafish embryos was never published at the time this project started.

3.1.1 The Buc-GFP transgenic line

Several zebrafish transgenic lines were generated in the past years to label embryonic PGCs (Blaser et al., 2005; Krøvel and Olsen, 2002; Meyen et al., 2015). However, a common drawback was that expression of zygotic transgenes initiate after ZGA, only allowing visualisation of the PGCs after high stage (Blaser et al., 2005). In zebrafish, the PGCs become identifiable for the first time around the 32-cell stage, when the four germ plasm masses ingress into the cytoplasm of four blastomeres. At this point, the germ plasm-carrying cells keep on dividing and only one daughter cell inherits the germ plasm (Yoon et al., 1997). This process is uniquely observable in the offspring of transgenic zebrafish females, where the *buckyball* (*buc*) and *gfp* genes are fused or via in-situ hybridization (Riemer et al., 2015). In fact, one of the main advantages of the Tg(Buc-GFP) line compared to the rest of available PGC-tagging transgenic lines was that the maternal protein allows visualisation of the germ plasm from the very beginning of zebrafish development.

The *buc* gene is required for maintaining animal-vegetal polarisation and germ plasm assembly in the oocyte and in the zygote. Importantly, the Buc protein is restricted to germ cells and it was shown to co-localise with the Balbiani body in mature and embryonic germ cells (Bontems et al., 2009; Marlow and Mullins, 2008). It was shown that, when fused to GFP, Buc could be used to label the germ plasm the early embryo. This was achieved by injection of in-vitro synthetized RNA into the 1-cell embryo (Bontems et al., 2009). However, when RNA is injected into the embryo, a variable amount of time is required for the translation of the ectopic, fluorescent protein in order to be detected. Therefore, the Buc-GFP could be detected around

128-cell stage at the earliest time point when injected as RNA in the zygote. In 2015, the transgenic Tg(Buc-GFP) fish line was generated (Riemer et al., 2015); the *buc* 3'UTR, required for *buc* localisation within the germ plasm, was cloned downstream the *gfp* gene. Upstream of the *gfp* sequence, the whole *buc* gene and the endogenous distal promoter (1000 bp upstream the TSSs) were cloned (Figure 3.1). The construct was then inserted into a WT genome by Tol2 transgenesis. As expected, embryos of *buc-gfp*$^{+}$ mothers all showed fluorescent germ plasm from the 1-cell stage (Figure 1.4B) until prim-5 stage (Bontems et al., 2009). This is advantageous because it allows germ plasm detection much earlier than with any other available labelling method. The transgenic Buc-GFP embryos were kindly gifted by Dr. Roland Dosch, grown and used for the experiments described in this thesis.

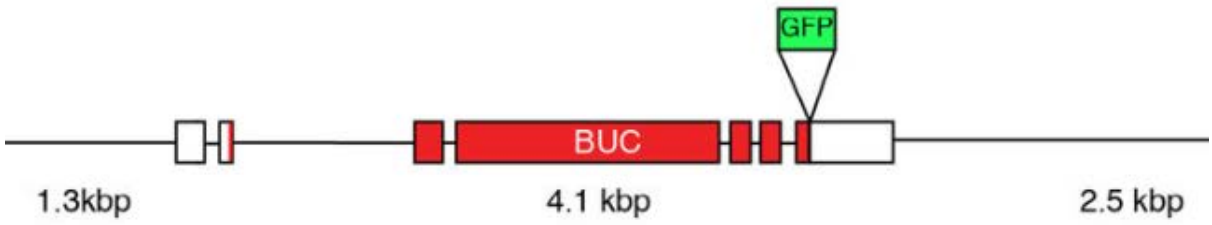

Figure 3. 1: the Buc-GFP construct.
The *Buc* gene made of 6 exons (red), the 3' and 5'UTRs (white) and the upstream and downstream regions were cloned into a vector. The *gfp* gene was added between the last exon and the 3'UTR (green). The construct was inserted into the zebrafish WT genome by Tol2 transgenesis (Riemer et al., 2015).

3.2 Results

3.2.1 PGC and germ plasm dynamics during the first day of zebrafish embryogenesis

In order to confirm the expression of *gfp* in the offspring of Tg(Buc-GFP) mothers and verify their use in late applications, I used fluorescent microscopy to track the distribution of the germ plasm and position of the PGCs in early zebrafish embryos by fluorescent microscopy. Immediately after the first and the second cell division, the germ plasm segregates with the cleavage furrows. The first cleavage pulls the germ plasm along the division axis and the second splits it along the orthogonal plane. The germ plasm is therefore brought to four poles of the animal cap and keeps this distribution for three other cell cycles (Figure 3.2). It is noticeable that smaller and weaker GFP localisations are observed at non-canonical positions, generating sometimes three-four extra germ plasm masses, in accord with previous reports (Bontems et al., 2009).

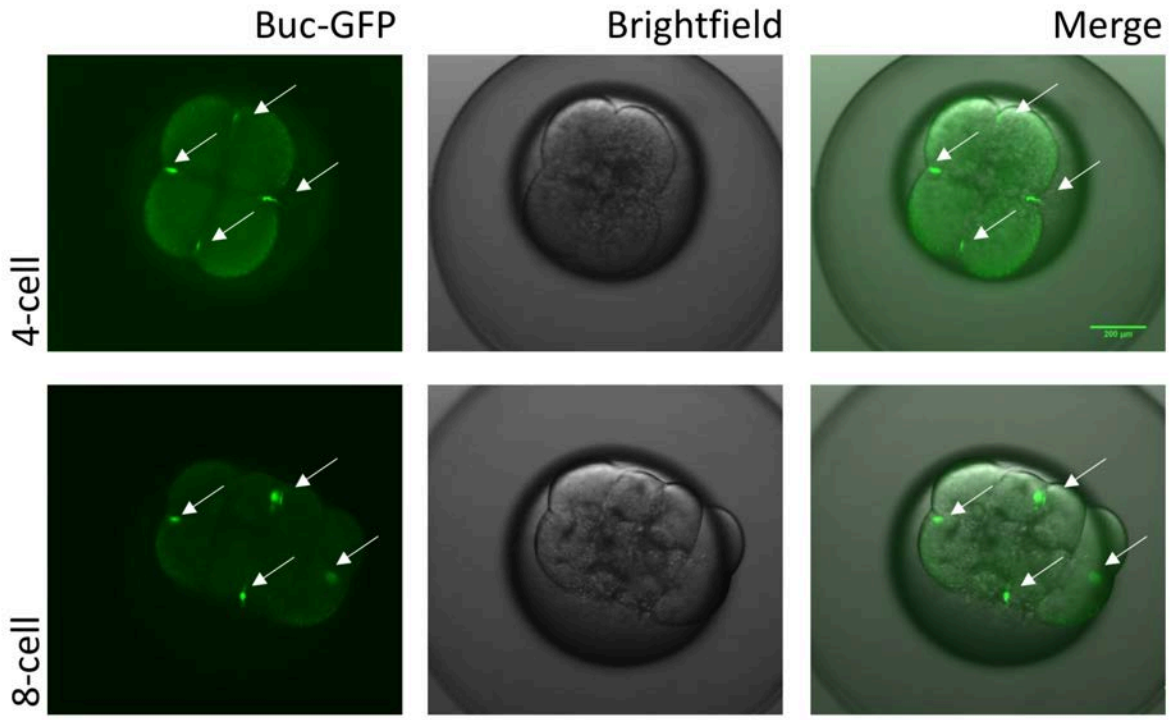

Figure 3. 2: Animal view of the offspring of Buc-GFP^{+} transgenic females.
Fluorescent microscopy and brightfield of Buc-GFP at 4-cell stage (top) and 8-cell stage (bottom). White arrows indicate green fluorescent germ plasm segregating with the cleavage furrows. Scale bar is 200µm.

At the 32-cell stage, the germ plasm transitions from extra- to intra-cellular. At this point, from three to seven germ plasm-carrying cells can be identified (Figure 3.3A, B).

In order to visualise early dynamics of smaller germ granules, light sheet microscopy was used. Analysis of whole embryo with labelled PGCs showed that smaller germ granules were spread across the cytoplasm of many cells with no specific pattern (Figure 3.3A). However, these granules fade with time, probably due to diffusion of GFP accumulations.

From 32-cell to post-MBT stages, the germ plasm-carrying cells showed indirect movements triggered by division of nearby cells, pushing the PGCs towards the yolk syncytial layer (YSL) (Figure 3.3C). At this stage, the main germ plasm masses segregate with only one cell at each division. However, when the first wave of zygotic transcription ends (around 3.5hpf), the germ plasm is inherited by the two daughter cells and eventually the number of germ plasm-carrying cells increases. I noted that the total number of germ granules per embryo (defining a germ cell) at high stage was between five and nine (Figure 3.3B).

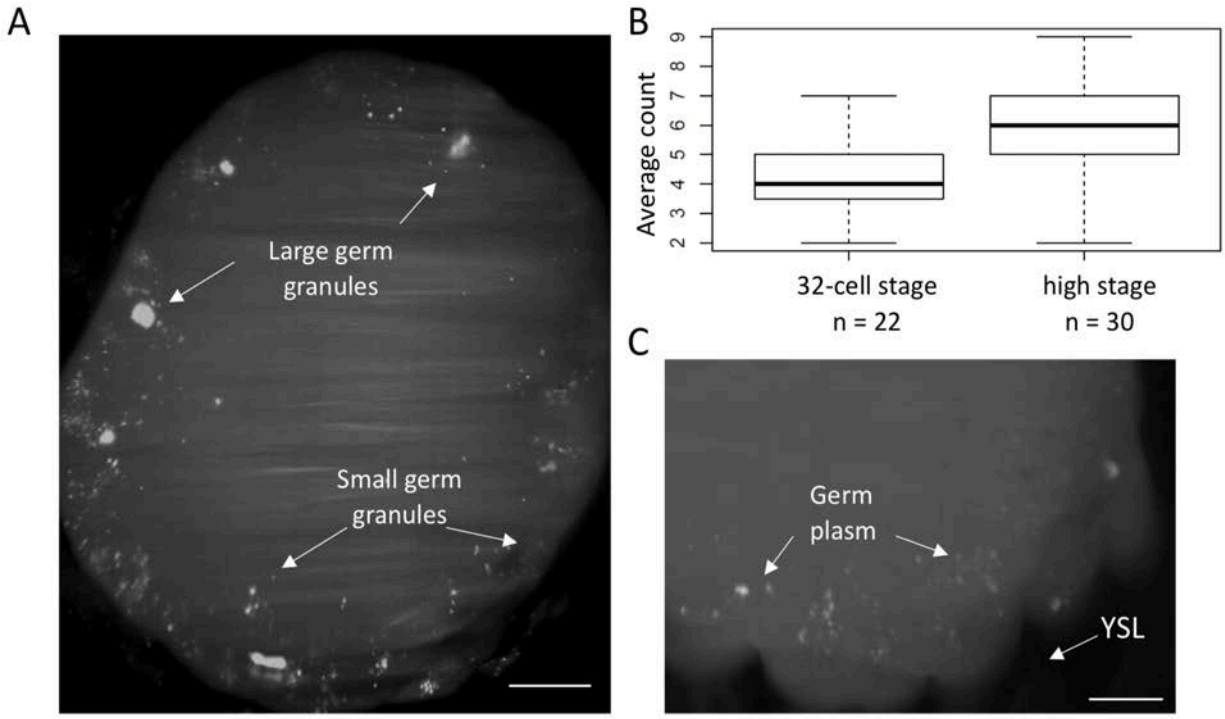

Figure 3. 3: Germ plasm distribution and number of identified germ plasm-carrying cells during the early stages of zebrafish embryogenesis.
(A) Top view of a zebrafish embryo at 16-cell stage in which the germ plasm is labelled by GFP reveals large and small germ granules at the periphery of the animal cap (white arrows). Scale bar is 50 µm. (B) Boxplots showing the median of the number of detected germ granules per embryo at 32-cell and high stages. n is number of embryos counted. (C) Lightsheet microscopy image of a 128-cell stage embryo. Germ plasm is indicated by the white arrows. Position of the yolk syncytial layer (YSL) is shown. Scale bar is 50µm.

As migration of PGCs has been extensively studied by other laboratories, it will not be described here. The germ plasm-carrying cells initiate independent and active movement around 4 hpf (dome stage). Nevertheless, for the purpose of this thesis, it is important to report the change of germ plasm shape. At 30%-epiboly (4.5 hpf), the germ plasm is partially fragmented while a dominant mass is still distinguishable in the cellular cytoplasm (Figure 3.4A). The fragmentation of the germ plasm completes during the end of epiboly and the germ granules spread evenly in the cytoplasm of the carrying cells at this point (Figure 3.4B). At 1-4 somites stage (10.5 hpf), the PGCs are mainly located dorsally and are migrating towards the genital ridge. Around this time, the germ plasm is perinuclear. The germ granules that previously were distributed evenly within the cytosol, are found in proximity of the nuclear

membrane (Figure 3.4C). Importantly, this perinuclear distribution is stable and is preserved at least for five days after this stage (data not shown).

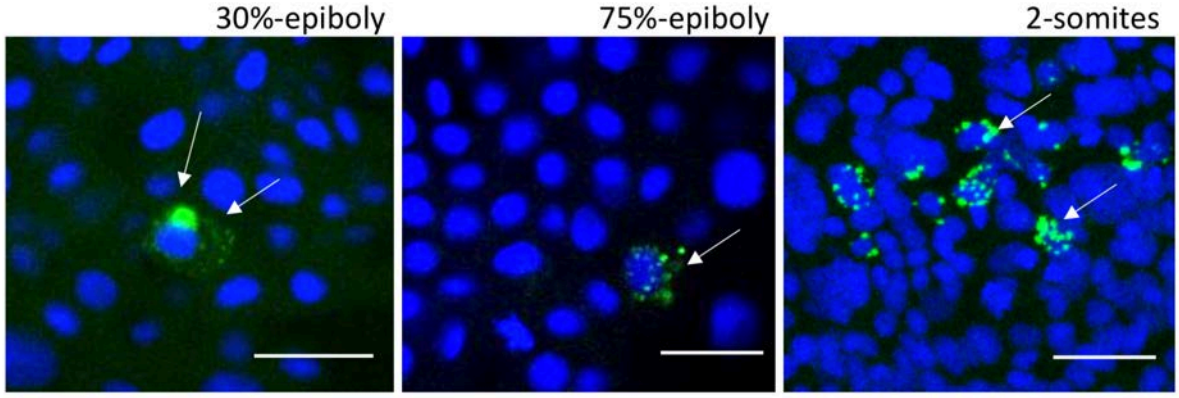

Figure 3. 4: High resolution confocal microscopy of germ granules distribution during early embryogenesis.
At 30%-epiboly, most of the germ plasm is found as a single cytoplasmic mass, however, smaller fragments are detectable in the cytosol. Fragmentation of the germ plasm has completed at 75%-epiboly and large germ plasm aggregates are not observed. At 2-somites stage, the germ granules are perinuclear. Scale bar is 20µm. White arrows indicate germ granules.

In summary, the described results confirmed that the germ plasm can be visualised and tracked in the offspring of Tg(Buc-GFP) mothers. Moreover, high resolution microscopy allowed in-depth studies of germ granules localisation and highlighted a general pattern, which brings those in proximity of the nuclear membrane during PGC migration.

3.2.2 Isolation of PGCs via FACS

3.2.2.1 Introduction

Zebrafish is considered by the scientific community as an excellent model organism for germ cell studies as they are easily trackable in their embryos. Of great contribution is the fact that germ cells restrict maternal molecules within the germ plasm, allowing easy tagging of the accumulated proteins. Additionally, the zebrafish embryos exhibit transparency and extra uterine development, making this organism suitable for identifying, tracking and studying germ cells behaviour. In the past, studies on zebrafish PGCs were exclusively dependent of

microscopy. In fact, the very low number of PGCs compared to the whole embryo makes downstream molecular analysis challenging.

Cell isolation via fluorescent-activated cell sorting (FACS) is a standard procedure used to distinguish, count and physically separate populations of cells with distinct light-emitting patterns. It allows efficient and fast isolation of cells based on size, shape and fluorescent signal by differentially charging the cell-carrying droplet which meets the desired features. The charged cell is then loaded in the collection tube after passing through a two electromagnet which split the cell flow.

For the purpose of this project, efficient isolation of PGCs at various developmental stages was needed. We decided to use a FACS-based isolation approach to separate GFP-expressing PGCs from the rest of the embryo.

3.2.2.2 Optimisation of PGC isolation from zebrafish embryos

In order to study the transcriptional and epigenetic features of PGCs at different developmental stages a FACS-based PGC isolation procedure was tested and optimised.

As reported in 2.2.6.2, after having grown up homozygous transgenic Tg(Buc-GFP) fish, the line was kept in heterozygosity for improving embryo yield. In fact, the homozygous adults immediately showed poor sexual interest for their partners (although healthy appearance), probably due to the fact that this line was inbred for many years. However, heterozygous adults were excellent breeders and more than suitable for our germ plasm-tagging goal, as each heterozygous mother generates GFP-positive embryos. Therefore, our experiments were carried out on homozygous, heterozygous and negative *buc-gfp* embryos, as mendelian law would predict for the offspring of two heterozygous adults. This breeding strategy was crucial for the success of the described work and we highlight the relevance in view of future studies requiring high number of zebrafish PGCs.

Usually, from 1000 to 5000 Buc-GFP^{+} embryos were used for each FACS isolation. To ensure developmental synchrony, embryos were collected within 5-minute intervals after being fertilised. At the time of collection, the embryos were promptly cooled on ice in order to slow any developmental/molecular process. In fact, we noted that PGCs undergo asynchronous cell division even when dissociated at room temperature (data not shown). Therefore, to avoid PGC asynchronous division or progression of PGC development, the cell suspension was kept on ice for the whole duration of the procedure. Embryo dissociation was diversely performed depending on the developmental time. In fact, the morphology and texture of the zebrafish embryo greatly changes within few hours during the first day of development. For early blastula and gastrula stages, mild pipetting in a saline solution was enough to dissociate the whole embryo in single cells. However, the higher yolk/cells ratio in the early stages compared to the late stages required more careful cell clean up. During somitogenesis, the formation of a tough notochord can induce embryo hardening and enzymatic digestions may be required. However, we did not note any significant change in PGC retrieval with and without enzymatic treatment probably because PGCs at late stages are positioned within the genital ridge, a soft ventral structure that is not affected by the hardening of the dorsal side.

After dissociation, cells were mildly spun down in order to remove most of the yolk granules. In fact, the presence of yolk granules leads to un-necessary extensions in the overall duration of the sorting procedure. Moreover, there is no efficient way to discriminate between yolk granules and GFP-negative cells, which were also used for the experiments. We found that two washes in saline solution was generally enough to remove yolk granules and debris from the cell suspension at early stages, while only one wash was performed for the late stages.

As a single cell suspension from whole embryo comprises many cell types with different shapes and sizes, during sorting we used a low speed of about 7000 droplets/second and a wide nozzle of 100 μm to reduce cellular damage. The totality of events (All events) was size- and shape-

selected upon application of forward and side scatters in order to filter out candidate debris or dead cells (Figure 3.5A). This resulted in retrieval of about half of the droplets (P1), which were then assigned to two gates (P2 and P3) based on the fluorescent emission (Figure 3.5A). The droplets were simultaneously illuminated with green and red lasers, which allowed us to distinguish true GFP-emitting droplets from autofluorescence-emitting debris. Prior to the sorting, the correct power for both green and red lasers was determined upon compensation of a GFP-negative sample. GFP-positive cells were therefore collected upon stringent gating conditions in order to increase the purity of the sample.

The efficiency of FACS isolation was measured by calculating the percentage of GFP^- cells falling in the $FITC^+$ gate (yield) after re-sorting the isolated cells. The standard validation procedure of estimating sorting yield is to run the isolated cell collection with the same parameters and count how many events occur within the selected $FITC^+$ gate. During this validation process, the yield of isolated PGCs was estimated as 71%, with 1379 cells out of 4754 cells falling outside the $FITC^+$ gate (data not shown). We hypothesised that the low yield was due to a significant amount of doublets causing fluorescence-emitting cells to carry along attached, non fluorescent cells. In order to increase retrieval of postive cells, an enriching cell sort was performed (Figure 3.5A, right). After a first cycle, the collected cells underwent a second round of isolation in order to increase the purity of the sample. However, this resulted in a further reduction in yield to 65% when re-analysed by FACS.

One potential drawback of this approach was the several rounds of FACS the cells experience. In fact, each cycle requires a pressurized enviroment and the cells are propelled at high speed in the collection buffer. In this situation, increased cell death would potentially lead to loss of germ plasm with consequent reduction of the GFP^+ cell population. To overcome this limitation, the sorting time was shortened by performing only one round of FACS and sorting

efficiency was validated via microscopy (Figure 3.5B). At pre-gastrulation stages, the yield was estimated around 80%, while at post-gastrulation stages it was 95% (Figure 3.5C). With this strategy, an average of one PGC per embryo was retrieved at high stage, while 6 PGCs per embryos were collected on average at prim-5 stage. These results were satisfactory and led us to proceed with downstream assays.

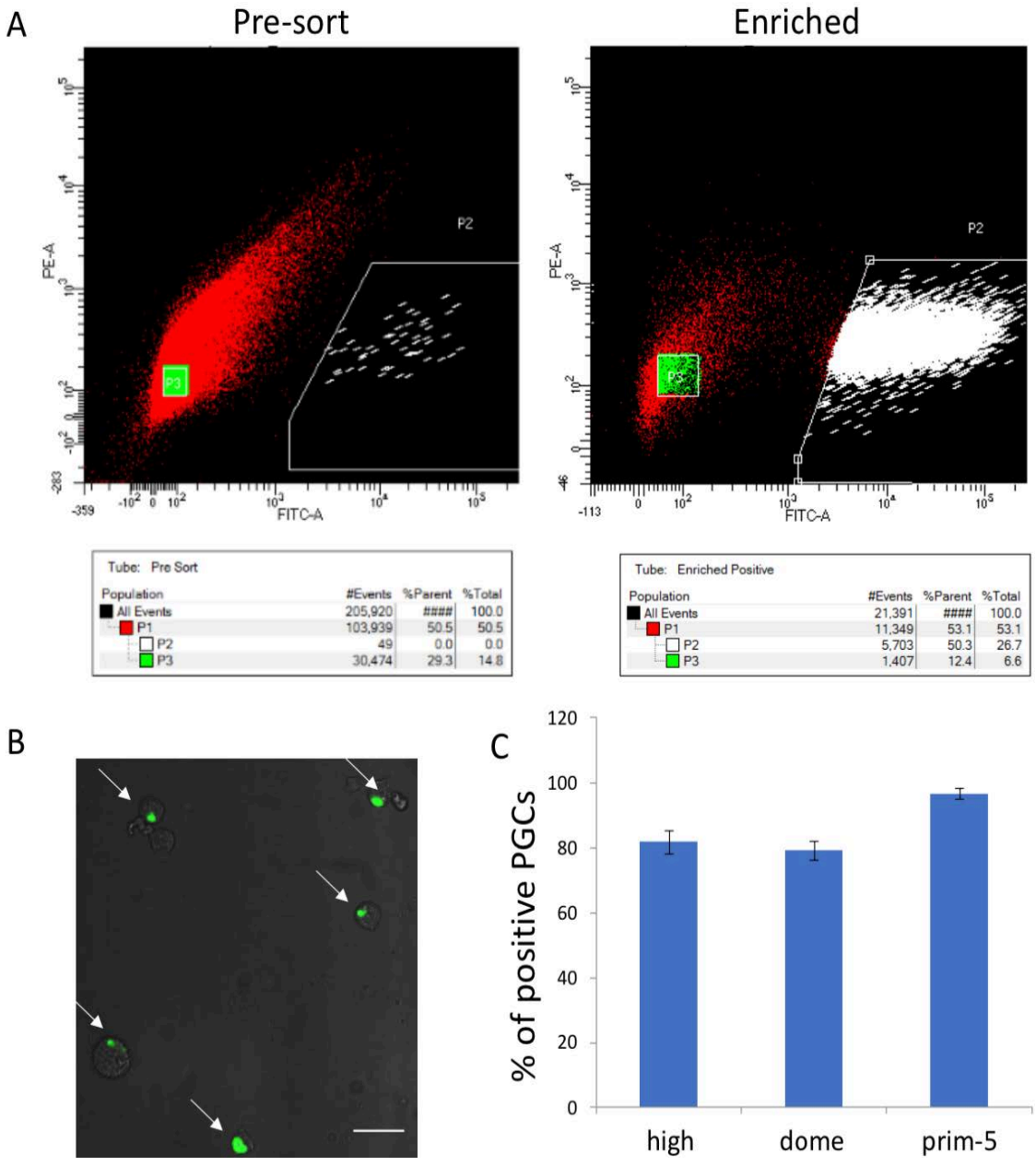

Figure 3. 5: Purity validation of isolated cells via FACS.
(A) Panel scatterplots of FITC-A vs PE-A before (left) and after (right) enrichment. PGCs are in white (P2) and somatic cells in green (P3). Grids report the counts of events/cells falling into each gate in respect to the parent gate and to the total. (B) Confocal microscopy of isolated GFP-positive cells. Scale bar is 20 µm. White arrows indicate germ plasm-carrying cells. (C) Percentage of GFP-positive cells after FACS isolation at high, dome and prim-5 stages. Bars indicate standard error.

3.2.2.3 Validation of cell purity by qPCR

The tested isolation protocol was performed on germ plasm-carrying cells and somatic cells at high, dome and prim-5 stages. Because our end goal was to retrieve usable DNA and RNA from pure populations of PGCs and somatic cells, I further verified purity and yield via RT-qPCR. In the analysis I included PGCs and somatic cells isolated from prim-5 stage. In order to avoid bias caused by comparison between a pure population of cells (PGCs) and a mix of many different cell types (somatic cells that include every GFP-negative cell), I have also attempted to manually purify brain tissues from prim-5 embryos. Because no transgenic line allowing selective labelling of the brain tissue was available, entire brains from prim-5 stage embryos were dissected (Figure 3.6).

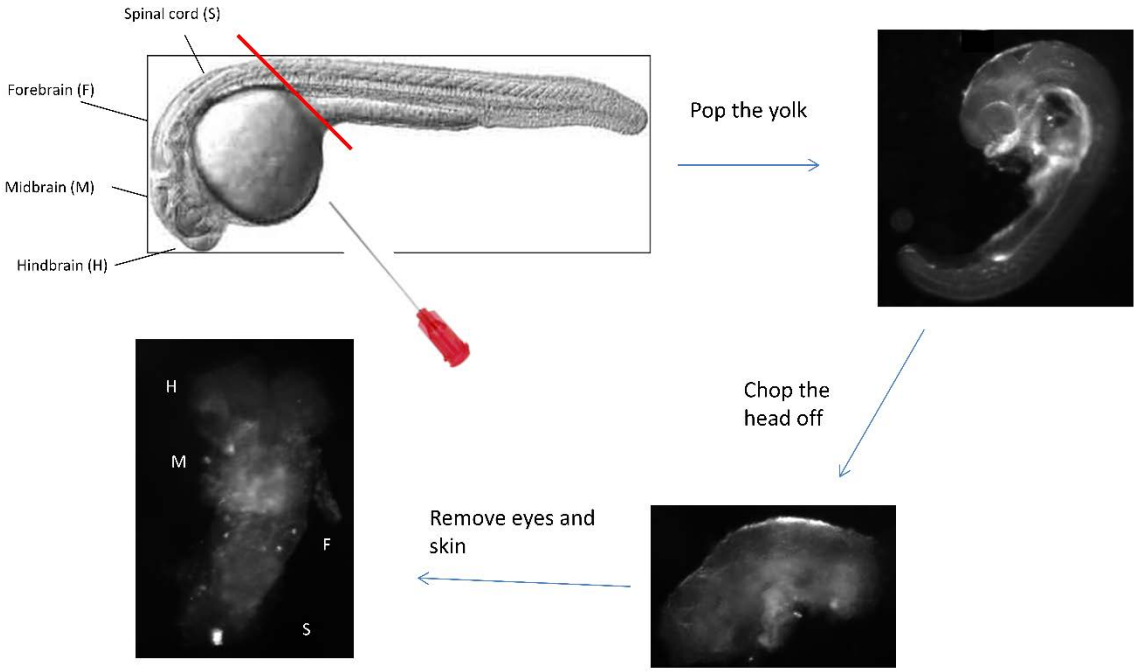

Figure 3. 6: Dissection of the brain from a prim-5 stage embryo.
The yolk was removed with a serological needle and the head was cut off at the far end of the forebrain (red line). The head was further dissected by removal of the eyes and skin prior to dissociation. Hindbrain (H), Midbrain (M), Forebrain (F) and Spinal cord are marked (S).

Once the brain was dissected, FACS was used in order to remove debris and to allow reliable comparison with the other samples.

The total RNA from sorted cells was extracted and converted to cDNA for qPCR analysis. Enrichment of gene expression was checked by amplification of selected markers for PGCs, brain and muscles (Table 3.1). Relative gene expression was estimated by double-delta fold change calculated by normalising the raw Ct values obtained with each primer pair by the Ct of a standard housekeeping gene (*beta* actin). The fold change gene expression relative to a reference marker was then plotted.

Table 3.1: List of primers used for purity validation.

Name	Forward sequence	Reverse sequence
Dnd1	TTCACTCTTCATGGCTCGTG	GTCAACAGACTCGGCTCTCC
Nanos3	AGACTGAGGCCGTGTACACCTCTCACTACT	GAGCAGTAGTTCTTGTCCACCATCG
Myod1	TCCGTCTTCTCGTCTGACAC	AGTCCGAGATCCAACTGCTC
Nanog	TCACAGCGGGCTACTTTACC	TGGTCTGCTCTGCATCTTTG
Bact	TCTCTCTGTTGGCTTTGGGA	CCTGACCCTCAAATACCCCA

As expected, both *nanos3* and *dnd1* showed significant 5- and 3-fold enrichment in the PGCs when compared to the somatic tissues (Figure 3.7), supporting previous validation on isolation of pure PGC population. As confirmation, the muscle marker *myod* was significantly depleted in the PGCs when compared to the somatic cells and less expressed in the brain tissue as well. Interestingly, the marker of pluripotent-stem cells *nanog* was equally enriched in the brain and in the PGCs when compared with the somatic cells, in accord with previously published in-situ hybridization and qPCR data (Han et al., 2016).

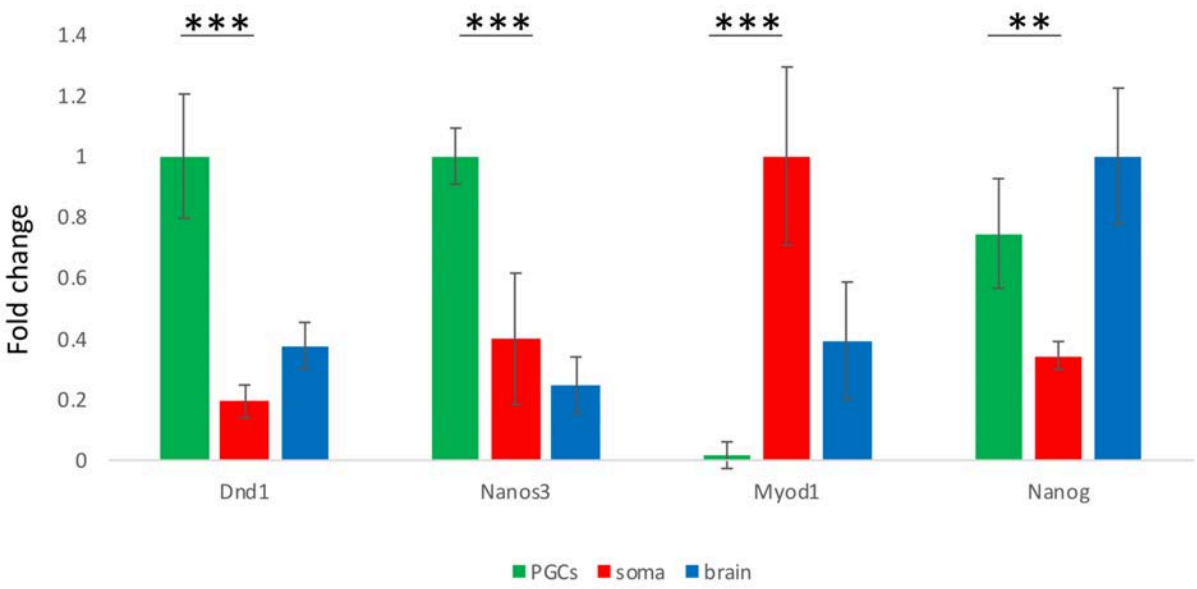

Figure 3. 7: Relative gene expression for PGC, brain and muscle markers.
Relative gene expression is shown as fold change after normalisation for the housekeeping gene *beta-actin* and the reference sample whose fold change is set to 1. Error bars indicate standard error. T-test was used to assess p-values: ** = $p < 0.01$, *** = $p < 0.001$, based on Wilcoxon test.

Taken together, these results confirmed that PGCs and somatic cells were efficiently separated by the described FACS procedure and showed significant enrichment for known tissue-specific marker. Having verified that the isolation procedure allowed us to obtain clean PGCs and control cells at several developmental stages, we next attempted more detailed studies of each cell type.

3.3 Discussion

In this chapter, I described the formation of zebrafish early germ cells by exploiting the Tg(Buc-GFP) line. My principal aim was to familiarise with germ plasm dynamics and optimise the physical separation of GFP-positive PGCs from zebrafish embryos for downstream genome-wide epigenetic and transcriptomic assays. Importantly, tracking of germ cells and germ plasm over time allowed us to identify three main phases of PGC development: 1) passive inheritance of germ plasm during cell division and germ cell movement; 2) spread of the germ plasm within the cytosol and active migration onset; 3) perinuclear germ plasm distribution and end of migratory movements.

I validated that the germ plasm localisation can be tracked over the beginning of zebrafish embryogenesis, during epiboly, gastrulation and early organogenesis. In order to follow GFP-tagged PGCs, a combination of standard and high-resolution fluorescent microscopy was used. In line with previous publications (Bontems et al., 2009; Riemer et al., 2015), we observed the formation of four initial germ plasm-carrying cells, whose development and migration is trackable. Notably, the fluorescent germ plasm largely spreads across the early blastomeres and the presence of several minor granules, in addition to the four main masses, complicates the definition of early germ plasm-carrying cells. This was in accord with what reported previously (Bontems et al., 2009), showing that the *buc* transgene contributes to formation of extra germ granules and candidate germ cells. However, these granules located in the somatic cells fade away between 64- and 256-cell stages and do not affect the detection of GFP-expressing cells by the FACS when compensation and correction for autofluorescence are applied. This observation raises questions regarding the nature of the germ cell. In fact, it indicates that there is no a predetermine set of embryonic cells fated to become germ cells but rather any blastomere acquiring enough germ plasm mass will be directed towards germ line

developmental pathways. A mechanism of germ plasm clearance in the early embryo has not been described yet, however, it is evident that small germ granules fade away when ZGA occurs in spite of few larger granules. Because of the vast spreading of the granules throughout the embryo, disappearance due to fusion seems unlikely. Although an apparent even spread of germ plasm could be caused by self-aggregation of GFP proteins exclusively, studies of *vasa* and *dnd* transcript localisations at early embryonic stages showed similar spread of germ granules in the embryo, suggesting that the detected GFP-positive aggregate contain additional germ plasm markers (Knaut et al., 2000; Weidinger et al., 2003; Yoon et al., 1997).

In order to better understand the role of the germ plasm at its various phases, we seek to optimize an isolation procedure for GFP-expressing cells. Here, we have reported a detailed protocol for PGC isolation from Buc-GFP-positive embryos, which was not attempted previously. Additionally, during the optimisation of PGC isolation procedure, we highlighted potential drawbacks of purity cell sorts, which may sacrifice cell viability in exchange of higher purity. Importantly, our in-depth selection of PGC stages associated with optimisation of FACS isolation provides the community with additional tools to study germ cell development.

In conclusion, although previously described, the different endeavour of the germ plasm and its consequences on cell fate have not been looked at in details so far. We highlight that the subcellular distribution of the germ granules reflects PGC dynamics and suggests biological relevance. Therefore, having verified feasibility of PGC visualisation and isolation, we asked which roles the germ plasm plays in PGC formation and how these relate to its subcellular appearance.

4 PGCs do not delay transcriptional activation and selectively retain zygotic transcripts

4.1 Introduction

Germ fate is gained in multiple ways among animals, often during the early phases of embryogenesis, through cooperative mechanisms involving germ plasm-mediated transcriptional blocks, post-transcriptional regulation and epigenetic reprogramming. Although recent studies have highlighted the importance of epigenetic reprogramming of the germ line (Gkountela et al., 2015; Guo et al., 2015; Hill et al., 2018; Skvortsova et al., 2019), a clear understanding of the relation between germ plasm and epigenetic control is missing.

In general, the acquisition of chromatin identity is a crucial step during differentiation. In many organisms, chromatin remodelling begins soon after fertilisation and involves combinatorial events leading, on one side, to preserve some parental epigenetic information and, on the other, to rebuild a new epigenetic code (Haberle et al., 2014; Zhang et al., 2018). Ultimately, the main role of an early establishment of epigenetic landscapes is to prepare the embryo for the elevated requirement of timed transcription it undertakes during ZGA, although the role the gametes play remains unclear. While maternally-provided TFs, histone proteins and enzymes have been proven to affect early chromatin regulation, the existence of an 'epigenetic inheritance' is still debated. Lines of evidence are recently supporting cases of epigenetic inheritance between gametes and embryos. Studies on early mouse embryos, for example, highlighted the relevance of histone grammar during ZGA. Low scale ChIP-seq on murine ESC and gametes revealed that oocyte-inherited H3K4me3 mainly tends to be preserved around the TSS and lost at other regulatory regions (Dahl et al., 2016). As suggested by the retention at the TSSs of a transcriptional-enhancing histone modification, it was shown that reacquisition of H3K4me3 at promoters is rapid after fertilisation and coincides with the major wave of transcribed genes during ZGA. Additionally, H3K4me3-rich DNA regions at the ZGA are mostly

hypomethylated and actively transcribing. On the other hand, regions of bivalent chromatin are rare and H3K27me3 settles later during development (Liu et al., 2016). Strikingly, comparison between gametic and embryonic H3K4me3 sites indicates that epigenetic information from both the parents is transmitted and found on the corresponding chromosomes during the early phases of embryogenesis (Zhang et al., 2016).

Although these data suggest that parental information is required to initiate embryonic transcriptional programmes, the role of the epigenetic memory during later developmental phases remains unknown. Inheritance of gametic histone modifications was reported also in zebrafish, however, these are promptly replaced by de-novo zygotic marks. On promoters, parental H3K4me3 and H3K27me3 are erased upon fertilisation and re-deposited after chromatin priming via H3K27ac (Zhang et al., 2018). Also, enhancers are reprogrammed following total loss of epigenetic marks. Strikingly, these processes occur in a highly proliferative cell, where it was shown that the chromatin is more compacted and less organised compared to post-ZGA embryos (Blythe and Wieschaus, 2016; Hug et al., 2017).

Another recent study suggested that nucleosome 'placeholders' could be required to bridge parental and zygotic epigenomes. These nucleosomes are H2A.Z (H2AFV) and H3K4me1 and resolve into bivalent H3K4me3/H3K27me-marked nucleosomes or active H3K4me3-marked nucleosome upon ZGA (Murphy et al., 2018).

Along this line, PGCs are an interesting system to deepen the understanding of early parental-inherited and zygotic chromatin. Indeed, PGCs single out and show specific behaviours (migration) during early embryogenesis. Moreover, the continuity of germ material throughout the generations via transmission of the germ plasm could suggest additional inheritance of specific chromatin marks.

The chromatin features and epigenetic requirements of germ plasm-dependent PGCs are unknown. As discussed, mammalian germ cells undergo dramatic DNA demethylation and show peculiar epigenetic features such as chromatin bivalency. It is unsure whether germ plasm-dependent animals require bivalent histone marks to preserve pluripotency. In fact, while prompt chromatin reprogramming is necessary to initiate PGC formation in mammals, in a germ plasm-carrying cell this could be superfluous (Lesch and Page, 2014). Interestingly, PGC formation in zebrafish requires both oocyte-inherited factors and stem cell-like chromatin features (Wu et al., 2011), suggesting that both parental and embryonic contribution could play a role during germ line specification. Moreover, other germ plasm-dependent mechanisms of germ fate acquisitions have been described. In *D. melanogaster* and *C. elegans,* transcriptional delay during ZGA is a conserved requirement in PGCs to skip the somatic fate and preserve pluripotency (Hanyu-Nakamura et al., 2008; Mello et al., 1996).

To study the regulation of chromatin and transcription during the onset of ZGA in the PGCs, we aimed to compare the transcriptome and chromatin accessibility of germ plasm-containing cells and somatic cells using NGS and imaging techniques. Our main goal was to understand how intense genomic activation and gene transcription occur in a germ plasm-carrying cell in order to gain insight into the molecular function of the germ plasm. In particular, we questioned whether a specialised transcriptional state is observed in a germ plasm-carrying cell during or immediately after ZGA, to better investigate mechanisms linking germ factors and transcriptional regulation. For these reasons, we aimed to compare germ plasm-carrying cells to the rest of the embryo before and during ZGA. To this end, we exploited a newly developed tool to image transcript accumulation in-vivo with single cell resolution, in combination with immunohistochemistry for histone marks. Additionally, to more comprehensively study transcriptional and post-transcriptional regulation of zebrafish PGCs during ZGA, RNA-seq

and ATAC-seq were performed on isolated germ plasm-carrying cells and somatic cells. By performing whole-transcriptome and epigenome profiling, our goal was to understand which are the key mechanisms contributing to early germ cell specification and development.

4.2 Results

4.2.1 H3K4me3 and H3K27me3 are detected in nuclei of pre- and post-MBT PGCs

We first looked for any evidence of parental epigenetic memory or signs of early epigenetic differentiation in the germ plasm-carrying cells. As discussed, early embryos lose epigenetic marks soon after fertilisation and rebuild the epigenome concomitantly with transcription onset. As PGCs were shown to delay transcriptional activation, one could assume that transcriptional delay could be caused by a specific, pre-set epigenetic profile. Therefore, one potential mechanism of pre-setting the germ fate before transcription starts would be the acquisition/depletion of a determined histone mark set.

We hypothesised three scenarios: 1) The germ plasm-carrying cell undergoes histone mark acquisition alongside the rest of the embryo. This would support alternative mechanisms of epigenetic germ fate acquisition. 2) The germ plasm-carrying cell differentially rebuilds its epigenome as observed in mouse. 3) The germ plasm-carrying cell acquires or retains a gamete-like epigenetic profile. In our opinion, this hypothesis is the less likely as there is a phase when the germ plasm does not belong to any cell.

To verify these possibilities, I took advantage of two highly-represented histone modifications in the early zebrafish embryo (H3K4me3 and H3K27me3) (Lindeman et al., 2010; Vastenhouw et al., 2010) and asked if differences in their nuclear expressions were detectable between the germ plasm-carrying cells and the somatic cells via immunohistochemistry (IHC). First, I checked what was the earliest time point when H3K27me3 and H3K4me3 were detectable during embryogenesis. Previously published datasets suggest that H3K27me3 is present in the nuclei of pre-ZGA embryos, while it is debated whether H3K4me3 is present prior to ZGA (Murphy et al., 2018; Zhang et al., 2018). When enrichment of H3K27me3 in blastomeres is inspected, I could observe detectable nuclear signal starting from 16-cell stage (Figure 4.1A). On the other hand, I was unable to image H3K4me3 prior to 256-cell stage.

As H3K27me3 appeared stably present in pre-ZGA nuclei, we aimed to co-image H3K27me3 staining and Buc-GFP starting from 64-cell stage. Interesting, H3K27me3 is detected in nuclei of just-formed germ plasm-carrying cells and somatic cells (Figure 4.1B). Although the weak staining did not allow reliable quantification of the fluorescent signal, no evident differences were detected between germ plasm-carrying cells and somatic cells. Our results are in accord with what reported earlier, suggesting that mild incorporation of H3K27me3 in zebrafish genome occurs from the beginning of cell division and increases over time (Murphy et al., 2018; Zhang et al., 2018).

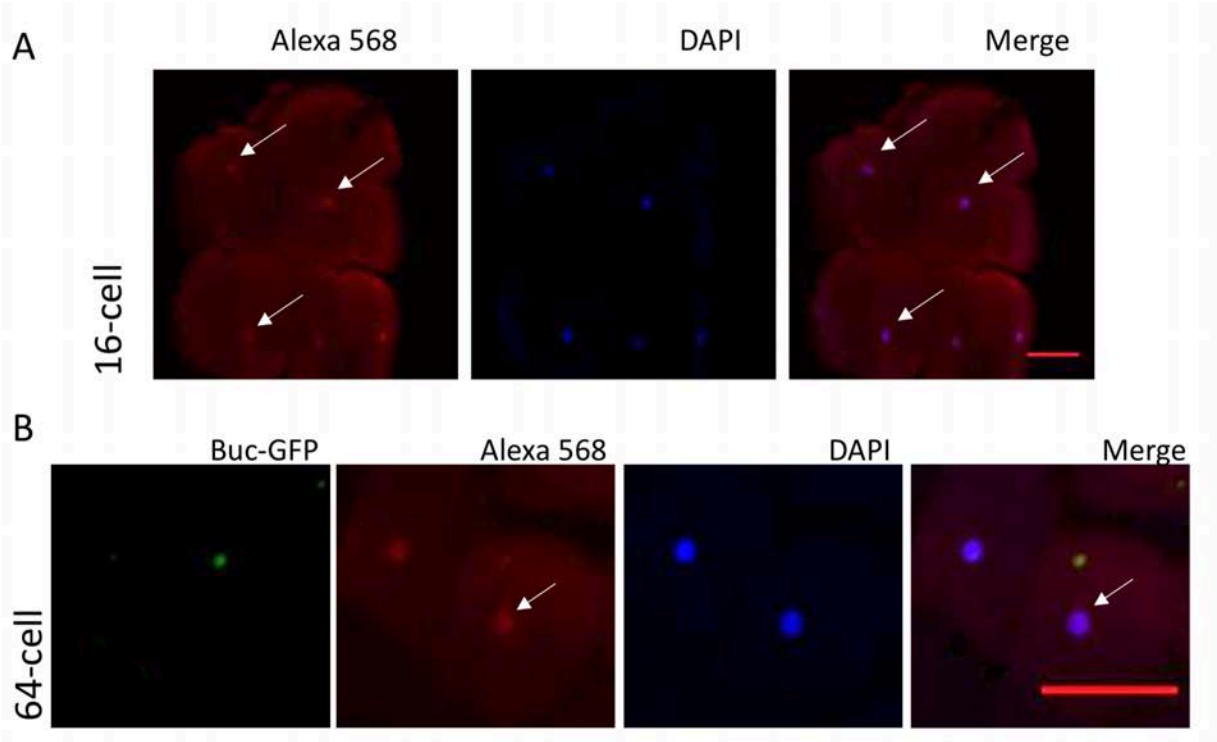

Figure 4. 1: H3K27me3 is detected in pre-ZGA embryos and germ plasm-carrying cells.
(A) Top view of a zebrafish embryo at 16-cell stage. The H3K27me3 immunostaining of PFA-fixed embryos shows nuclear enrichment of the fluorescent signal (white arrows). The germ plasm is in green. Scale bar is 50 µm. (B) Two cells of a 64-cell embryo are shown upon PFA fixation and immunostaining. H3K27me3 is found in both germ plasm-carrying cells and somatic cells at 64-cell stage White arrow indicates the nucleus of a germ plasm-carrying cell. Scale bar is 50 µm.

In order to verify whether different chromatin regulation is achieved in PGCs and somatic cells while development proceeds, both H3K4me3 and H3K27me3 marks were imaged and the

signal quantified before and after the first wave of ZGA (from 256-cell to 1k stage) (Figure 4.2A). Due to the variation in efficiency of antibody binding, comparison of signal intensities was carried out only within the same experimental samples. Interestingly, no significant difference was observed between germ plasm-carrying cells and somatic cells at any of the stages included in the experiment (Figure 4.2B). Overall, a parallel trend of signal increase over time was detected in PGCs and somatic cells for both H3K4me3 and H3K27me3, consistent with continuing acquisition of histone marks around ZGA.

In conclusion, these results show that early embryonic histone marks are gradually acquired in the blastomeres, with no significant PGC-specific pattern observed.

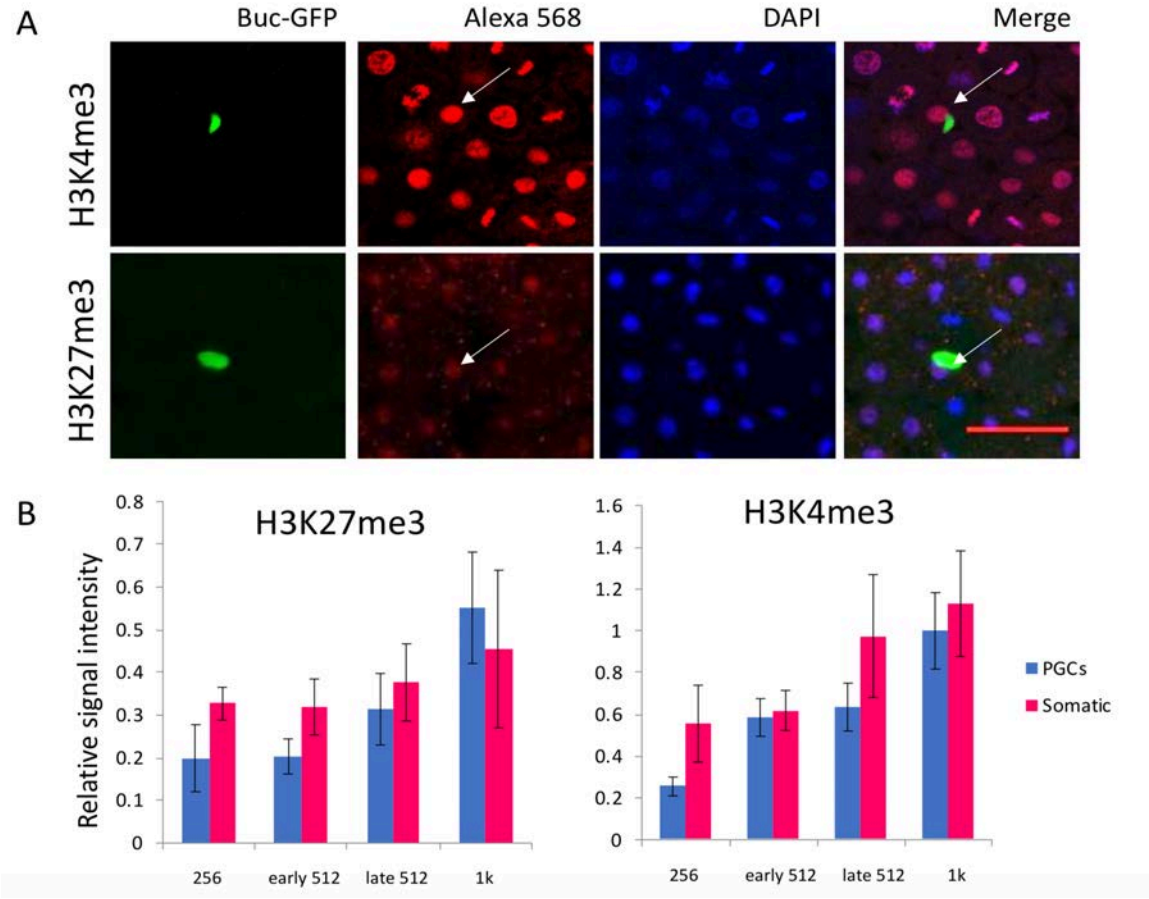

Figure 4. 2: H3K4me3 and H3K27me3 show comparable levels in early germ plasm-carrying cells and somatic cells.

(A) Immunostaining of H3K4me3 and H3K27me3 at 1k stage detected both the histone marks in germ plasm-carrying cells and somatic cells Scale bar is 50 µm. (B) Relative intensities of H3K4me3 and H3K27me3 immunostaining normalised by DNA signal (DAPI) show non-significant variation between germ plasm-carrying cells and somatic cells from 256-cell stage to 1k stage. Bar indicate standard error. The signal from 40 cells was used per each experiment.

4.2.2 *miR-430* is expressed in PGCs by 512-cell stage

Immunohistochemistry of H3K4me3 and H3K27me3 did not show specific PGC epigenetic features during the onset of zygotic transcription. Moreover, although the presence of germ granules, pre-migrating PGCs do not show behavioural nor phenotypical differences compared to the rest of the embryo. In fact, the PGCs divide at the same pace of the non-germ plasm-carrying cells. Accordingly, each germ plasm-carrying cells can generate a somatic cell after asymmetric cell division.

These observations contrasted the hypothesis that the germ plasm may affect nuclear activity already at these early stages. As explained before, the germ plasm of many organisms inhibits zygotic transcription during canonical ZGA, therefore we asked whether this applied to zebrafish as well and how.

In order to determine whether PGCs delay ZGA in zebrafish, we exploited a morpholino-based approach to detect transcriptional activity with single cell resolution. This method relies on detectable morpholinos, which can be targeted towards any nucleic acid and visualised when the signal intensity is boosted by subcellular regionalisation. In case of morpholinos targeting nascent RNAs, sites of intense transcription can be observed in the nucleus (Hadzhiev et al., 2019) (Figure 4.3A). Importantly, this approach has been successfully used to detect transcription of *miRNA-430*, one of the earlier transcripts produced by a zebrafish embryo, required to clear maternal RNAs (Giraldez et al., 2006; Hadzhiev et al., 2019). In fact, the high demand of *miRNA-430* results in massive preduction/storage sites in the nuclei of an early embryo, allowing selective detection of this transcript.

Table 4.2: Sequence of fluorescently-labelled morpholinos

Name	Target Sequence
miR-430 MO	CACACGCATCTTGTTGTCTGCTGTT
5-mismatch	CCCACTCATATTGTTGTATGCTTTT

Injection of fluorescently-labelled morpholinos targeting the 5' UTR of nascent *miR-430* showed transcription occurring by 512-cell stage in PGCs, in line with the canonical first wave of ZGA in zebrafish (Figure 4.3B,C). The percentage of PGCs and somatic cells expressing *miR-430* showed no significant variation (Figure 5B). In accord with previous data (Knaut et al., 2000), this result confirmed that zebrafish PGCs do not delay ZGA in contrast with most of other model organisms. *miR-430* is one of the first transcripts expressed during zebrafish transcriptional activation and it is in charge of degrading maternal mRNA when the zygotic genome is activated (Giraldez et al., 2006). Detection of *miR-430* transcription suggested that PGCs do not differ from the surrounding ESCs at this stage, however maternal-inherited transcripts survive. In line with previous reports, this observation confirms the passive role of the germ plasm in initiating the germ fate when the embryo activates transcription.

In conclusion, we proved that PGCs do express *miR-430* around the canonical ZGA. This was surprising as maternal transcripts localised in the germ plasm are known to be retained for several days (Knaut et al., 2000; Köprunner et al., 2001; Mishima et al., 2006). However, we show that germ plasm and *miR-430* co-exist, suggesting that germ plasm-retained transcripts are shielded from *miR-430*-mediated degradation.

In order to test whether PGCs and somatic cells differentially activate the genome, we performed RNA-seq on isolated germ plasm-carrying cells at stages spanning ZGA.

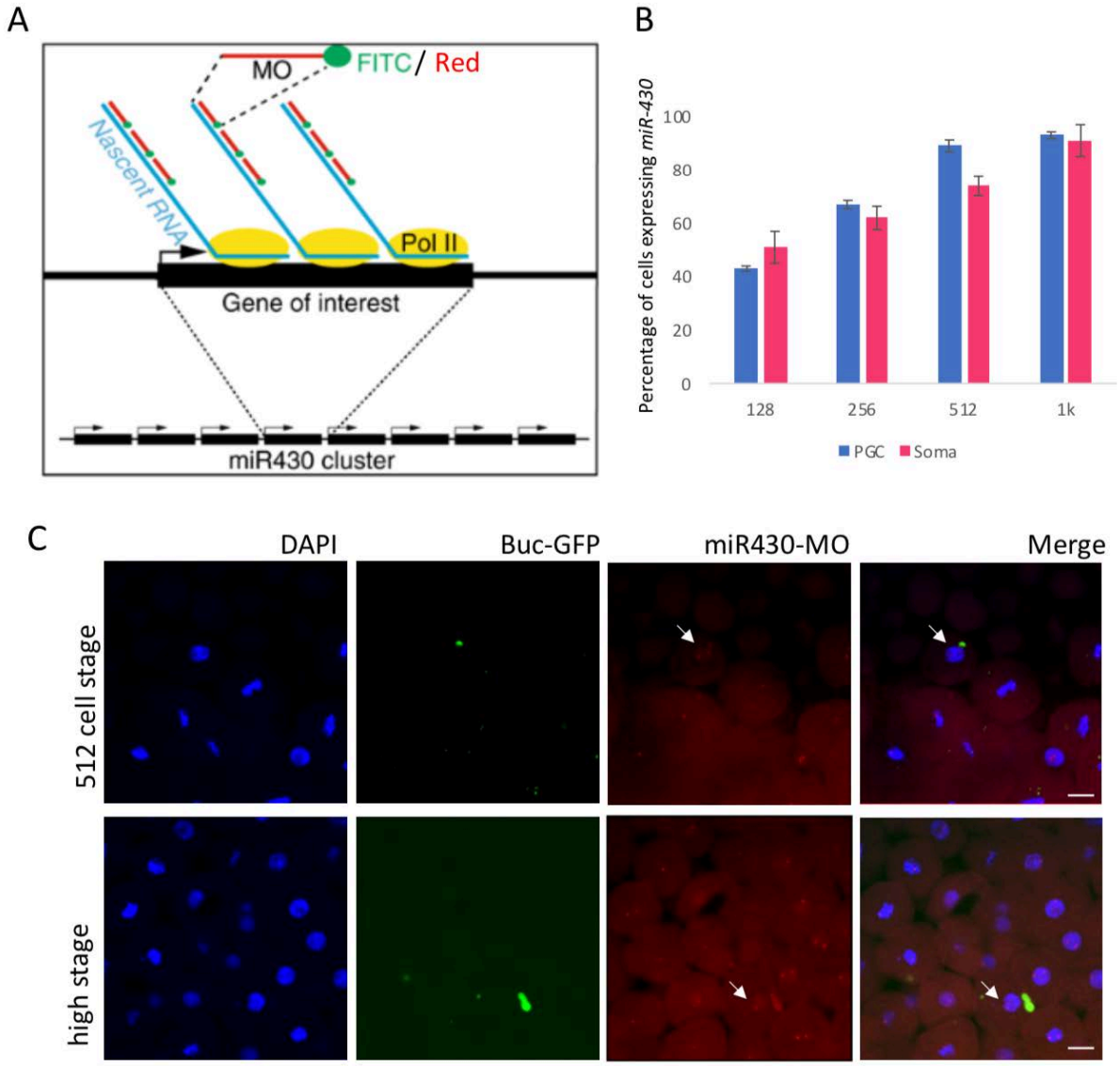

Figure 4. 3: Transcriptional activity is detected prior to the major wave of ZGA in the germ plasm-carrying cells.
(A) Mechanism of *miR-430* detection is based on the accumulation of morpholinos on highly-repetitive sequences, boosting the fluorescent signal (Hadzhiev et al., 2019). (B) Percentages of PGCs (blue) and somatic cells (pink) expressing *miR-430* at various stages. Bars indicate standard error. (C) Confocal microscopy of zebrafish embryos injected with *miR-430* morpholinos at 1-cell stage. White arrows point the nuclear foci representing sites of transcription. Germ plasm is labelled by GFP. Scale bar is 20 µm.

4.2.2.1 Transcriptome analysis of early PGCs

Although detection of *miR-430* accumulation in germ plasm-carrying cells proved that transcription of RNA occurs during the first wave of ZGA, imaging did not provide quantitative

and genome-wide information. Therefore, we aimed to perform RNA-seq on isolated PGCs before and after ZGA.

While sequencing of the total transcriptome would be preferred to study the role of regulatory RNAs, total RNA-seq was not applicable to small cell numbers at the time the experiment was performed, despite nowadays a few pioneering techniques achieving total RNA sequencing on picograms (pg) of RNA are available (Hayashi et al., 2018). When this project was initiated, the sequencing of coding RNAs exclusively was relatively straight forward even on limited amount of material. The Takara Smart Seq V4 kit allows efficient reverse transcription from one to one thousand cells of polyadenylated RNAs, allowing in-depth studies of coding RNAs. To better understand the global transcriptional features of the early germ line, I performed RNA-seq of polyadenylated transcripts at various developmental stages in PGCs and somatic cells. My first aim was to determine whether PGCs are transcriptionally globally active at the time of canonical ZGA or delayed. Because the variability in cDNA library yield was minimal between two hundred and one thousand cells according to manufacturer's instructions, two hundred cells were used for the described experiments. PGCs and somatic cells were isolated at a pre-ZGA stage (256-cell stage) and at a post-ZGA stage (high and dome stages). After cell isolation by FACS, cDNA was prepared from both GFP-positive and –negative cells at the developmental stages mentioned earlier. Polyadenylated RNAs selected and used as template for cDNA production. To enable inter-sample measurements of absolute RNA amount, equal amount of bacterial RNA was spiked into each sample according to the ERCC spike-in Mix recommendations (Jiang et al., 2011). DNA libraries were sequenced on a mid-output Illumina flow cell.

4.2.2.2 Computational analysis of genes differentially regulated in PGCs before ZGA

First, I validated RNA-seq quality and data reproducibility. Quality control of the sequencing read was overall satisfactory. The per-base sequence quality score was high for all the samples

(Figure 4.4A), low levels of PCR duplicated were found and no GC bias was detected. Raw reads were mapped to the reference genome Zv10 and to the ERCC spike-in reference as described in the methods. The mappability of all reads aligning to the reference genome or to the expected spiked-in transcripts was ranging between 85 to 93% (Figure 4.4B).

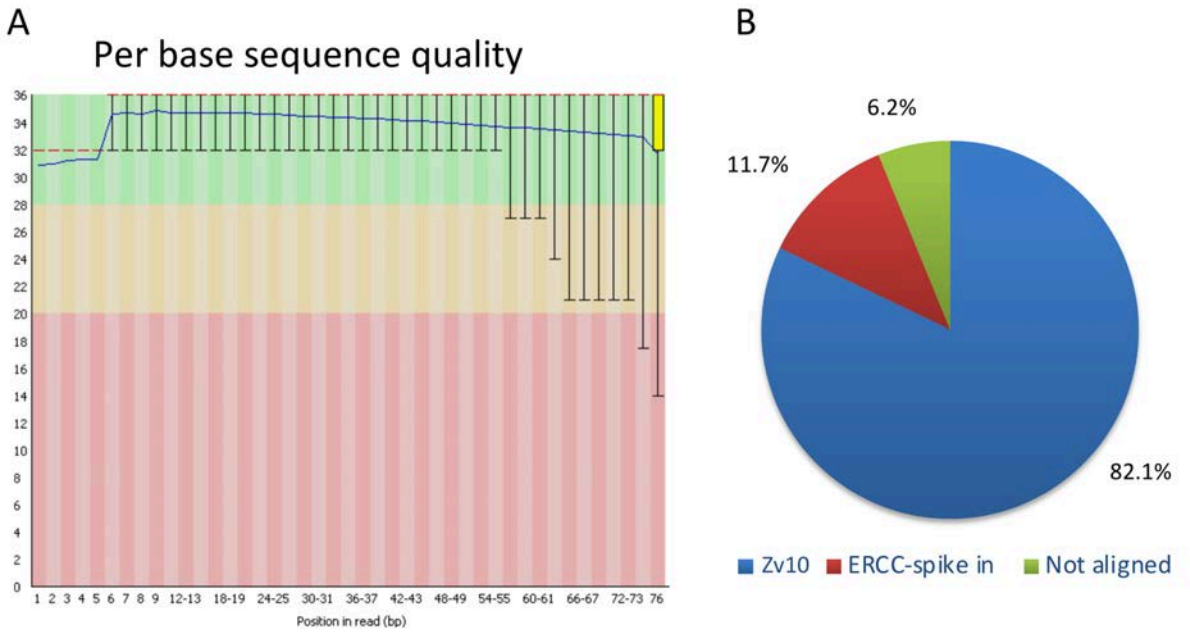

Figure 4. 4: Quality control and mappability of RNA-seq data.
(A) Per base sequence quality output of the Fastqc software shows overall high quality of bases. Drop of sequence quality for the last two bases is standardly observed. The y-axis reports the sequence quality score. Scores between 28 and 36 (green area) indicate that each base is assigned with 99% of confidence. (B) Pie chart showing percentage of reads mapped to the reference genome (blue), to the ERCC-spike in control (red) and not-aligned reads (green).

Reproducibility between replicates was checked by performing Principal Component Analysis (PCA). This is a standard statistical procedure during RNA-seq data analysis and provides a numerical overview of correlation among multiple samples. The numerical measure of correlation is the variance, which is calculated along the axis with the highest possible variability (Principal Component 1, PC1) and the subsequent orthogonal axes (PC2 and PC3). When PCA was plotted for the RNA-seq samples, minimal variation among replicates was observed (Figure 4.5). Importantly, all the sample distributed along the PC1 according to developmental time points and along the PC2 according to cell type. Importantly, as the PC1

axis corresponds to the highest variability, this analysis indicated us that variation among samples is greater between developmental stages than between PGCs and somatic cells.

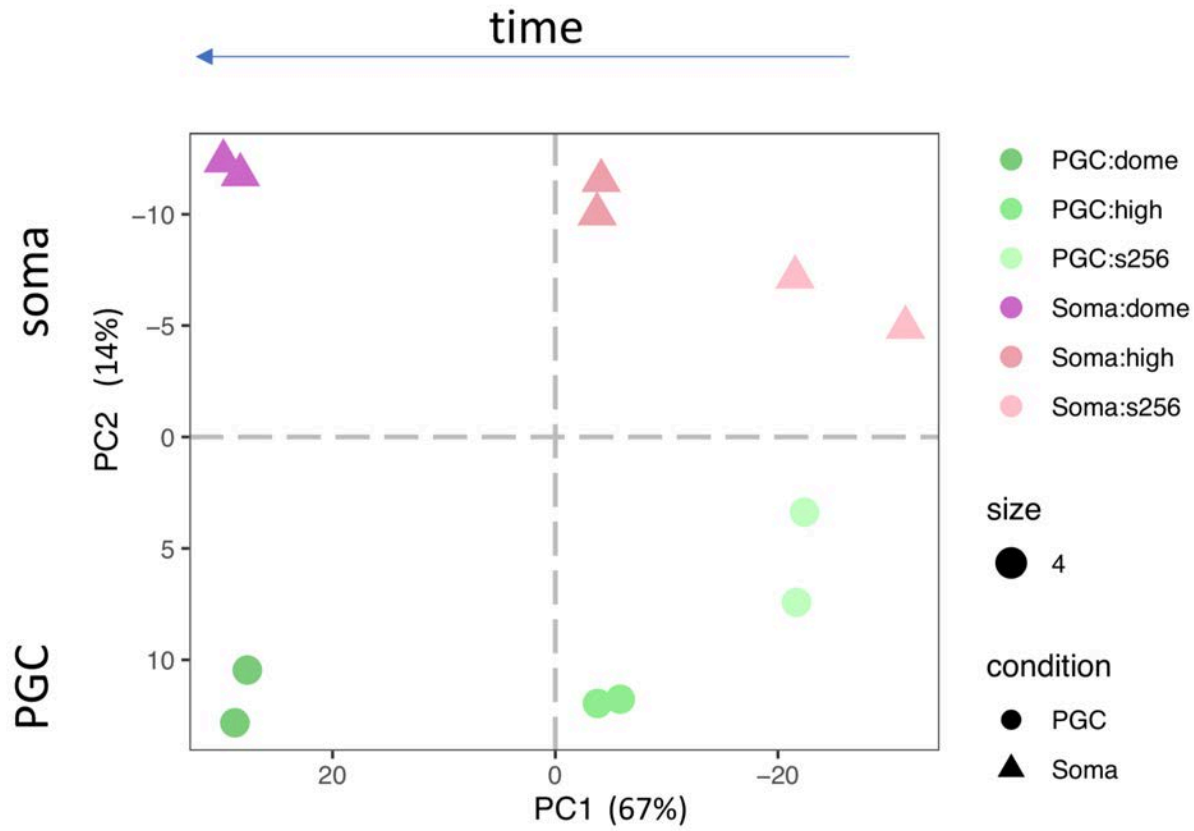

Figure 4. 5: Principal component analysis of the germ and somatic transcriptomes over early development.
Developmental time points distribute along the x-axis of the PCA, representing the greatest contributor to the variance (67% variance). Differences between cell types distribute along the y-axis (14% variance). Sample variance was calculated in R using DESeq2 package. PGCs and somatic cells are shown as circles and triangles in shades of green and purple, respectively.

4.2.2.3 Identification of new candidates for germ plasm marking by RNA-seq

Having assessed that the data quality was satisfactory and that the differences between time points were greater in comparison with the differences between cell types, I compared PGCs and somatic cells by performing Differential Gene Expression (DGE) analysis, with the aim of identifying germ cell-specific transcripts. When total polyadenylated transcriptomes of PGCs and somatic cells at 256-cell stage were compared, only twenty-three genes were found to be significantly upregulated in the PGCs (Table 8.1, Appendix), while there was no significant

upregulation in the somatic cells (Benjamini-Hochberg (BH) adjusted p-value < 0.1). This number was adjusted to twenty-two after application of an expression threshold based on average transcript per million score applied to all the stages and samples (Figure 4.6A). As expected, known germ plasm markers such as *ddx4*, *dnd1* and *nanos3* were strongly upregulated in the PGCs, indicating that the isolation procedure was successfully enriching for germ plasm-carrying cells.

It is noteworthy that ten out of twenty-two identified transcripts had not previously been associated with germ cell or PGC functions (Figure 4.6B, orange). The remaining twelve transcripts instead were either known germ plasm markers (*gra, tdrd7, rg514a, ca15b, dnd1* and *dazl*) or associated with germ cell development/survival (*hook2, tgfa, zswim5, b4galt6, camk2g1*) (Eno et al., 2018; Gazdag et al., 2009; Gerovska and Araúzo-Bravo, 2016; Levine et al., 2000). Of these ten transcripts whose association with germ fate is reported, four have a role in transmembrane transport of small molecules. The proteins Nkain4 and Camk2g1 are ion-transporters mainly localised on the plasma membrane (Gorokhova et al., 2007; Rothschild et al., 2013), while Golga3 and B4galt6 are found on the Golgi apparatus and their functions are linked to nuclear transport and glycolipid biosynthesis respectively (Bentson et al., 2013; Camacho et al., 2012). Moreover, Hook2, Ndel1 and Golga3 display cytoskeleton and structural functions.

In conclusion, DGE between non-transcribing PGCs and somatic cells allowed us to identify protein-coding RNAs composing the early germ plasm. Overall, the fact that only twenty-two transcripts were differentially regulated between PGCs and somatic cells was surprising. However, the described experimental procedure excludes the totality of non-polyadenylated transcripts, which may largely contribute to the germ plasm composition (Tiwari et al., 2017). Nevertheless, further validations are needed in order to verify localisation and functions of germ plasm markers.

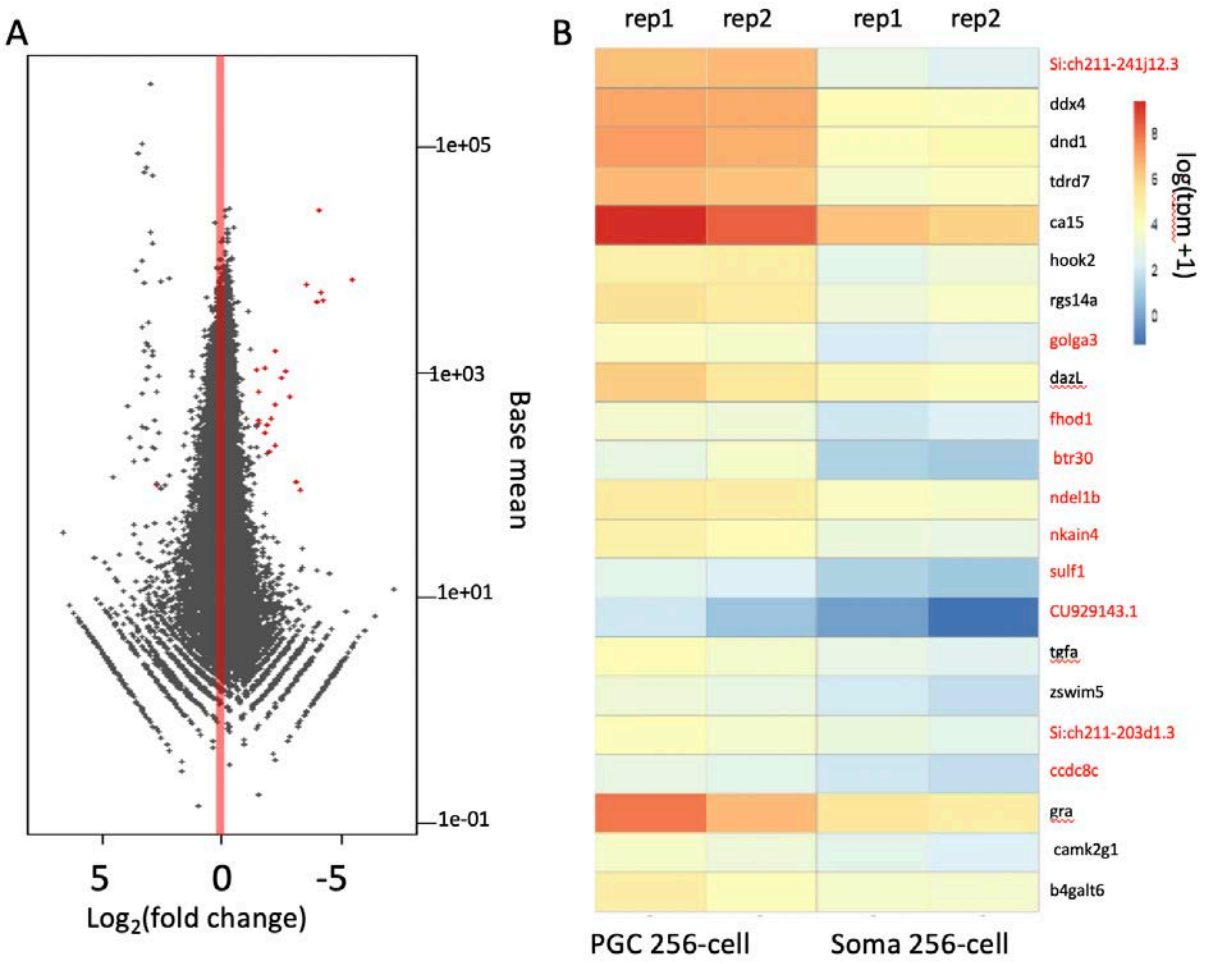

Figure 4. 6: Differentially expressed transcripts between germ plasm-carrying cells and somatic cells before ZGA.
(A) Scatter plot of fold change vs normalised expression value shows differentially expressed genes between PGCs (right) and somatic cells (left). Red-labelled dots represent significantly changing genes among the two groups. (B) Heatmap of normalised gene expression (log_2(tpm+1)) of genes differentially expressed between PGCs and somatic cells. Red names indicate those genes being previously associated with germ fate.

4.2.2.4 A small subset of early zygotic genes is expressed in the PGCs at high stage

In zebrafish, ZGA occurs in two 'waves' of gene expression. The first wave occurs between 2 and 3 hpf (64- to 1k-cell stages) and involves just a very small and specific number of genes. The second and major wave follows the first wave and keeps increasing, reaching the peak of expression at around 6 hpf (sphere stage) (Hadzhiev et al., 2019; Lee et al., 2013).

In many model organisms, the retention of the germ plasm or germ granules was correlated with a temporary inhibition of transcriptional activation at ZGA, which characterises the initial

hours of vertebrate and invertebrate development (Batchelder et al., 1999; Hanyu-Nakamura et al., 2008; Mello et al., 1996). Although the delay in genome activation applies quite widely in the germ line of other organisms, active transcription of *miR-430* already at 512-cell stage (Figure 4.3C) clearly supports transcriptional activity during the first wave in zebrafish.

4.2.2.4.1 Transcripts upregulation is detected in PGCs from 256-cell to high stage

To verify whether *miR-430* represents an isolated exception and to globally characterise the ZGA in germ plasm-carrying cells, the transcriptomes of pre-ZGA (256-cell stage) and post-ZGA (high and dome stages) PGCs were compared. Alongside, the same analysis was performed on the somatic cells in order to control for zygotic genes transcribed in the whole embryo.

Differentially expressed genes between 256-cell and high stages were retrieved as reported before using raw reads as input on the DESeq2 R package. When only reads from GFP-positive cells (PGCs) were taken into account, 152 transcripts were upregulated at high stage while 48 were downregulated (padj < 0.1) (Figure 4.7, bottom). When the analysis is extended to the GFP-negative cells (somatic), 289 transcripts are upregulated and 186 are downregulated at high stage (Figure 4.7A, top).

In order to gain insights into the mechanisms regulating the onset of transcription in germ plasm-carrying cells, I performed gene ontology clustering on differentially expressed genes between two stages in the same cell type. Interestingly, gene ontology clustering of genes upregulated in PGCs at high stage was enriched for germ plasm-associated transcripts (Figure 4.7B). The fact that germ plasm markers such as *dnd1*, *dazl* and *tdrd7* were upregulated in PGCs after ZGA was striking. In fact, this observation could indicate that PGCs already transcribe germ cell-specific genes at high stage. Overall, we report for the first time RNA-seq data from isolated germ plasm-carrying cells and somatic cells in zebrafish during ZGA. Our data allowed us to identify the first transcribed genes in an early zebrafish PGC.

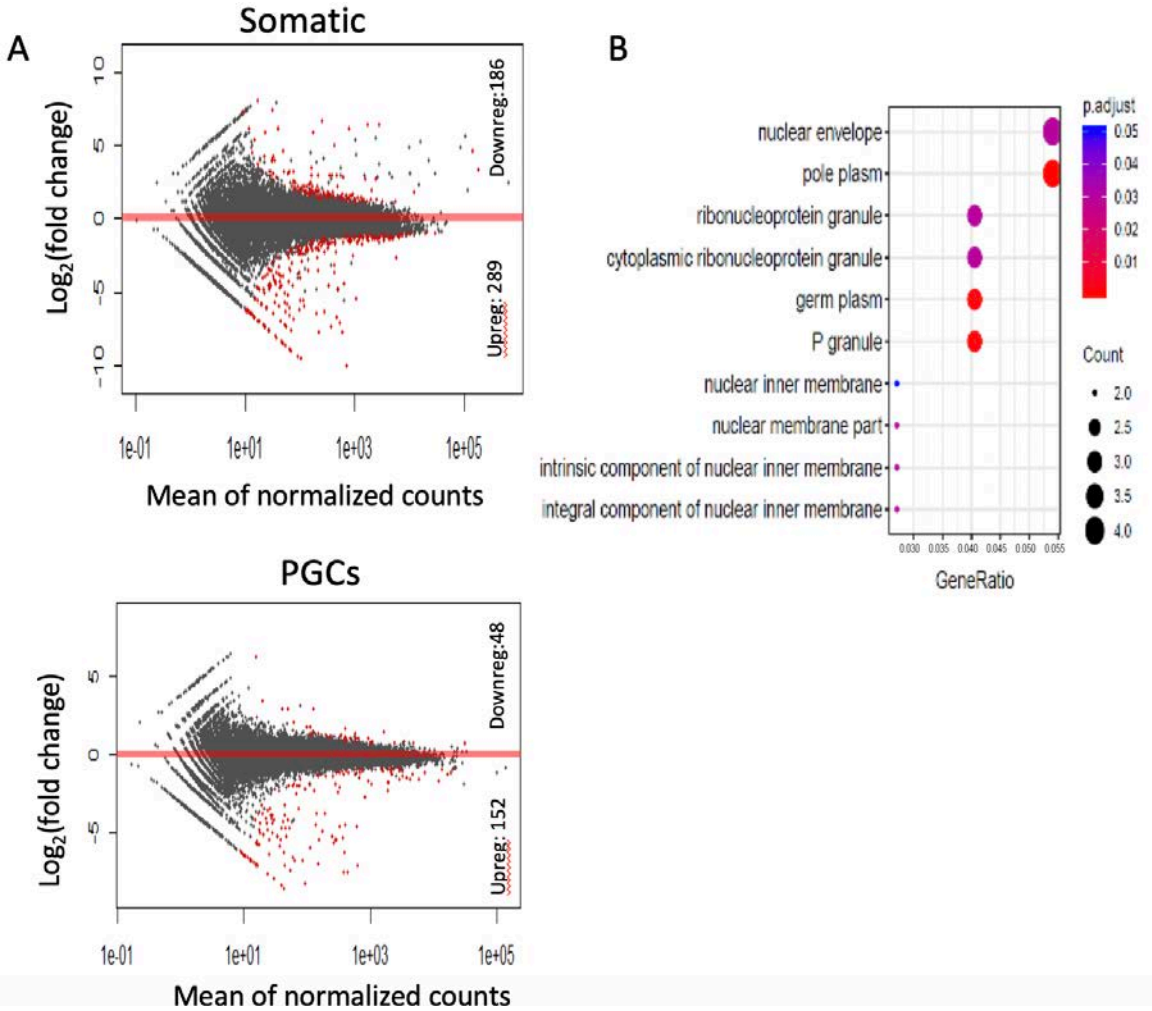

Figure 4. 7: Differentially expressed genes between 256-cell stage and high stage in PGCs.
(A) Log_2(Fold change) vs normalised expression value volcano plots for differentially regulated genes from 256-cell to high in the somatic cells (top) and in the PGCs (bottom). Number of significantly differentially regulated genes is reported. (B) Gene ontology for genes upregulated in PGCs at high stage ($p < 10^{-6}$). Count indicates the number of genes contributing to each GO category. The x axis shows the ratio between the total number of genes that have been tested for the GO term and those which matched the GO term.

4.2.2.4.2 PGCs upregulate zygotic genes

To verify whether the detected upregulated genes were known early zygotic transcripts, I distinguished between, purely zygotic and maternal/zygotic genes by applying a transcripts-per-million (tpm) threshold. Consequently, purely zygotic genes were considered as those with expression lower than 2 tpm at 256-cell stage. In order discard from the analysis genes showing low read counts throughout the experiment, an average tpm threshold for all the samples was set to 2. Of the 135 upregulated genes after ZGA after applying the average tpm threshold, 60

were identified as zygotic without maternal contribution (Figure 4.8A). To confirm our analysis, these results were compared with a previously published RNA-seq dataset (White et al., 2017). Raw read counts were downloaded and normalised to tpm. After overlapping the ENSEMBL IDs of predicted maternally-derived genes from the two datasets based on tpm values, we found a high degree of overlap between the two datasets (over 90% of each dataset was coherently assigned), confirming that our predicted distinction between maternally-deposited and zygotically-expressed transcripts was consistent with independent experiments (Figure 4.8B). In particular, out of 60 zygotic genes found upregulated in the PGCs at high stage, 56 matched predicted zygotic genes from analysis of RNA-seq (White et al., 2017) (data not shown). In conclusion, these data strongly evidence that active transcription occurs in germ plasm-carrying cells as soon as high stage.

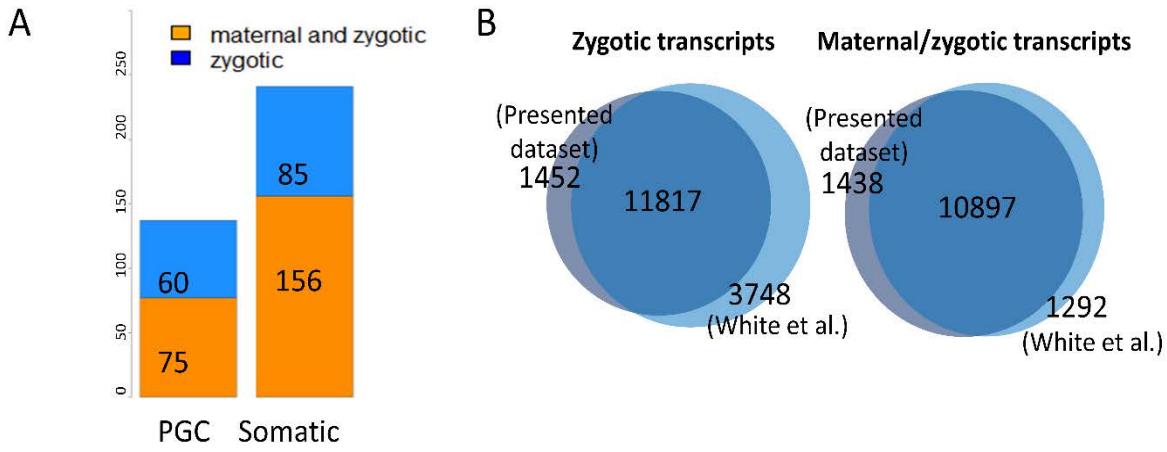

Figure 4. 8: Computational identification of zygotic genes expressed in PGCs and somatic cells after the first wave of ZGA.
(A) Stacked bar chart showing numbers of zygotic or maternal differentially expressed genes after ZGA in PGCs and somatic cells. Predicted maternal genes (orange) have tpm lower than 2 at 256-cell stage. Zygotic genes are in blue. (B) A major overlap is observed when predicted zygotic and maternal/zygotic genes are compared to a previously published dataset (White et al., 2017).

4.2.2.5 Differential gene expression between early PGCs and somatic cells

We then sought to better understand the transcriptomic differences between germ plasm-carrying cells and somatic cells at ZGA. To this end, DGE analysis was performed between

PGCs and somatic cells at each developmental stage and found a small amount of differentially expressed genes in respect to the total transcriptome (over 34000 transcripts), which gradually increases over time (Figure 4.9A). As expected, the increase in differentially expressed genes between PGCs and somatic cells indicates progressive fate acquisition and differentiation, however this effect could be caused by either the two cell types undergoing cellular differentiation or only one. Interestingly, we noted that only 327 genes were differentially regulated between PGCs and somatic cells at dome stage, although PGCs and somatic cell differentially express 4387 and 3098 genes between high and dome stage respectively (Figure 4.9 A, B). These numbers indicate that, while both the cell types experience remarkable changes along the temporal axis, there is a negligible amount of differentially expressed genes when PGCs and somatic cells are compared at the same stage. Therefore, we concluded that germ plasm-carrying cells and somatic cells undergo similar differentiation pathways. As further evidence, GO analyses on genes upregulated in PGCs and somatic cells from high to dome showed large overlap (Figure 4.9C).

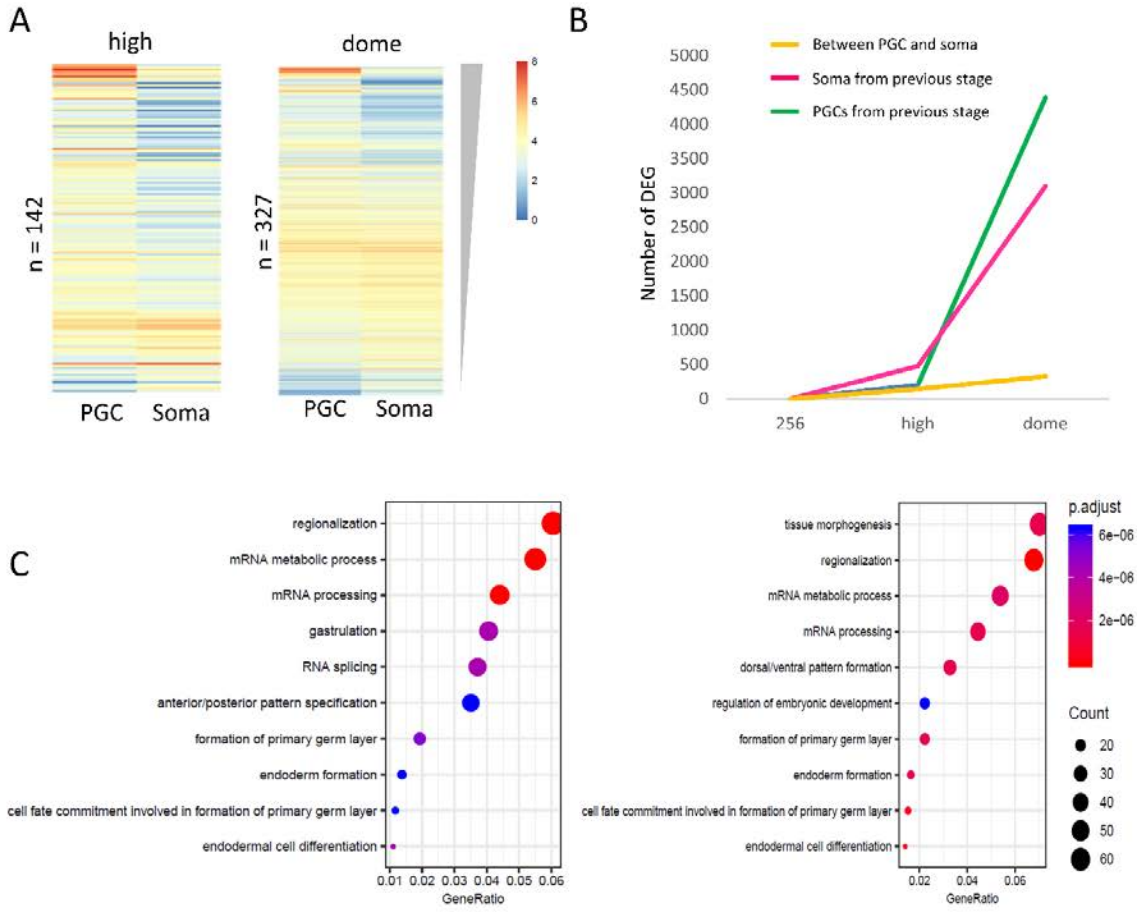

Figure 4. 9: PGCs and somatic transcriptomes are broadly similar during ZGA.
(A) Heatmap of normalised gene expression (log_2(tpm+1)) of genes resulting differentially expressed between PGCs and somatic cells at high and dome stages. (B) Line chart reporting the number of DEG between PGCs and somatic cells (orange) and from previous stages in PGCs (green) and somatic cells (pink). (C) GO analysis for DEG from high to dome in PGCs (left) and somatic cells (right) ($p < 10^{-6}$). Count indicates the number of genes contributing to each GO category. The x axis shows the ratio between the total number of genes that have been tested for the GO term and those which matched the GO term.

Although small, transcriptional variation between PGCs and somatic cells was detected, therefore we aimed to look in more detail at the variation between the two transcriptomes.

In order to study the nature of ZGA in germ plasm-carrying cells, a regression-based approach was used. This is achieved with the R package MaSigPro, which clusters genes with similar expression profiles between experimental groups during a time course experiment. The aim was to identify groups of genes that are differentially regulated between PGCs and somatic cells at ZGA.

MaSigPro was run on PGC and somatic cell datasets at 256-cell, high and dome stage, in order to cover three important phases of ZGA. Normalised read counts were fitted with a quadratic regression model as they cover three time points and then significant genes were discovered by applying a p-value threshold of 0.05. The p-value was calculated by applying the standard Benjamini and Hochberg False Discovery Rate procedure (FDR). Finally, the various clusters between experimental groups were generated by executing a stepwise regression among significantly expressed genes.

When executed, MaSigPro identified nine gene clusters. Globally, the result shows that PGCs and somatic follow similar trend of gene expression during ZGA for almost all the clusters (Figure 4.10). Of the nine identified clusters, only cluster 6 showed a distinct trend for PGCs and somatic cells. As expected, these genes matched putative zygotic genes upregulated in PGCs only from 256-cell to high stage and the majority of putative germ plasm-components were found in this cluster (18 out of 22 genes). Although this number is small when compared to the whole transcriptome, it becomes relevant at ZGA, when just a small set of gene is transcribed (289 genes in the somatic cells, Figure 4.7A). Moreover, it includes 42% of the genes upregulated in PGCs at high stage from the previous time point. Interestingly, many germ plasm-embedded RNAs are found in this class, indicating for potential mechanisms of post-transcriptional regulation. As expected, genes included in this clusters are assigned to germ plasm GO terms (data not shown). Interestingly, we noted that three clusters group genes whose expression is exponentially increasing after 256-cell stage, evidencing for active transcription being achieved in both PGCs and somatic cells (Figure 4.10, squared). Therefore, the data presented strongly suggests that PGCs have the ability to transcribe already at the early phases of development, supporting the occurrence of ZGA in PGCs. In conclusion, we found no evidence for zebrafish PGCs delaying genome activation as observed in other animals

(Hanyu-Nakamura et al., 2008; Mello et al., 1996; Venkatarama et al., 2010) and we highlight that somatic and PGC ZGAs have many similarities.

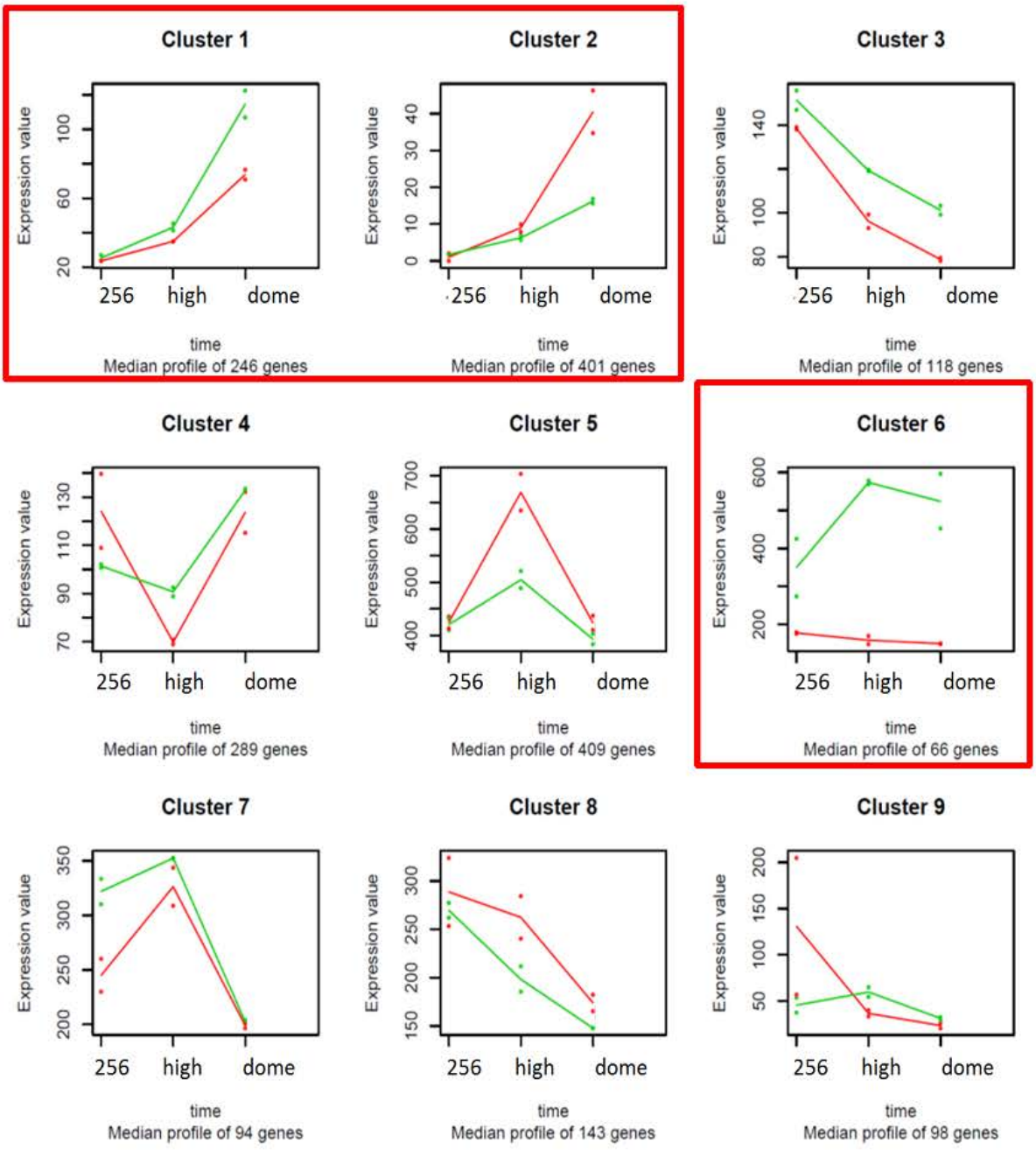

Figure 4. 10: Clusters of normalised gene expression trends among three developmental stages in PGCs and somatic cells.
Gene expression values are plotted for each trend in green (PGCs) and red (somatic cells) and represent the gene read counts normalised by the DESeq2-calculated size factor. Red squared highlight clusters supporting transcriptional activation in the PGCs.

4.2.2.5.1 Estimation of absolute RNA concentration in PGCs and somatic cells

The current sequencing-based methods of transcriptomes profiling can very precisely estimate the relative abundance of a determined transcript in reference to any other. Although this allows reliable analysis of differential gene expression, there is no information about absolute concentration of the transcripts. In fact, sequencing depth and library complexity can greatly affect the overall number of output reads.

Currently, the most common strategy to estimate absolute RNA abundance is via normalisation by internal controls. To this purpose, the External RNA Controls Consortium (ERCC) has developed and optimised a mix of polyadenylated transcripts of known concentration that, when added to any RNA sample prior to cDNA library generation, can be used to verify the lower limit of detection and assessing the platform response to known transcript ratios (Jiang et al., 2011).

We took advantage of this method to quantify the absolute concentration of RNA at three developmental stages in PGCs and somatic cells. Two reasons motivated this approach: 1) confirming that PGCs do not delay ZGA and 2) verifying whether PGCs accumulate more RNA in comparison to the somatic cells. In fact, as mentioned in the introduction, when the absolute amount of RNA was measured in isolated murine PGCs and compared to the somatic cells, significant higher levels of RNA were found (Percharde et al., 2017).

Two different ERCC mixes containing the same 96 transcripts at different ratios were added to the PGCs and to the somatic cells respectively prior to library preparation. After sequencing, raw reads from the ERCC mixes were counted and normalised to tpm as described before (Table 8.2, Appendix). In order to evaluate the degree of correlation between two ERCC mixes, the expected versus the calculated fold-change ratio was plotted (tpm > 1). The correlation ranged from 0.951 to 0.993 (Figure 4.11A), indicating that normalised reads were suitable estimators of transcript concentration.

I then generated a calibration curve in order to convert read counts to transcript concentration and extracted the slope and the intercept that were used to calculate RNA concentration for each transcript (Figure 4.11B).

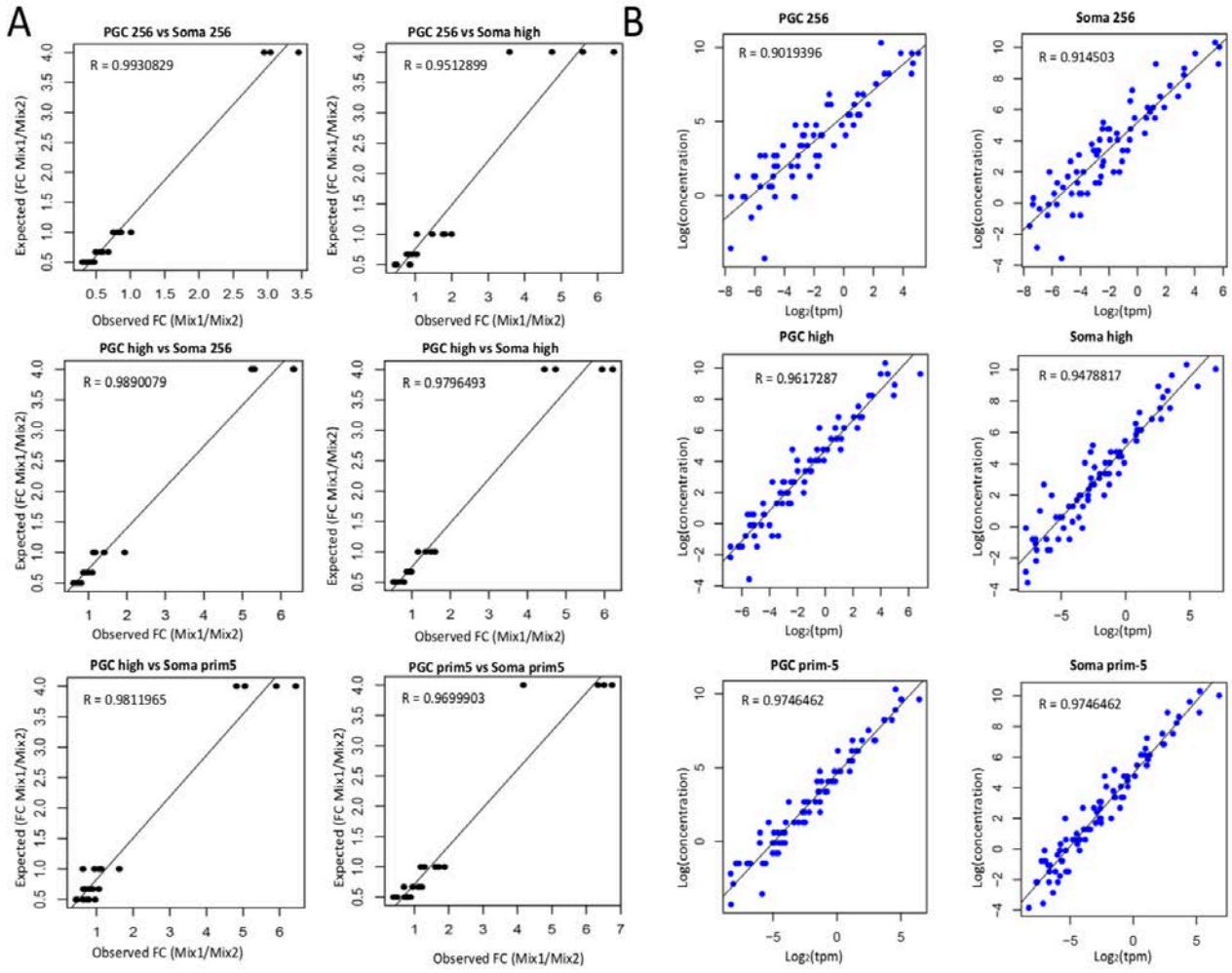

Figure 4. 11: Regression scatter plots between expected vs observed ERCC mix ratios.
(A) The expected vs observed fold change of normalised reads for all the ERCC transcripts passing the threshold was compared at each stage. Threshold tpm >= 1. (B) A reference curve was built for correlation between absolute RNA concentration and read counts. Best-fitting line was drawn and regression score calculated in R.

When the median of the RNA concentration was plotted for each sample, no significant difference was detected. As the RNA-seq libraries were generated from exactly two hundred cells, comparable RNA amounts were expected. However, when only the genes upregulated in the PGCs from 256-cell to high stage upon DGE analysis were considered, a significant increase of RNA absolute concentration was observed (Figure 4.12). In our opinion, the

increase of the absolute amount in RNA species after the first wave of ZGA is an indication of transcriptional activity. Importantly, a statistically significant increase for those genes was detected in the somatic cells as well, although less prominent. This prompted us to deepen our analysis of the transcriptional dynamics of the PGCs and the somatic cells at ZGA.

In conclusion, we have provided strong evidence that gene transcription occurs in the zebrafish PGCs as early as high stage by multiple and independent analyses. In combination with detection of *miR-430* transcription, we conclude that PGCs do not delay the canonical ZGA. Instead, when somatic and PGC ZGAs are compared, many similarities are found, suggesting that a PGC-specific ZGA may not occur. In order to better understand mechanisms of gene activation, we have explored pre- and post-transcriptional events leading to the acquisition of a PGC-specific transcriptome.

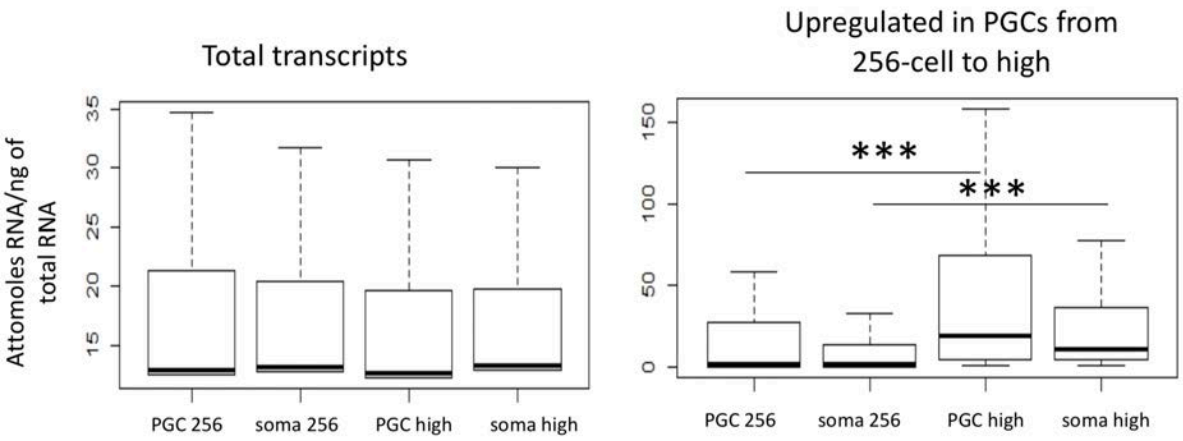

Figure 4. 12: Absolute RNA concentrations upon transformation of sequencing reads into attomoles of RNA/nanograms of total RNA.
Boxplots report the predicted average and median concentration of each transcript included in the analysis. Left panel shows concentration for all the transcripts. Right panel shows concentration of transcripts resulting significantly upregulated in the PGCs only from 256-cell to high stage (first wave of ZGA). p-value is shown as *** if < 0.0005, based on t-test.

4.2.3 Pre-migratory PGCs and somatic cells have similar open chromatin profiles

So far, sequencing and imaging analyses suggested that ZGA is broadly similar between PGCs and somatic cells. Alongside, weak signs of fate acquisition can be assumed by the gradual

increase of differential expressed genes between the two cell types over time. However, the mechanisms and processes leading to acquisition of cellular identity and germ fate can be multiple and complex. In our opinion, we could face two possibilities. The first is that the germ plasm already at the very early stages influences and drives transcription of the cell in which it is embedded through germ plasm-loaded transcriptional regulators, resulting in a PGC-specific ZGA. The second possibility, instead, would be that the same set of genes is equally activated throughout the embryo but the cytoplasmic germ plasm may function as an RNA-processing site in which certain zygotic transcripts are loaded. This possibility is supported by the fact that genes significantly upregulated in the PGCs from 256-cell to high stage are upregulated in the somatic cells as well, however discarded by the DGE analysis because not statistically significant. Moreover, one could think that, alongside genome activation, intense RNA degradation takes place at these stages with the primary aim of clearing maternal transcripts as quickly as possible. It is known that several maternal transcripts contained in the germ plasm are protected from the degradation by *miR-430* (Mishima et al., 2006). In this scenario, one could hypothesise that *miR-430* targets would be degraded everywhere in the embryo (somatic cells) but protected in the germ plasm-carrying cells, therefore were upregulated in the PGCs.
To verify which of the two scenarios would induce differential gene expression in PGCs and somatic cells, we first interrogated the chromatin state of the early transcribing PGCs. Asking whether differential transcription occurred between PGCs and somatic cells at ZGA, we assumed that diverse transcriptional outputs should be detected on the chromatin level. In fact, it is nowadays proven that active or inactive cis-regulatory elements show diverse chromatin organisation (Margueron and Reinberg, 2010; Tsompana and Buck, 2014). I therefore performed ATAC-seq on isolated PGCs and somatic cells at high stage and compared the chromatin profiles of the two cell types. ATAC-seq is a recently-developed technique to

identify regions of accessible chromatin and nucleosome positioning genome wide. The principle and technical approach are widely discussed in the next chapter and in Methods.

Surprisingly, the comparison between PGC and somatic cell open chromatin profiles showed that the two cell types are broadly similar at high stage. The correlation between PGCs and somatic cells was very high (r= 0.9809) (Figure 4.13A) and I was unable to find instances of significant differences when regions were manually inspected. In fact, even for genes that were differentially transcribed between PGCs and somatic cells at high stage, the ATAC profiles for early stages (high and dome) were overall similar (Figure 4.13B).

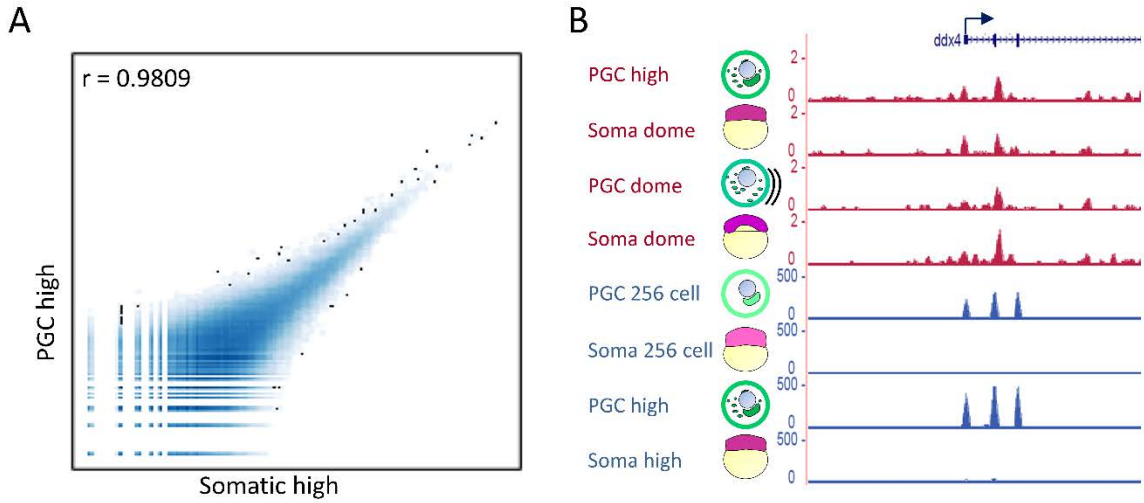

Figure 4. 13: ATAC-seq of isolated PGCs and somatic cells at high stage shows global similarities between the two open chromatin profiles.
(A) Correlation scatter plot of identified open chromatin regions between PGCs and somatic cells at high stage. Each dot corresponds to an ATAC-peak shared between the two cell types and the fold changes over the background were compared. (B) Genome browser view of ATAC- (magenta) and RNA-seq (blue) profiles of *ddx4* for PGCs and somatic cells at 256-cell, high and dome stages.

With the help of two collaborators at the Imperial College London (Dr. Piotr Balwierz and Dr. Boris Lenhard), we looked more comprehensively at the degree of openness on promoters and enhancers for subclasses of genes. To do this, the signal-to-background fold change of called ATAC peaks was centred to either the TSS for promoters or the centre of the peak for distal elements. Then, the average signal was plotted for genes upregulated in the PGCs from 256-cell to high stage and for genes which GO term was associated to germ plasm. Even with this

approach, PGCs and somatic cells appeared almost identical over all the gene classes (Figure 4.14). The lack of evidence for differential chromatin accessibility between early PGCs and somatic cells was an indication that the detected differential gene expression was caused by post-transcriptional regulation, directing us to hypothesise for the existence of alternative post-transcriptional mechanisms exploited by the PGCs to initiate lineage specification.

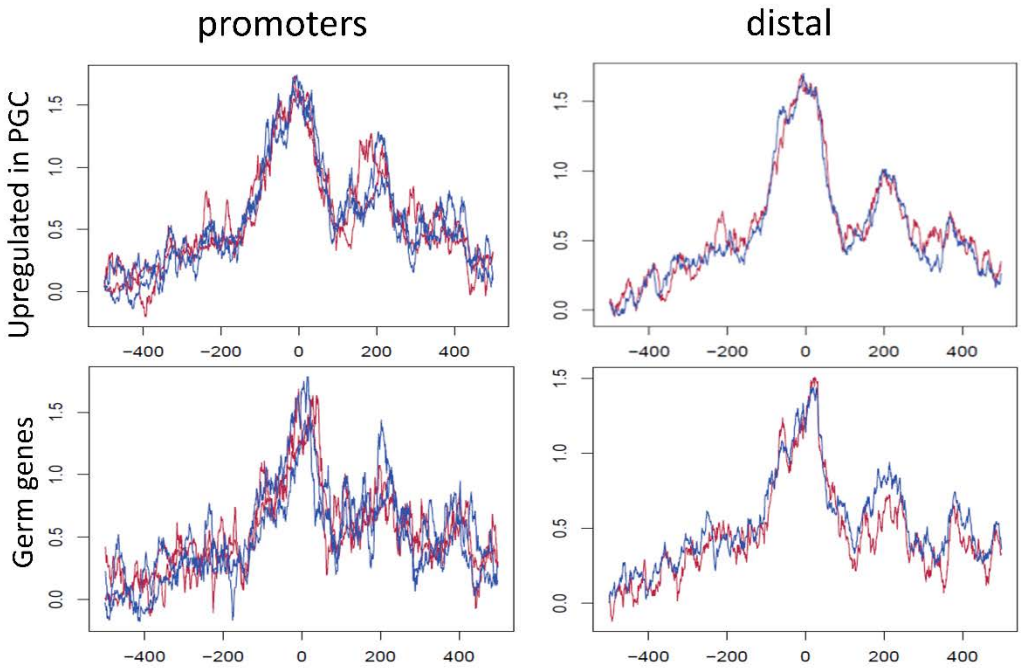

Figure 4. 14: Aggregation plots of ATAC-seq signal on promoter and enhancers for gene subclasses at high stage in PGCs and somatic cells.
ATAC signal (fold change over the background) was plotted in a 1000 bp window around the TSS (left) and the peak centre of non-promoter-associated ATAC peaks (right). ATAC signal from PGCs is in blue and from the somatic cells is in red. Genes upregulated in PGCs were obtained via DESeq ($log_2FC < 0$, padj < 0.1). Germ genes were retrieved under the gene ontology accession GO:0007281.

4.2.4 Germ plasm-localised genes are transcribed in the whole embryo

The fact that no significant PGC-epigenetic signature was detected through ATAC-seq, yet differential transcript regulation occurred was striking. Consequently, we hypothesised that, although transcription was not specifically regulated in the PGCs, post-transcriptional events were taking place to initiate germ cell specification. One possibility that we explored was that

differential transcript degradation occurs upon zygotic transcription of germ cell genes in the whole embryo (Table 4.3).

To verify whether somatic cells transcribe germ plasm-localised genes during early phases of development, qPCR on isolated somatic cells was performed. In fact, to avoid contamination, PGCs were removed via FACS prior to RNA extraction and cDNA generation. Because maternal RNA is in excess during early stages, enrichment for de-novo transcripts was achieved through nuclear isolation. RNA was extracted from both cytoplasmic and nuclear preparations and cDNA was generated. qPCR was performed on selected candidate genes which were significantly upregulated in the PGCs from 256-cell to high stage in the RNA-seq data. As a control, embryos were treated with triptolide to inhibit transcriptional activation.

Table 4.3: qPCR primers used to verify expression of germ plasm-localised genes in the somatic cells.

Name	Forward Sequence	Reverse Sequence
Ddx4	AGGATCCTTCAAGAGCGATGA	GGTATTGAAGAAGCTCGCACA
Myl12.1	CCAAGGTAAAGCTGCACTGT	CCACGAGAGCCCTGAACTTA
Nanos3	AGACTGAGGCCGTGTACACCTCTCAC TACT	GAGCAGTAGTTCTTGTCCACCATCG
Tdrd7	TCTACCCAGCGGAAGCTTTA	CTGGTGTCCCACTGGTCTTT
Tdrd9	GGTCTCCGATCCGTAATCAG	AGCCTCCATCTCATCAAAGC
Dnd1	TTCACTCTTCATGGCTCGTG	GTCAACAGACTCGGCTCTCC
Bact	TCTCTCTGTTGGCTTTGGGA	CCTGACCCTCAAATACCCCA

After double normalisation against the non-transcribing controls (triptolide) and the housekeeping gene *beta-actin*, the relative fold changes of both the cytoplasmic and nuclear

fractions were compared. Interestingly, we noted a consistent trend of relative fold change increase in the nuclei but not in the cytosol (Figure 4.15). Out of six tested germ plasm-genes, five were significantly upregulated in the nuclei of embryos at high stage, with an opposite or non-significant change in the cytoplasm (p-value < 0.05, Wilcoxon-test). These results suggested that some germ plasm-associated genes are actively transcribed in the non-germ plasm-carrying cells when transitioning from a non-transcribing stage (256-cell) to a transcribing stage (high).

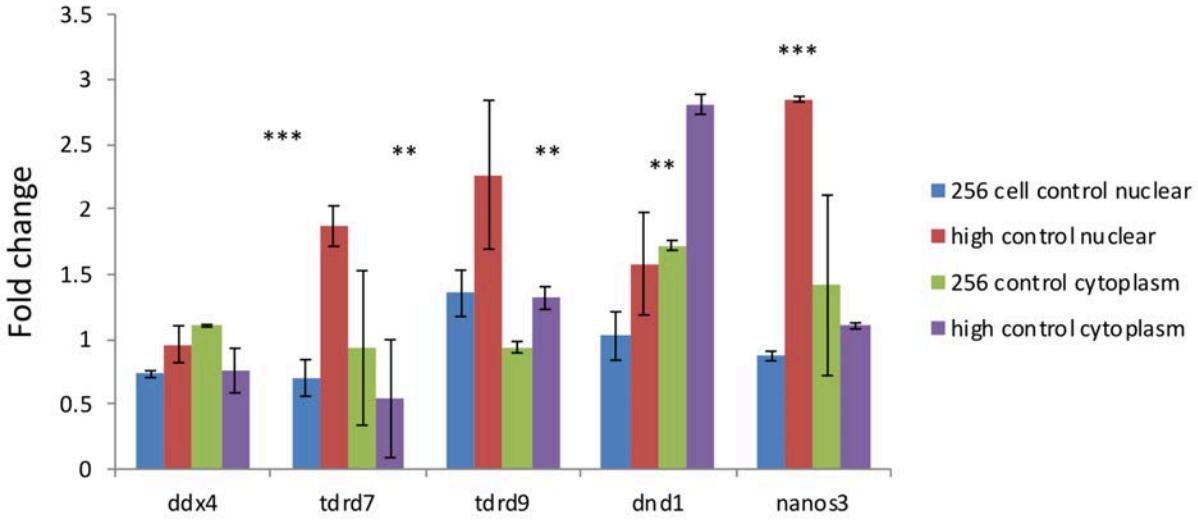

Figure 4. 15: Relative expression analysis of germ cell genes in the somatic cells at 256-cell and high stage.
Two fractions (nuclear and cytoplasmic) were used for each stage and treatment. Gene expression fold change was calculated upon normalisation for the housekeeping gene *beta-actin* and for the triptolide treated samples (double delta). Bar represent standard error among three biological replicates. p-value is represented as ** if < 0.005 or *** if < 0.0005 based on Wilcoxon test.

In order to discriminate between an early specific transcription and a selective degradation in the germ plasm-carrying cells, we sought to test whether we could visually localise those transcripts highly upregulated in PGCs after ZGA. Our aim was to check if non-PGCs also were transcribing these genes. For this purpose, antisense RNA probes were designed and synthesised to target *dazl* RNAs, which was one of the highest differentially expressed genes from 256-cell to high stage in PGCs. As these transcripts are maternally provided and spread

across the embryo before maternal transcript degradation starts, we aimed to target intronic sequences of each transcript. Although this does not exclude the eventuality of unspliced inherited maternal RNAs, we increased the chances to detect zygotic transcription.

As, most of the splicing events occur immediately after the transcript is generated, and the lifespan of the introns is usually relatively short, designing probes spanning both exons and introns can be beneficial (Figure 4.16A). ISH was performed on pre-MBT (256-cell stage) and post-MBT (high stage) embryos in order to discriminate between maternal and zygotic transcripts. Additionally, the post-MBT stage was treated with 1 μM triptolide to inhibit gene transcription as a control. Interestingly, *dazl* appeared to be broadly expressed in the whole embryo, not precluded to germ plasm-carrying cells as instead expected (Figure 4.16B). The fact that no nuclear enrichment was detected in the pre-MBT stage and in the triptolide treated post-MBT stage, strongly suggested the occurrence of zygotic transcription of a germ cell-specific transcript in somatic cells.

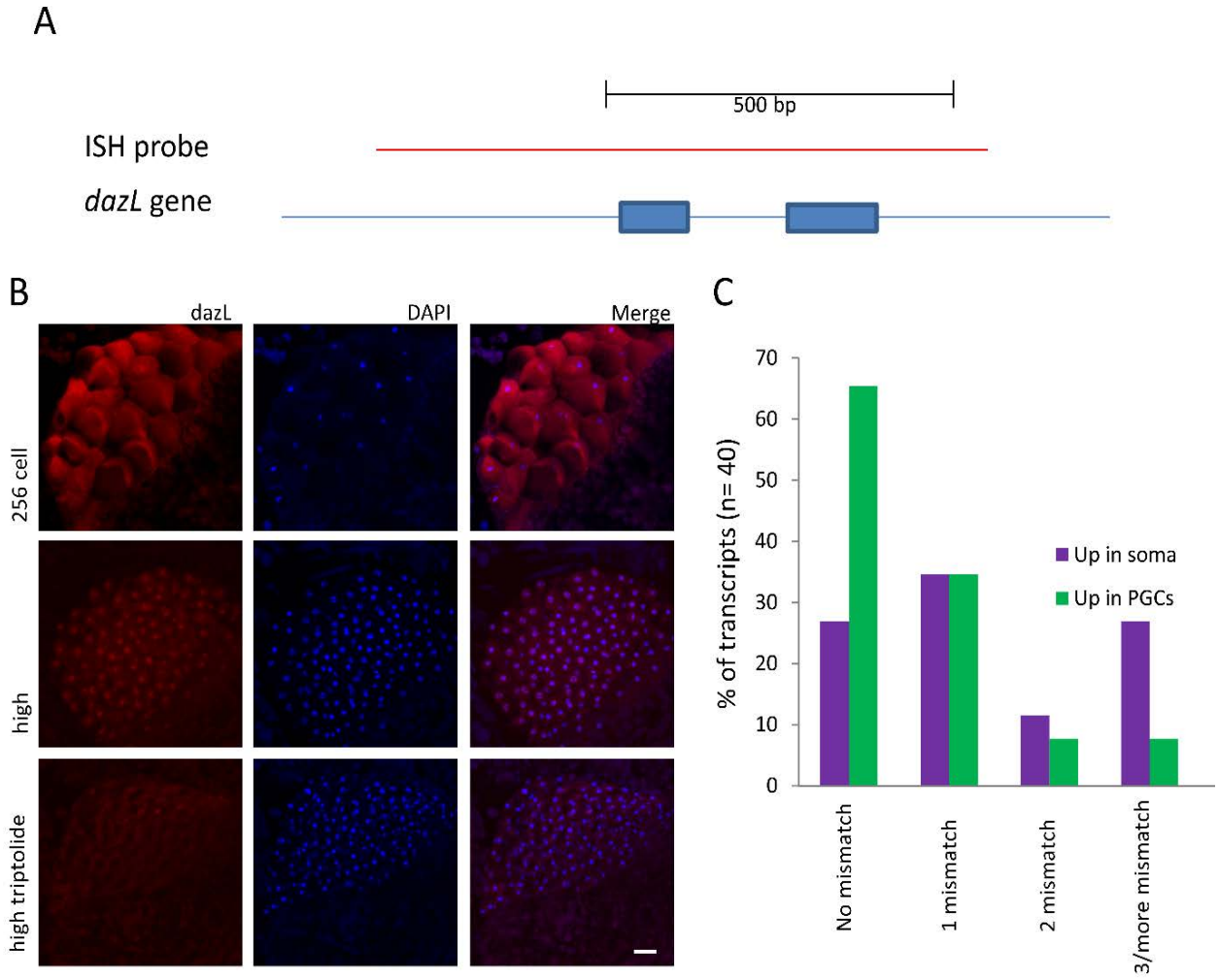

Figure 4. 16: The germ cell-specific gene *dazl* is transcribed the zebrafish blastomeres.
(A) RNA probes for in-situ hybridization were designed in order to target nascent dazl transcripts. (B) Fluorescent in-situ hybridization shows localised nuclear signal after ZGA (high stage) in all the nuclei, which is lost when transcription block is applied (triptolide). (C) Percentage of transcripts with a matching *miR-430* target sites on the 3'UTR. Genes are divided in those with no mismatches (perfect match), 1, 2 or more mismatches. Genes upregulated in PGCs from 256-cell to high stage are in green. In purple the genes upregulated in the somatic cells from 256-cell to high stage are shown.

Finally, to further support our model, I performed in-silico analysis of *miR-430* target sites on two subclasses of genes: upregulated in the PGCs and upregulated in the somatic cells from 256-cell to high stages. If the upregulation in the PGCs was due to differential transcript protection, we assumed that these genes should be efficiently targeted by *miR-430* in the somatic cells. Importantly, genes upregulated in PGCs from 256-cell to high stage tended to be better targets for *miR-430*, compared to genes upregulated in the somatic cells from 256-cell to

high stage (Figure 4.16C). In fact, I counted the number of *miR-430* target sequences in the 3'UTR of the two gene subsets and found that overall, over 60% of genes upregulated in the PGCs after ZGA have at least one perfect match with the *miR-430* target sequence. If sequences with one mismatch are also included, over 90% of the analysed genes is a candidate *miR-430* target. In contrast, genes upregulated in the somatic cells were predicted to be less likely targeted by the *miR-430* (Figure 4.16C).

4.3 Discussion

In this chapter, a combination of imaging and NGS techniques were utilised to study the onset of ZGA in the germ plasm-carrying cells. Importantly, we show several lines of evidence suggesting that PGCs and somatic cells are not epigenetically nor transcriptionally distinct before PGC migration starts (dome stage). This is in contrast with most of the germ plasm-dependent organisms, which rely on a delayed transcriptional activation of the PGCs.

Here, I used visual transcript detection to follow the early wave of ZGA in the PGCs, showing that PGCs undergo transcriptional activation alongside the rest of the embryo. It is noteworthy that our results confirm that transcriptional activation in zebrafish is not delayed. In fact, this was previously suggested upon immunostaining of active RNA polymerase II in the PGCs (Knaut et al., 2000), however, direct evidence for transcriptional activation was missing so far. These results support a model according to which no epigenetic nor transcriptional difference exist between a germ plasm-carrying cell and a somatic cell during ZGA. This is further confirmed by immunohistochemistry (IHC) analysis of H3K4me3 and H3K27me3, which indicates similar epigenetic dynamics in both PGCs and somatic cells prior to ZGA, while disproves the existence of epigenetic inheritance. Nevertheless, recent work has evidenced that signature of epigenetic inheritance in form of histone modifications exist (Liu et al., 2016; Zhang et al., 2016). However, these modifications usually characterise the very beginning of embryogenesis and are lost upon differentiation and development (Dahl et al., 2016), therefore probably required for a short period of time during the early phases of embryogenesis. Importantly, our results provide insights into dynamics of early histone modifications in the developing germ line, however showing no significant difference in the global levels of H3K27me3 and H3K4me3 between germ plasm-carrying cells and somatic cells. Together with RNA-seq and *miR-430* expression data, we interpret these results as indication that a germ

plasm-carrying cell is not epigenetically programmed towards the germ fate prior and after ZGA in zebrafish.

To fully characterise the ZGA in PGCs and to provide further details of the transcriptional activity of the early germ line, RNA-seq of three developmental was performed. Comparison of gene expression between early stages spanning ZGA allowed us to report the earliest expressed genes in the PGCs. Surprisingly, these seem to be mainly shared with the somatic cells, suggesting that ZGA is overall similar between PGCs and the rest of the embryo. Additionally, by including the transcriptionally-silent 256-cell stage in the analysis, we were able to identify new potential germ plasm markers. Of the differentially regulated genes between PGCs and somatic cells at pre-ZGA, more than half were known germ plasm markers. This observation further confirmed that the isolation procedure allowed us to efficiently separate PGCs from the rest of the embryo. To date, 11 of the identified transcripts upregulated in the transcriptionally-silent PGCs have not been associated to the germ plasm, making these interesting novel candidate markers of the germ granules.

It has to be acknowledged that differential gene expression analysis between highly-similar samples is challenging (Lamarre et al., 2018) and the statistical requirement to identify mildly upregulated genes over the background of maternal transcripts is usually high. On the other hand, due to the elevated costs of library preparation and sequencing, the experiment was designed with two replicates, exposing the analysis to common drawbacks such as inapplicability of identifying outliers by Cook's distance. Nonetheless, DESeq2 models the distribution of the read counts with the negative binomial distribution, which allows to relatively accurately assume mean or variance for samples with small number of replicates (Anders and Huber, 2010). For the reasons described above, the RNA-seq analysis was performed under different stringency conditions (log_2FC threshold = 1 or 2 and padj < 0.1 or

0.01). Although the number of significantly up and downregulated genes varied, the observation that transcription occurs in both PGCs and somatic cells stands.

Even though the discussed limitations of the method may have reduced our ability in fully identify differential expressed genes, we confirmed the reliability of our datasets by detecting differentially expressed PGC-specific transcripts. Even when the fold change and p-value thresholds were increased, the number of differentially regulated genes remained low in respect to the total transcriptome, indicating that a very small amount of RNAs contributes to induce PGC migration in zebrafish. In line with this observation, it was shown that transcription is not necessary for early PGC mitosis and initial migratory movements, however, it is crucial during epiboly (Blaser et al., 2005; Kane et al., 1996). Indeed, even nucleus-lacking PGCs undergo symmetric division and attempt migration (Knaut et al., 2000).

Interestingly, although few zygotic transcripts were differentially regulated between PGCs and somatic cells, the open chromatin profile did not show remarkable differences both globally and locally. In the context of epigenetic regulation, comparable chromatin profiles between PGCs and somatic cells suggest similar transcriptional output. The interpretation of the ATAC- and RNA-seq results led us to hypothesis that PGC-specific transcription does not occur at those time points. Instead, the mild detected differential gene expression was exclusive consequence of germ plasm-dependent post-transcriptional regulation. This possibility assumes that zygotic transcripts are differentially retained in the PGCs, while degraded in the somatic cells. Notably, the fact that the germ plasm protects transcripts from physiological degradation of maternal RNAs after ZGA was already described (Mishima et al., 2006). However, to our knowledge, the evidence that germ plasm-located transcripts are equally expressed throughout the embryo and selectively retained within the germ plasm was missing. This was concluded based on relative gene expression measurements in the somatic cells,

detection of nuclear *dazl* signal in the whole embryo and in-silico prediction of *miR-430* target sites. Altogether, our results support a model where equally-transcribing cells are directed towards different fates by subcellular germ plasm granules. Interestingly, according to this model, it is not important or decided which blastomere inherits the germ plasm, but the intrinsic properties of the germ plasm will affect the fate of the blastomere.

5 PGCs gain epigenetic and transcriptional identity after migration and retain pluripotent features

5.1 Introduction

In mouse, following activation of *BMP4*, *BLIMP1* and *PRDM1* at E6.5, PGCs activate a specific transcriptional program resulting in migratory behaviour and epigenetic reprogramming. In fact, migrating germ cells undergo intensive DNA demethylation and acquisition of promoter-associated poised histone marks. The epigenetic reprogramming continues throughout embryogenesis and requires reshaping the methylome and chromatin landscape (Hill et al., 2018).

Although several recent studies have shed light on the epigenetic reprogramming of the developing germ line in mammals, it is still unknown what mechanisms drive PGC formation in a germ plasm-dependent animal. The murine BMP4/BLIMP1-mediated transcriptional activation in PGCs does not seem to play a role in zebrafish. In zebrafish, mutation of the *BLIMP1* homologue *u-boot* causes truncation of the body axis and head defects, however, no evident effect on the germ line was reported (Baxendale et al., 2004). On the other hand, *nanog* was shown to be expressed in the cytoplasm of post-migratory PGCs and its inhibition caused over-proliferation and aberrant localisation of PGCs (Wang et al., 2016). In contrast, conditional *nanog* knockdown mice show apoptotic germ cell death during migration (Yamaguchi et al., 2009), suggesting two diverse roles for *nanog* in presence or absence of the germ plasm. Therefore, it remains unknown whether epigenetic reprogramming is achieved in zebrafish migratory PGCs.

Organisms in which epigenesis is utilised to form the germ line need prompt establishment of epigenetic landscapes in order to form the germ cells and to initiate the typical germ line pathways. Consequently, it was postulated that epigenetic reprogramming in the early mammalian PGC may serve as a replacement for the germ plasm (Lesch and Page, 2014). To

this end, organisms relying on germ plasm may undergo a later epigenetic reprogramming of the germ line, while germ factors support the initial phases of PGC formation. On the other hand, the reported transcriptome profile of the post-migratory PGCs showed intense variation between the somatic and the post-migratory germ cells, which could be hardly explained by post-transcriptional regulation only.

To study the relationship between germ plasm and epigenetic reprogramming, we followed PGC development at migratory and gonadal stages in a developing zebrafish. We interrogated the transcriptome and the epigenome in stages corresponding to reprogramming PGCs in human and mouse. As we noted a remarkable change in germ plasm localisation during migration of PGCs in zebrafish, we attempted to provide insights on the function of the germ plasm and link its localisation to eventual epigenetic and transcriptional changes by temporal analysis of PGC epigenome and transcriptome. To do this, both RNA-seq and ATAC-seq were performed on migratory and post-migratory PGCs, in order to understand the relationship between chromatin and transcriptional state and what role the germ plasm could play in the acquisition of germ line fate.

5.2 Results

5.2.1 Overall characterisation of pre- and post-migratory PGC transcriptomes

Transcriptome analyses of early PGCs and somatic cells showed a weak but gradual branching of the two cell types over development. To further understand PGC development in zebrafish, RNA-seq was performed on PGCs collected a 12.5hpf (10-somites stage) and 24hpf (prim-5), which are referred as late PGCs. These two developmental stages correspond to migrating and gonadal PGCs respectively and coincide with perinuclear localisation of the germ plasm. Isolation of PGCs and somatic cells (non-PGCs) were performed as in 4.2.2.1.

Global variance in gene expression between stages and replicates was analysed using PCA. Interestingly, while early PGCs and somatic cells show minimal differences, the variance at late stages is higher. We observed a positive shift towards principle component 1 (PC1) following developmental time and an elevated separation along both the PC1 and PC2 axes between PGCs and somatic cells at prim-5 stage (Figure 5.1A). Expectedly, the global transcriptome of the late PGCs was more similar to the early somatic cells (high and dome stages) in comparison to the late differentiated somatic cells (prim-5 stage), as indication of pluripotency. When the number of significantly differential expressed genes from one stage to the previous was checked, an exponential increase between high and 10-somites stages for both the cell types was noted (Figure 5.1B). When comparing PGCs and somatic cells, 9847 and 9094 genes were found differentially expressed at dome and 10-somites stages, respectively. Interestingly, the number of differentially expressed genes between PGCs and somatic cells (orange line) greatly increases after dome stage, indicating segregating differentiation pathways.

In conclusion, transcriptome analysis of PGCs over time infers that germ fate is mainly acquired during migration. Although few PGC-specific transcripts are detected before

migration, a remarkable increase in the number of differentially expressed genes from the somatic cells is observed in migratory (10-somites) and gonadal (prim-5) PGCs.

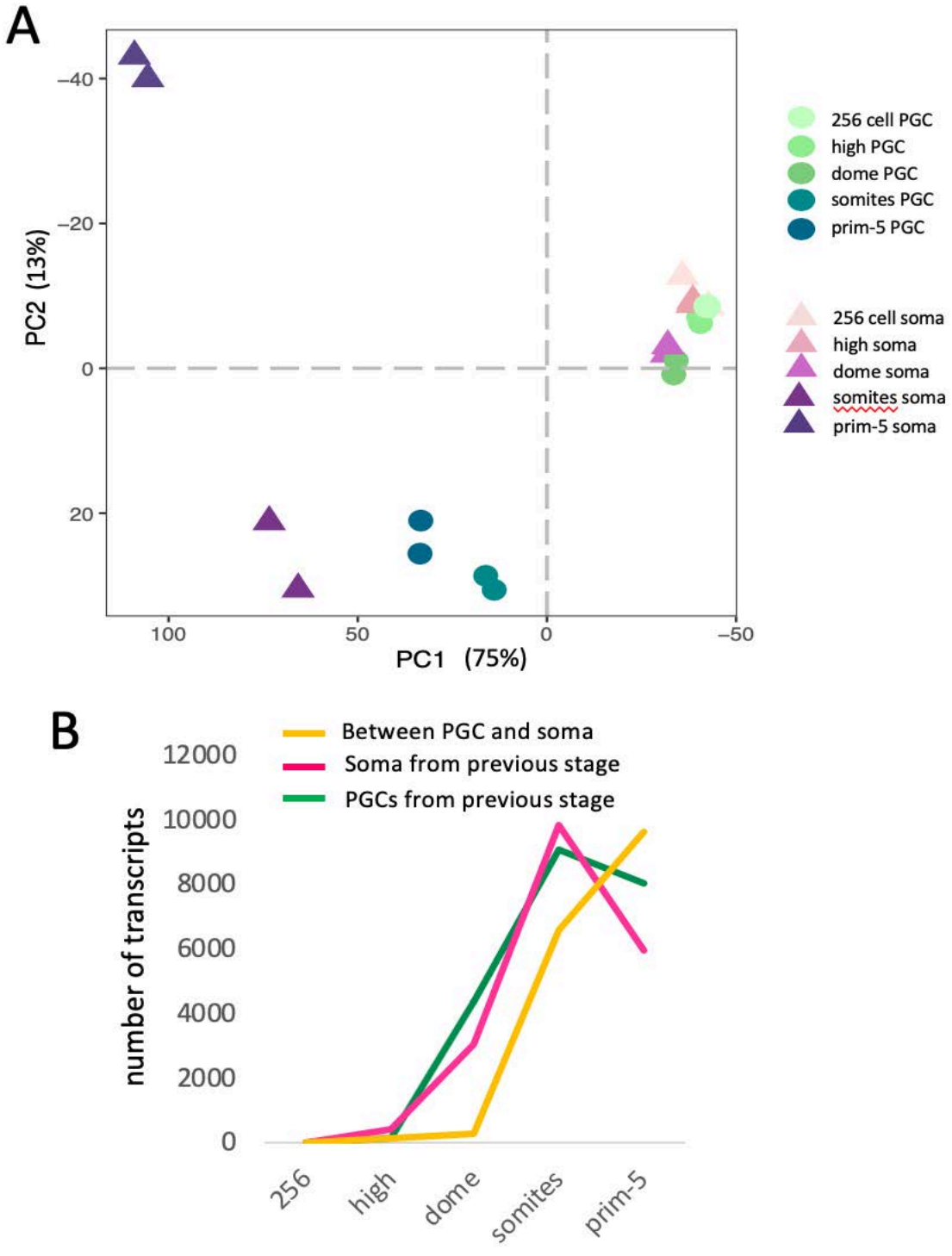

Figure 5. 1: Global transcriptomic analysis of pre- and post-migrating PGCs.
(A) PCA plot shows clustering of early stages (256-cell, high and dome) and separation of late germ and somatic transcriptomes (10-somites and prim-5). PGCs and somatic cells are shown as circles and triangles respectively (B) Line-chart for gene counts. Differentially regulated genes from previous stage are in pink (somatic) or dark green (PGCs). Yellow line reports the number of genes differentially expressed between PGCs and somatic cells at each stage.

5.2.2 Post-migratory PGCs downregulate somatic and developmental genes

Following this, we next investigated which gene classes contribute to the elevated divergence in the transcriptomes of late PGCs and somatic cells (10-somites and prim-5). To this end, GO analysis was used to analyse genes that were significantly upregulated in the PGCs or upregulated in the somatic cells for prim-5 embryos (Table 8.3, 8.4, Appendix). As expected, the vast majority of GO classes were associated with biological processes corresponding to tissue development and morphogenesis for genes upregulated in the somatic cells when compared to the PGCs (Figure 5.2, left). Interestingly, muscle development was greatly over-represented, probably due to the fact that muscle cells comprise the vast majority of GFP-negative cells at prim-5 stage and that muscle growth requires activation of several molecular pathways (Egerman and Glass, 2014). On the other hand, upregulated genes in the PGCs were classified as germ plasm factors, gametogenesis, regulation of nuclear processes and cytoskeleton as expected in migrating cell types (Figure 5.2, right). Interestingly, an under-representation of factors involved in cellular division confirmed that these cells are barely proliferative as described previously (Houston and King, 2000; Tang et al., 2016).

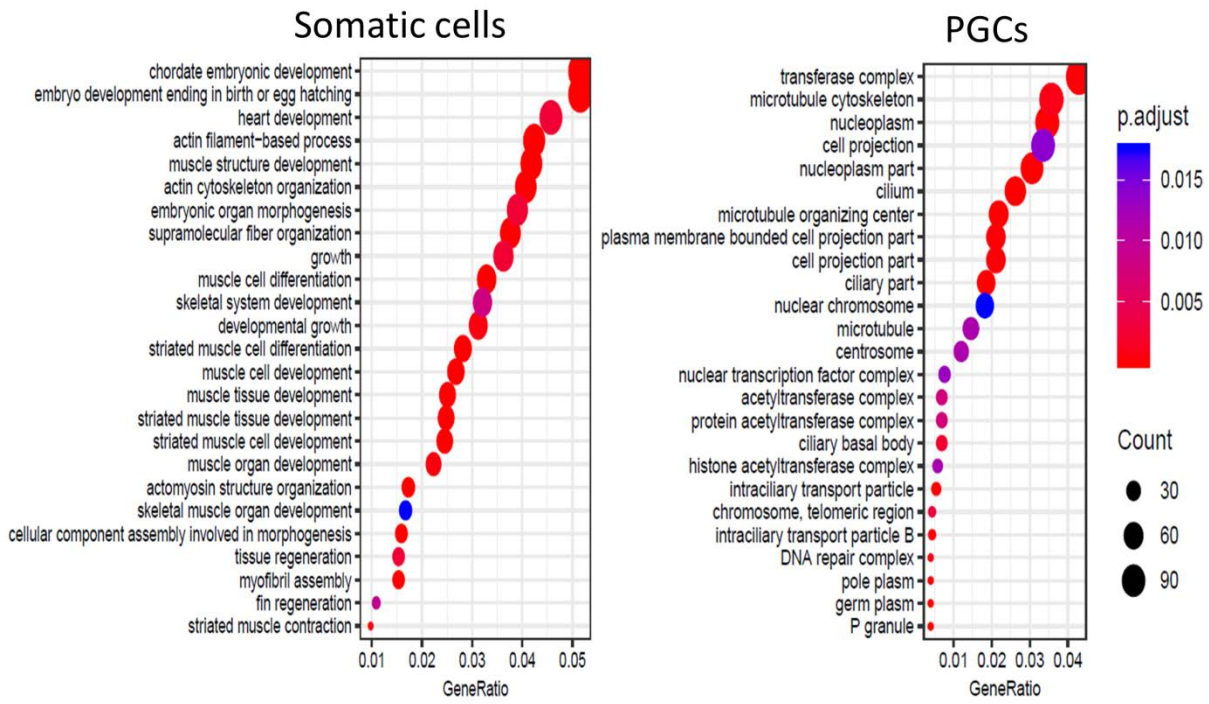

Figure 5. 2: GO analysis for biological processes upon differential gene expression analysis between PGCs and somatic cells at prim-5 stage.
GO was performed on genes upregulated in the somatic cells (left) and in the PGCs (right) at prim-5 stage (post-migration). p-adjusted < 0.005. Count indicated the number of genes contributing to each GO category.

Notably, PGCs overexpress chromatin re-modellers. Among these, we found enzymes involved in histone packaging and DNA methylation: *dnmt3bb.1* is upregulated over somatic cells, while *tet3* is downregulated. Concomitantly, histone regulation was evidenced by upregulation of arginine methyltransferases (*prmt6*), lysine demethylases (*kdm7* and *kdm8*) and bromodomain-containing proteins (*brdt* and *brd9*) (data not shown). Among these genes, upregulation of genes involved in histone acetylation could indicate epigenetic regulation upon chromatin relaxation (Eberharter and Becker, 2002). Unsupervised hierarchical heatmap of transcripts associated with histone acetylation (GO:0016573) evidenced a general tendency of PGCs to upregulate these genes (Figure 5.3). One interpretation of these data would suggest that PGCs could have initiated specific epigenetic programmes at this stage of embryogenesis.

The expression of chromatin regulation mediators prompted us to question whether germ fate acquisition would occur through post-transcriptional regulation as for early stages or would also be observable at epigenetic levels.

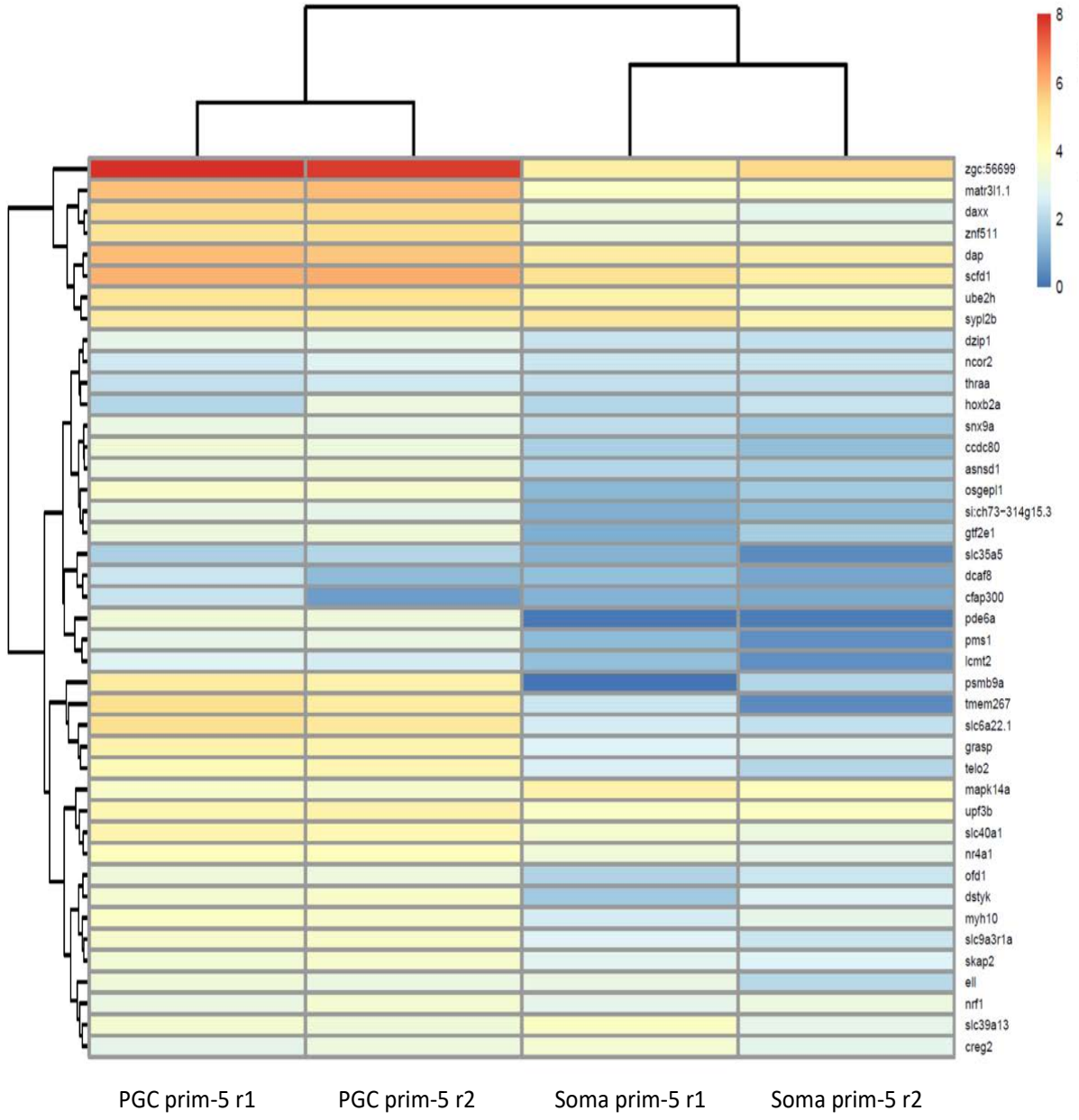

Figure 5. 3: Heatmap of transcript levels for genes associated with histone acetylation.
Raw read counts for gene expression were normalised to transcripts per million (tpm). The scale shows the $\log_2$(tpm+1) of genes associated with the GO term GO:0016573 for histone acetylation. Genes were grouped and normalised expression values (tpm) plotted as colour coded heatmap.

5.2.3 Dynamics of chromatin accessibility in PGCs during early embryogenesis

In the mouse, PGCs undergo dramatic epigenetic rearrangement while migrating towards the gonads (Hill et al., 2018). This is a requirement for allowing PGCs survival after they reached the target location. As zebrafish relies on the germ plasm, one could propose that the epigenetic reprogramming would not be strictly necessary at these early stages of development. However, significant increase in differential gene expression between PGCs and somatic cells after gastrulation indicates that PGC-specific epigenetic regulation may occur.

In order to verify whether a PGC-specific epigenetic signature exists during and after migration, we aimed to investigate the open chromatin profile of isolated PGCs. Chromatin accessibility is indicative of active genomic sites and/or cis-acting elements (Tsompana and Buck, 2014). If integrated with histone marks and transcriptomic data, the open chromatin profile can accurately predict the position of cis-acting elements. To investigate open chromatin states, ATAC-seq was performed on PGCs and somatic cells at four time points (high, dome, 10-somites and prim-5 stages). The main aim was to link any characteristic change in the germ plasm behaviour with the chromatin landscape of the germ plasm-carrying cells. DNA library preparation was carried out as described in Buenrostro et al., 2013 from isolated, Buc-GFP-positive and -negative cells.

After chromatin preparation and tagmentation (Figure 5.4A), DNA fragments tagged with sequencing adaptors were amplified by standard PCR. The total number of PCR cycles used for library amplification was determined for each sample by monitoring the reaction by qPCR (Table 5.4).

Table 5.4: Summary of ATAC sample preparation

Sample	24hpf PGCs	24hpf Soma	24hpf PGCs	24hpf Soma	high PGCs	high Soma
Collection	FACS	FACS	FACS	FACS	FACS	FACS
# of cells	7000	5000	7000	7000	2780	5000
T.R volume	50ul	50ul	50ul	50ul	50ul	50ul
Adapter	Ad2.3	Ad2.4	Ad2.1	Ad2.2	Ad2.5	Ad2.6
PCR cycles	11	11	11	11	11	11

Only samples which met the expected library size and profile (Figure 5.4B) were selected for sequencing.

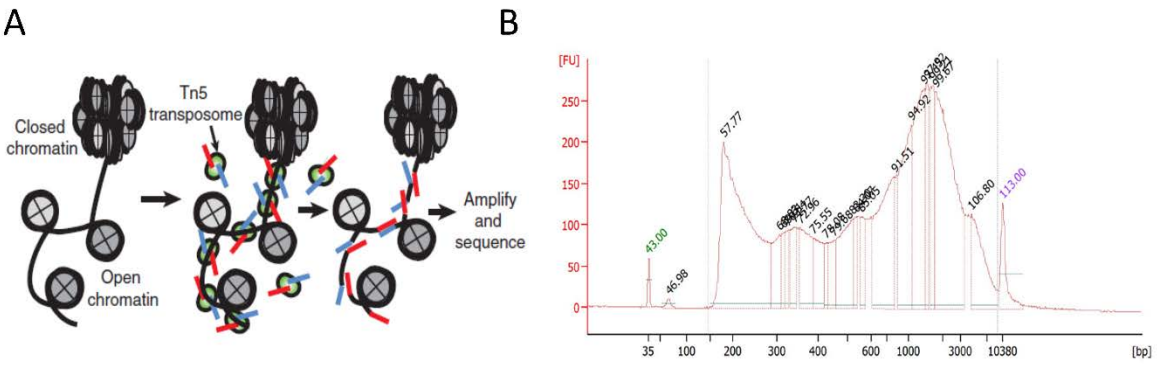

Figure 5. 4: ATAC-seq rational and DNA library profile.
(A) A modified Tn5 transposase targets open chromatin sites and inserts sequencing adaptors while cutting the DNA. The generated tagged DNA fragments can be PCR amplified and sequenced. From (Buenrostro et al., 2013). (B) Bioanalyser track of an amplified ATAC library prior to sequencing. The profile reflects the nature of the DNA packaging, showing open chromatin, mono-, di-nucleosomes and packed DNA. The y axis shows the fluorescent signal intensity (FU).

The sequencing was performed as described in Buenrostro et al., 2013, paired-end and with an expected read count of at least 25 million reads per sample. After library generation and sequencing as described in the methods, raw reads were aligned to the reference genome for

visualisation purposes and peaks were called for downstream analysis. Alignment and peak calling were performed as described in the methods chapter. As the peak calling software (MACS2 v.2.2.4) is designed for determining peaks for ChIP-seq experiments, an input sample was required for normalisation over the background. MACS2 performs better when an input is provided because it calculates the enrichment over the background by estimating a genome-wide local lambda (λBG) for Poisson distribution. However, it would not make sense to include an input control for an ATAC-seq experiment as no antibody-mediated pull down occurs. In this case, MACS2 generates a local λ based on the vicinity of peak centres, which could be less stringent. To filter out lowly represented peaks, an arbitrary fold enrichment threshold was applied to 4 for each called peak upon visual peak inspection.

For better visualisation, reads were shifted by 4 bp to account for Tn5 overhang and extended by 50 bp on both ends to smoothen the peaks. The tracks were then uploaded on the custom UCSC genome browser mirror (http://browser.genereg.net) by our collaborators Piotr Balwierz and Boris Lenhard at Imperial college London (Figure 5.5).

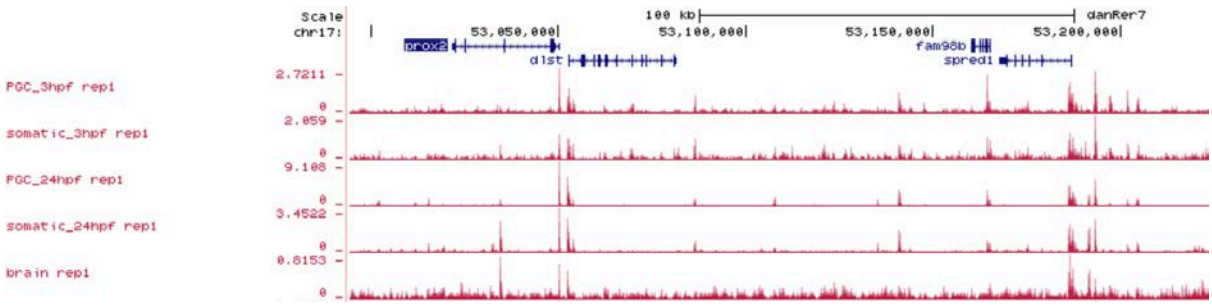

Figure 5. 5: Example of genome browser view of ATAC-seq tracks aligned to the reference genome after peak calling.
ATAC peaks were uploaded into the reference genome browser and showed as cumulative, mapped signal at each genomic location (magenta). Annotated genes are in blue. Open chromatin regions are expected to mark promoter regions and cis-regulatory elements surrounding the genes (distal element).

After grouping shared peaks between samples, sample-sample correlations were plotted. For all replicates, the correlation ranged from 0.776 to 0.981 (Figure 5.6, left), with the lower

correlation observed between the two replicates of PGCs at high stage, which were obtained from low cell number (1800 cells). Overall, the data were considered satisfactory and suitable for further analyses.

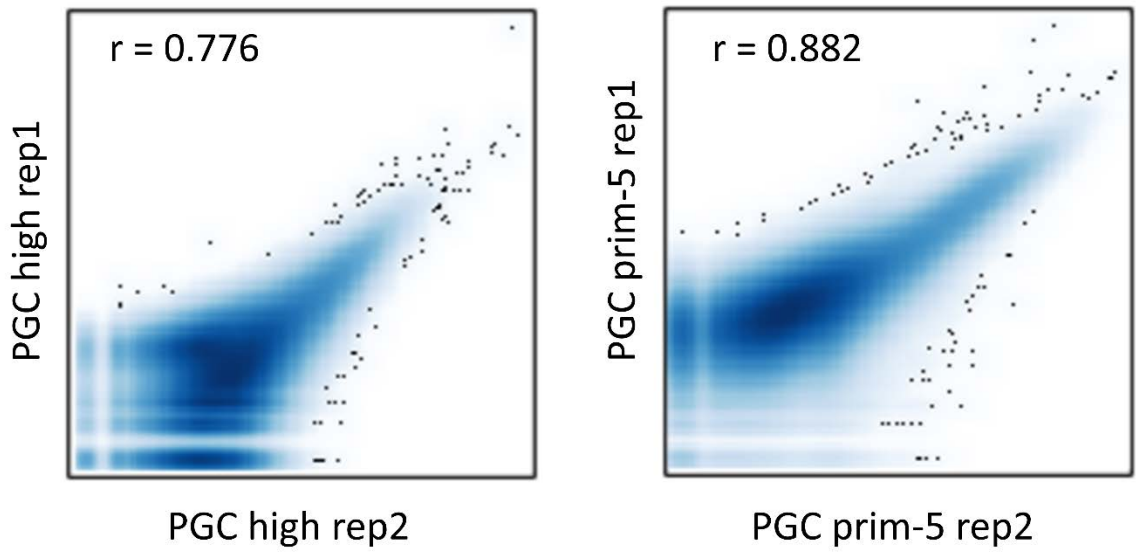

Figure 5. 6: Correlation scatterplots of ATAC-seq replicates.
Smoothed signal of ATAC-peaks significantly enriched over the background after threshold was applied. Colour shade indicates density. Black dots represent samples which cannot be fitted by the smoothing function. Only overlapping peaks between replicates are included in the analysis.

5.2.3.1 Epigenetic variation between PGCs and somatic cells is observed after gastrulation

Following this, global ATAC profiles were analysed comparing PGC and somatic cell groups at different developmental stages. In agreement with the transcriptomic analysis (Figure 5.1A, B), hierarchical unsupervised clustering of open chromatin regions showed that the inter-sample variance is greater between developmental stages rather than cell types during early embryogenesis (Figure 5.7). The fact that PGCs and somatic cells cluster together at high and dome stage is indicative of similar chromatin accessibility, suggesting that both undergo coordinated routes during early phases of epigenetic rearrangement. Notably, the variance between PGC and somatic chromatin accessibility profiles increases after migration has started (10-somites and prim-5).

As described earlier, the dome stage coincides with the onset of PGC migration. As there has been no observable evidence for a PGC-specific epigenome at dome stage, this would imply that initiation of germ cell migration may not be epigenetically driven. In fact, the most marked difference between samples was observed for ATAC profiles from prim-5 PGCs and somatic cells (Figure 5.7 and 5.8A). This separation was further confirmed by PCA plot when only PGCs and somatic cells at prim-5 stage were taken into account (Figure 5.8A).

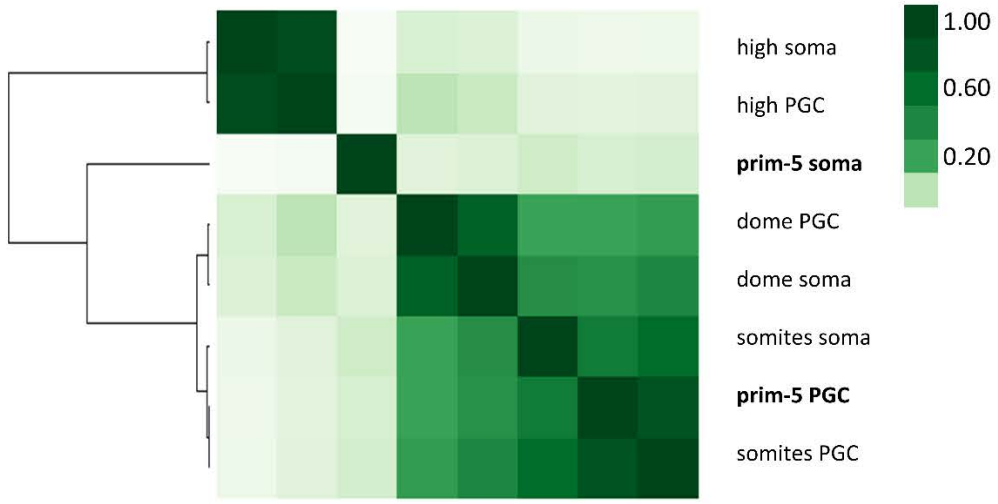

Figure 5. 7: Hierarchical unsupervised heatmap of variance between ATAC-seq samples after peak calling.
Sample correlation for open chromatin profiles across stages. Shades of green indicate the correlation coefficient between samples.

In order to find differential accessible chromatin regions between PGCs and somatic cells at prim-5, DESeq2 was run on counts of Tn5 insertion occurrence. When comparing PGCs and somatic cells at prim-5 stage, 3056 regions were significantly more accessible in the PGCs, while 14512 were more open in somatic cells (FDR < 0.05) (Figure 5.8B). This data indicated that most of the putative cis-acting elements identified by ATAC-seq in the somatic cells are less opened in the PGCs, while only a smaller proportion of PGC-specific genomic sites are selectively more open in the PGCs (Table 8.5, 8.6, Appendix).

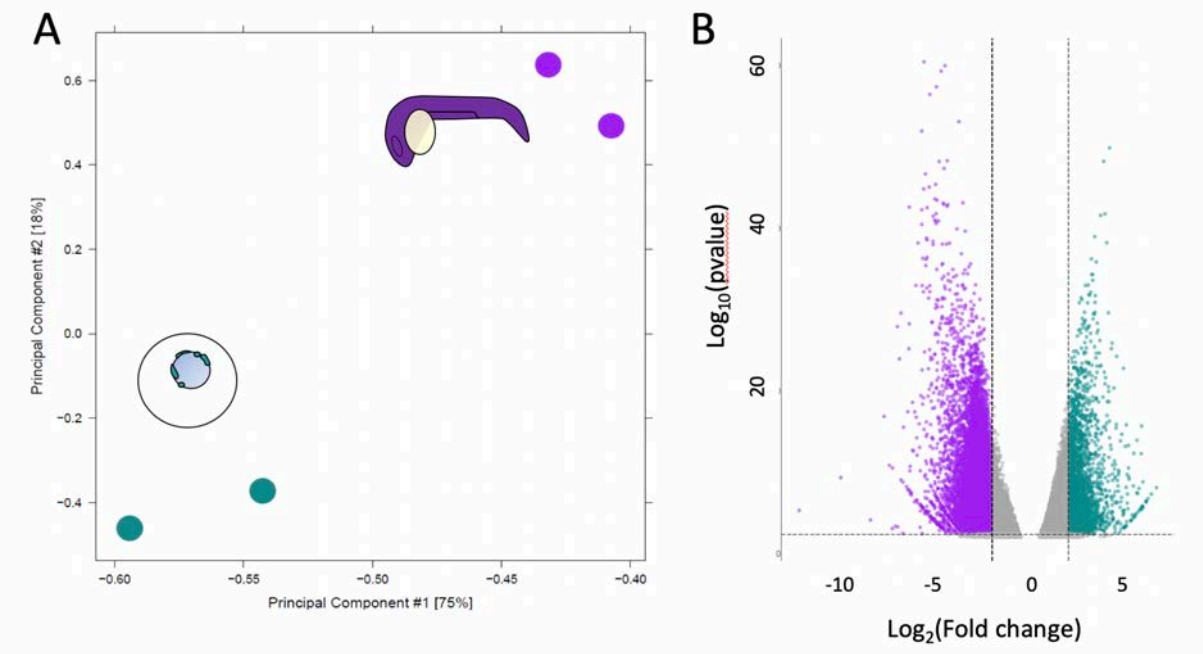

Figure 5. 8: Global differences in the open chromatin profiles of PGCs (dark cyan) and somatic cells (purple) at prim-5 stage.
(A) PCA plot of open chromatin profiles shown as raw coverage highlights clustering of replicates and divergence between the open chromatin profiles of PGCs and somatic cells. (B) Volcano plot of $\log_2$(fold change) vs $\log_{10}$(p-value) for differentially opened regions between PGCs and somatic cells. Upregulated regions in the somatic cells are in purple, while upregulated regions in the PGCs are in dark cyan.

In order to check which classes of genes showed differential epigenetic regulation, each significantly up- or downregulated region was assigned to the closest TSS. This approach allowed us to estimate genes that are likely regulated by the nearby identified open chromatin region. When GO analysis was performed on both genes associated with regions more opened in PGCs or more opened in the somatic cells, we noted a remarkable enrichment for development-associated pathways downregulated in the PGCs (Figure 5.9, bottom). Interestingly, GO terms for genes associated with ATAC peaks upregulated in the PGCs were enriched for ion and transmembrane transporters (Figure 5.9, top). While this result was unexpected, it was not followed up, and it could be interpreted as caused by biased analysis of long genes (Zylka et al., 2015).

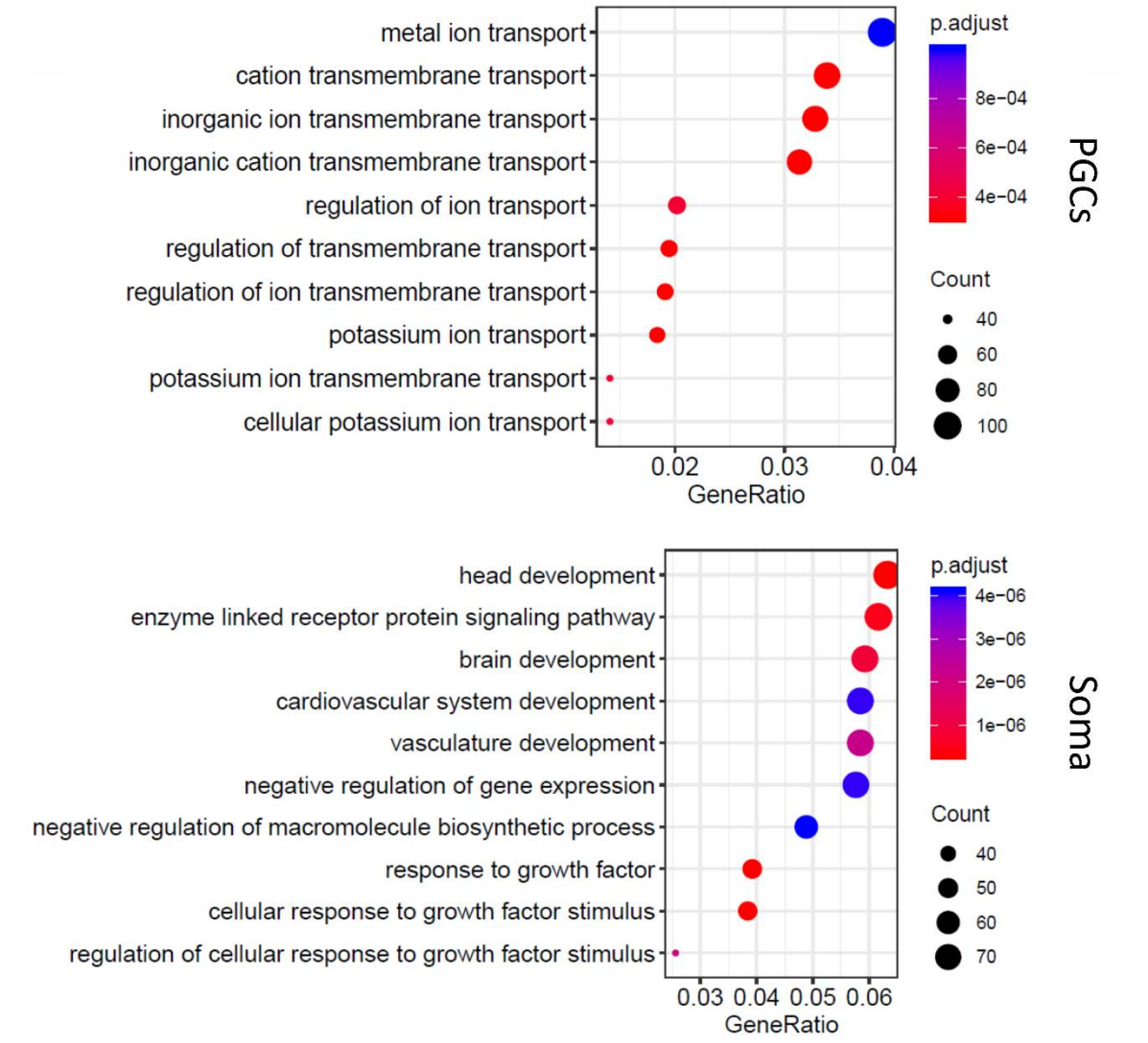

Figure 5. 9: GO analysis for biological processes associated with genes in proximity of open chromatin regions.
Ensembl IDs for the analysis were obtained by linking differentially regulated ATAC-peaks between PGCs and somatic cells with the closer gene. p-adjusted < 0.005. Count indicates the number of genes contributing to each GO category.

5.2.4 PGCs have less open chromatin regions on the gene bodies compared to the somatic cells

Next, in order to study how the open chromatin distributes within the genome in respect to gene elements (promoters, gene bodies, intergenic regions), the percentages of ATAC peaks matching each gene element were plotted. Seven genomic elements were considered and the

number of ATAC peaks upregulated in PGCs and somatic cells coinciding with each element was retrieved. As expected, only a small percentage of peaks was found in regions downstream of the promoter, in exons and at the 3'/5'-UTRs in both PGCs and somatic cells (Figure 5.10). Gene bodies are usually organised in compact chromatin and are not expected to show accessible sites (Wolffe and Guschin, 2000).

Contrasting this, we noted that the percentage of ATAC peaks overlapping promoters and introns were significantly different between the two cell types. In fact, 43% of those peaks upregulated in PGCs overlapped promoter regions, while 11% was associated with introns. On the other hand, a significant majority of peaks upregulated in the somatic cells (41%) were found within introns, compared to 6% associated with promoters (Figure 5.10).

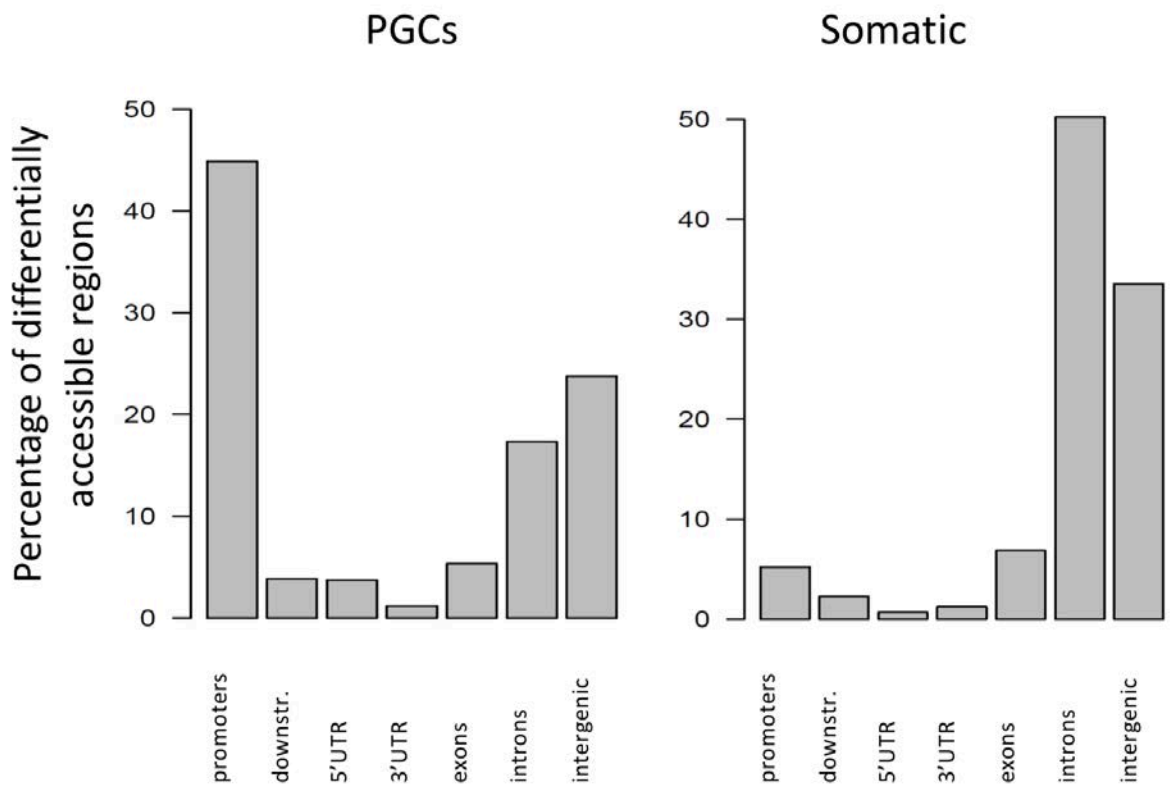

Figure 5. 10: Distribution of ATAC peaks based on gene features.
Histograms showing the percentage of differentially regulated ATAC peaks between PGCs (left) and somatic cells (right) matching a certain genetic element. Promoters include regions 1000 bp up- and down-stream the TSS.

This result was surprising mainly because tissue-specific epigenetic variation is rarely observed on promoters, which tend to be stable across development and cell types (Kubik et al., 2017). An early interpretation may suggest that somatic cells have a tendency to control chromatin accessibility at distal regulatory regions. This interpretation led us to hypothesis that in PGCs only a few distal regions are accessible in comparison to the somatic cells. As regulatory regions in proximity of the TSS often sit within introns, one could therefore expect a significant effect on intronic peaks.

Nevertheless, an initial interpretation of this output could suggest that PGCs positively regulate transcription at the level of the promoters and negatively at the level of putative enhancers (distal cis-acting elements).

However, the fact that the percentage of differentially regulated intergenic regions was comparable between the two cell types led us to propose a different scenario. We checked with help from collaborators at Imperial college London (Dr. Piotr Balwierz and Dr. Boris Lenhard) the density of ATAC peaks within a 100kb window around the TSS of PGCs and somatic cells. To do that, promoters were excluded by setting a cut off of 500 +/- around the TSS. Interestingly, although the basal ATAC signal is comparable between putative somatic and PGC peaks upstream of the gene body, we observed a significant depletion in ATAC signal downstream of the gene promoter in the PGCs (Figure 5.11). Overall, our data suggested that PGCs are depleted in chromatin accessible regions in the gene body (introns) compared to the somatic cells, while regulatory regions upstream of the TSS do not show significant difference.

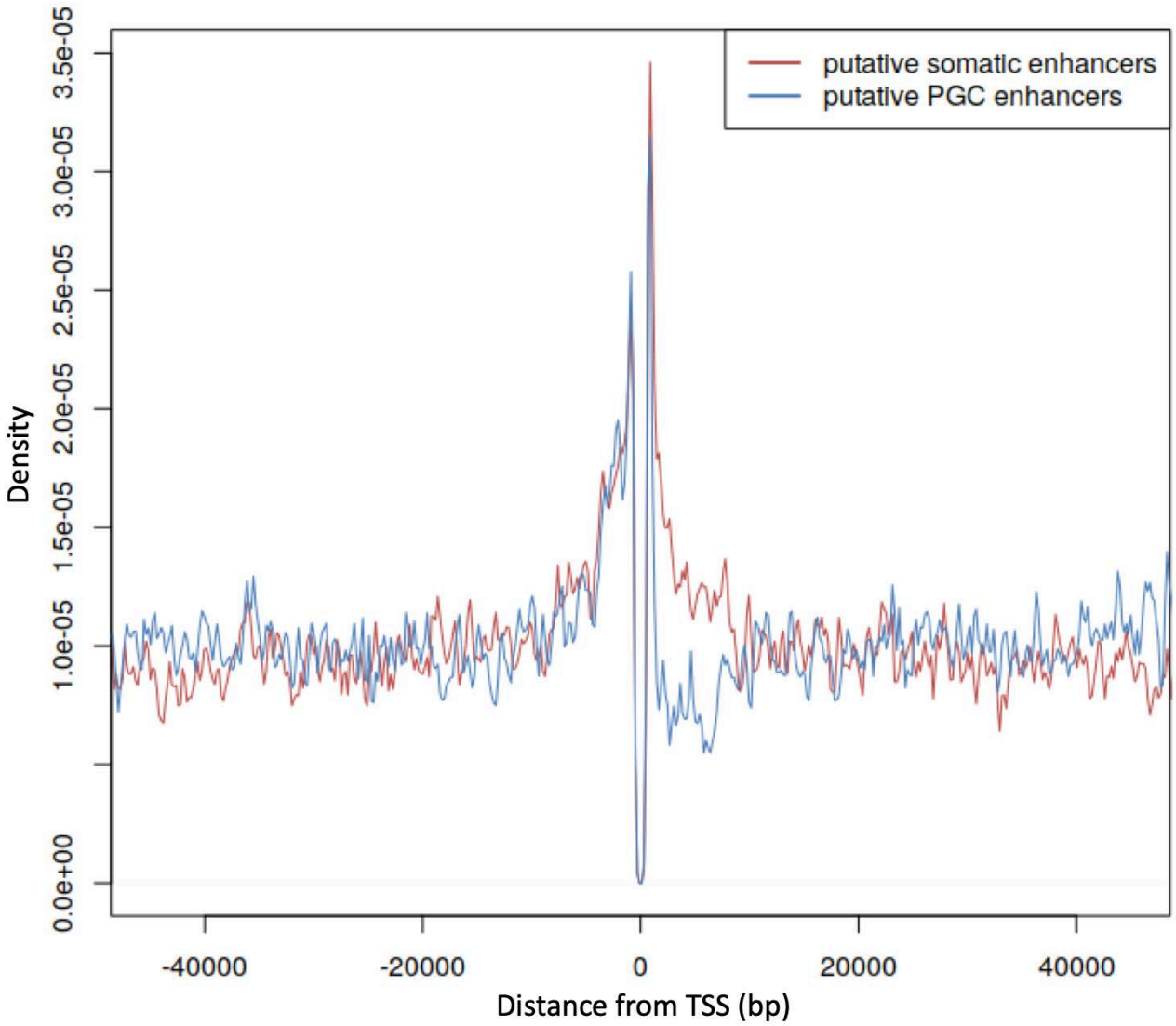

Figure 5. 11: Cumulative density of ATAC signal around promoters.
Density plot for putative enhancers location in PGCs (blue) and somatic cells (red) at prim-5. The average fold change of each peak upregulated in one cell type was plotted on each location in a 100 kb window centred on the TSS. Promoter is depleted in signal because excluded from the analysis.

Taken together, these results highlight that there is a tendency in the PGCs to keep chromatin compacted downstream the gene body. We were unable to link a observed biological function for this phenomenon, however, this may indicate a consistent PGC-specific type of transcriptional regulation, or a conserved feature of gene classes specifically downregulated in the PGCs (for example developmental genes).

5.2.5 Correlation between chromatin opening and transcriptional activity

Next, we asked whether the differential chromatin accessibility between PGCs and somatic cells correlates with gene transcription. The fact that regions in proximity to developmental genes showed a tendency of being more compacted in the PGCs was already indication that PGC fate could be linked to inhibition of developmental and differentiation pathways.

Transcription is the final outcome of a series of complex regulatory mechanisms, which occur at the chromatin level and the global understanding of these events on a specific locus is crucial to predict transcriptional dynamics in space and time. Nowadays, it is established that heterochromatin regions are less likely to accommodate active genes, while euchromatin promotes transcription. In fact, chromatin compaction functions as a physical barrier for transcription factors and polymerase recruitment. Compaction of the chromatin is a highly regulated process, which has been shown to be dependent on various factors including DNA primary sequence, histone/DNA modifications and regulatory RNAs. Importantly, although chromatin compaction is sufficient to block polymerase recruitment, open chromatin does not always result in transcriptional activation (Zhang et al., 2018).

Therefore, transcriptional control is accomplished via complex regulated mechanisms and often chromatin packaging does not reflect the transcriptional capacity of a certain locus. However, it was shown that comparative chromatin accessibility profiling could provide a reliable estimation of different regulation between two samples (Song et al., 2011; Tsompana and Buck, 2014). To this end, we utilised both ATAC- and RNA-seq readouts in order to predict truly active loci essential for germ cell development. As reported, a remarkable gene upregulation was observed between PGCs and somatic cells, therefore detection of cell type-specific chromatin signature for these genes was expected. To verify to what extent the observed differential chromatin opening was causing differential gene expression in the PGCs, the overlap between upregulated genes and upregulated chromatin regions in both the cell types

was determined. Differentially regulated genes were retrieved as described in 4.2.2.3, by performing differential expression analysis between PGCs and somatic cells at the same stage (padj < 0.1). In addition, the gene list was further filtered for genes upregulated from the previous developmental stage. The reason for filtering was that the pool of RNA detected by RNA-seq does not reflect actively transcribed genes, but only the steady-state RNA at a certain time. To account for this, I intersected the genes differentially regulated at prim-5 between PGCs and somatic cells with the genes upregulated from dome to prim-5 stage as these would well represent the change between maternal to zygotic transcription.

In order to intersect ATAC- and RNA-seq data, all the differentially opened regions were assigned to a gene based on distance. Each peak was associated to the nearest gene within a 50kb region. In total, 899 differentially transcribed genes and 9171 open chromatin regions unique to the PGCs were overlapped. Of these, 205 (22.8%) overlapped (Figure 5.12A). In contrast, a total of 1542 (41.3%) overlapped when 3732 differentially transcribed genes and 8861 open regions significantly more accessible in the somatic cells were intersected (Figure 5.12B).

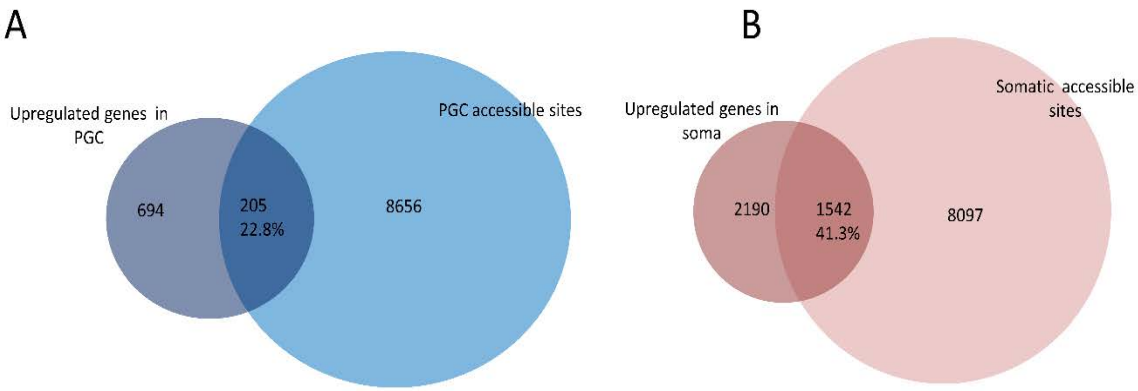

Figure 5. 12: Correlation between chromatin profile and transcription via intersection analysis. Intersection between Ensembl IDs of upregulated genes in PGCs (left) or somatic cells (right) and Ensembl IDs associated to upregulated ATAC peaks in each cell type.

Overall, in both the cell types the overlap was lower than expected, based on visual inspection of the genome browser tracks. However, this type of approach had two main limitations: first,

assigning ATAC peaks to genes based on distance can be unreliable for reasons explained earlier. Second, multiple open chromatin regions can act on the same gene (ATAC peaks). Unfortunately, due to lack of chromosomal conformation/capture data in the PGCs, we had to rely on the correlation between ATAC peaks and gene expression based on genomic distance. To better evaluate the degree of correlation between transcription and chromatin accessibility, RNA and open chromatin levels were compared by identifying classes of genes based on differential expression analysis and measuring the chromatin accessibility of regions associated to those genes. First, the chromatin accessibility score and the transcript levels were compared in the somatic cells at prim-5. To correctly account for the cumulative effect of multiple peaks, the accessibility scores for all ATAC peaks associated to a gene were summed. As gene expression fold change is not indicative of RNA levels, normalised RNA-seq reads (transcript per million, tpm) were used. In fact, a high fold change could result from lowly expressed genes, which are more likely to be inconsistent across replicates, potentially generating unreliable results. A standard way to overcome this problem is to set a lower threshold for expression and discard those transcripts. However, to avoid potential biased pre-selection of the datasets, we decided to follow a different approach. All transcripts were assigned to a bin based on tpm values. Four bins were generated: tpm between 0 and 1, tpm between 1 and 10, tpm between 10 and 100 and tpm bigger than 100. When, the cumulative chromatin accessibility score for each bin was calculated, a general correlation between transcription and open chromatin was observed (Figure 5.13). Interestingly, discrimination of the total peaks at promoter and distal regions showed that the correlation was higher for promoter-associate peaks (Figure 5.13). This result was expected for two reasons: first, promoters-overlapping peaks cannot be mis-assigned. Second, it is known that opening of the chromatin at the promoter level is generally more effective than on enhancers, whose role often is to provide tissue- or time-specificity (Melamed et al., 2016; Shlyueva et al., 2014).

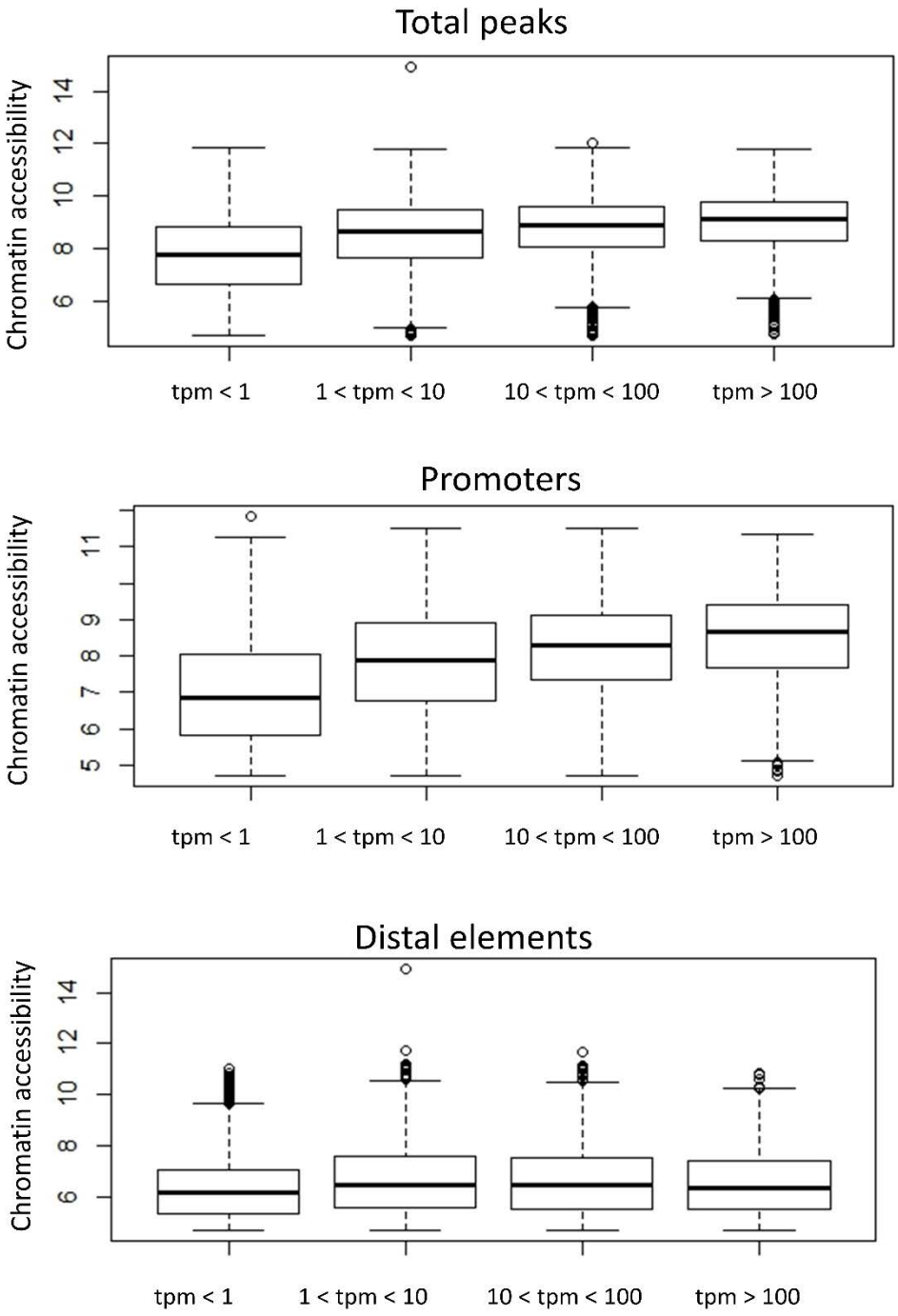

Figure 5. 13: Chromatin accessibility scores relative to transcript levels in somatic cells.
Normalised transcript levels (tpm) were subdivided into four categories and the chromatin accessibility score (fold change over background, threshold = 4) was plotted for each category. Accessible regions were further divided in promoters and distal regions.

These results indicated that in the somatic cells the gene expression levels correlate with the global chromatin accessibility of the surrounding genomic region.

Next, to verify whether the chromatin accessibility could be linked to transcription for both cell types, the fold change values obtained upon differential expression analysis were checked for genes with cumulative more accessible chromatin in the PGCs and in the somatic cells. Similarly, differential accessible regions and differentially expressed genes correlated in both PGCs and somatic when total peaks were accounted (Figure 5.14). As PGCs appeared to be more proximally regulated than the somatic cells, all the peaks within 5 and 50kb from the TSS were analysed separately. As expected, more proximal accessible regions were more correlated with gene transcription in comparison to distal regions (Figure 5.14).

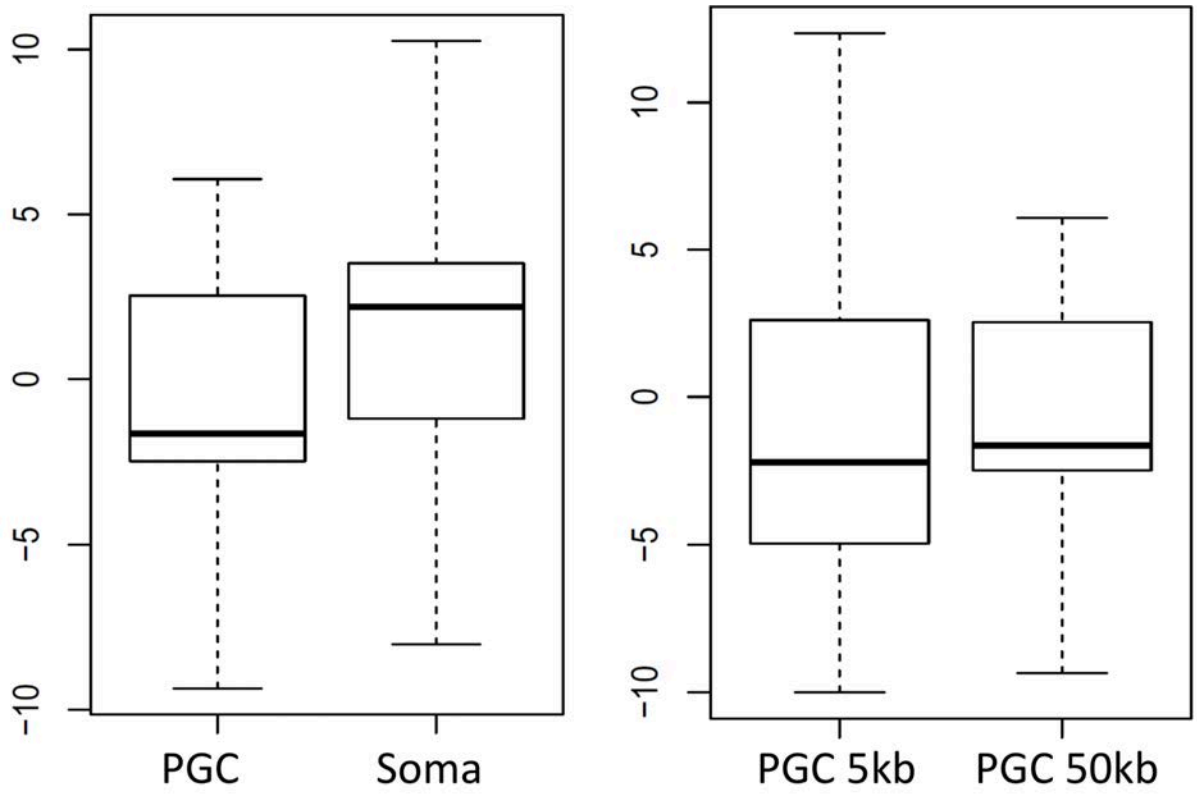

Figure 5. 14: Correlation between open chromatin and differential gene expression in PGCs and somatic cells.
The boxplots show the $\log_2$(fold change) of gene expression between PGCs and somatic cells at prim-5 stage for differentially regulated ATAC peaks. Median values are reported under the boxplot and drawn as black lines. On the left panel, total peaks are accounted for PGCs and somatic cells. On the right panel, a distance filter was applied.

Finally, to further verify the estimated association between chromatin accessibility and transcription, GO analysis was carried out on the differentially opened regions between PGCs

and somatic cells. Each differentially regulated region was assigned to the closest TSS independently of orientation. This approach allowed us to estimate genes that are likely regulated by the nearby identified open chromatin region. When GO analysis was performed on both genes associated with regions more opened in PGCs and more opened in the somatic cells, it was noted that development-associated pathways were less present in the PGCs (Figure 5.15, bottom). The strong enrichment for developmental GO terms suggested that PGCs downregulate chromatin accessibility in proximity of genes involved in somatic development. On the other hand, GO terms for genes associated with ATAC peaks upregulated in the PGCs were enriched for transmembrane transporters, signal transductors and factors involved in ectodermal development (Figure 5.15, top). Overall, GO analysis confirmed that PGCs actively regulate signalling and cellular homeostasis. It was also important to highlight that developmental genes are depleted in open chromatin regions in the PGCs, suggesting important potential mechanisms of resisting somatic differentiation and retention of pluripotent features. Then, to better understand the nature of differential chromatin regulation, promoter-associated and distal open chromatin regions were treated separately.

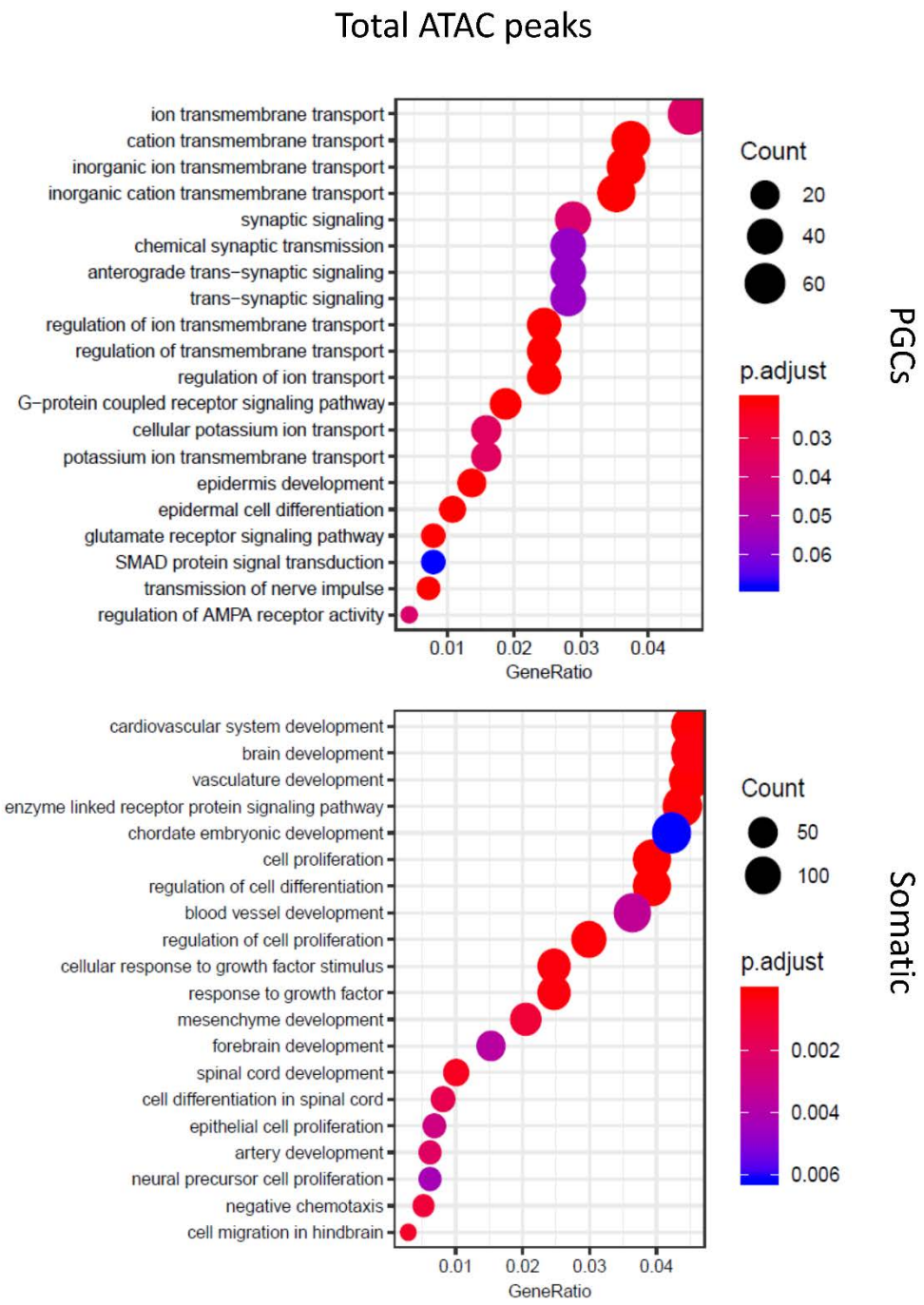

Figure 5. 15: GO analysis for biological processes associated with genes in proximity of open chromatin regions.
Ensembl IDs for the analysis were obtained by linking differentially regulated ATAC-peaks between PGCs and somatic cells with the closer gene. p-adjusted < 0.005. Count indicates the number of genes contributing to each GO category.

5.2.5.1 Opening of chromatin around developmental genes is reduced in the PGCs

In the effort to determine how the PGCs epigenetically commit and develop, cis-acting elements that show significant differential regulation between PGCs and somatic cells were further characterised. Visualisation of the genome browser suggested that diverse classes of genes were differentially regulated at the chromatin level in PGCs and somatic cells (Figure 5.16). On a significantly upregulated gene in the PGCs (*dnd1*), ATAC-seq detects increased promoter accessibility specifically in post-migratory PGCs. Concomitantly, a putative cis-regulatory element about 10kb upstream of the promoter appears during somitogenesis. Oppositely, the promoter of the developmental gene *slit1a* and an intronic region become more accessible in the late somatic cells but not in the PGCs.

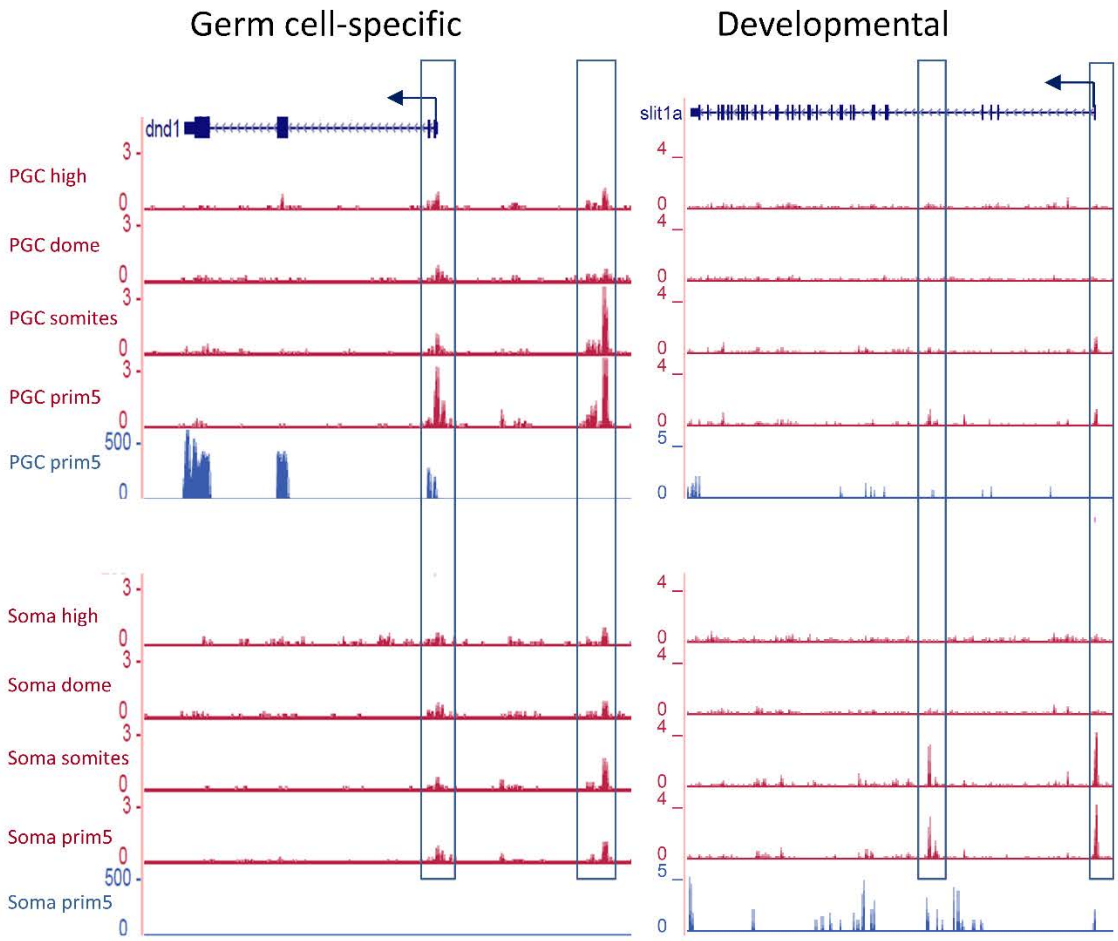

Figure 5. 16: Genome browser view of ATAC- and RNA-seq tracks for two representative genes in PGCs and somatic cells.
ATAC-seq signal is shown over early development (magenta). RNA-seq is shown for prim-5 stages only (blue). Blue boxes highlight promoter regions and candidate distal cis-regulatory elements based on ATAC peaks.

To further validate this observation, the fold enrichment over the background for peaks overlapping and non-overlapping with promoters was measured. The fold enrichment score was normalised for the average peak height of ten housekeeping genes and compared between PGCs and somatic cells. Interestingly, we noted that PGCs show significantly higher chromatin accessibility at the promoter of germ line-specific genes. Of the nine tested germ genes, only *piwil1* was an outlier, possibly caused by low ATAC signal on the promoter (data not shown). Interestingly, measurement of peak intensities at the promoter of different gene classes showed a tendency of the PGCs to have promoters opened also on developmental genes. Housekeeping

genes did not show significant variation among the two cell types, while tissue-specific genes have reduced promoter accessibility in the PGCs compared to the somatic cells.

Surprisingly, this analysis evidenced that promoters of developmental genes are more accessible in the PGCs than in the somatic cells (Figure 5.17A). Because this observation clashed with what was reported earlier, open chromatin on promoter-distal regions was investigated further (50kb up and downstream the TSS). Strikingly, an opposite trend was observed. Distal elements in PGCs had significantly lower chromatin accessibility at regions around developmental genes (Figure 5.17B). In this analysis, master regulators of axial and tissue development were selected. Conversely, the pluripotency-associated genes *nanog* and *sall4* showed an opposite trend (Figure 5.17B, top).

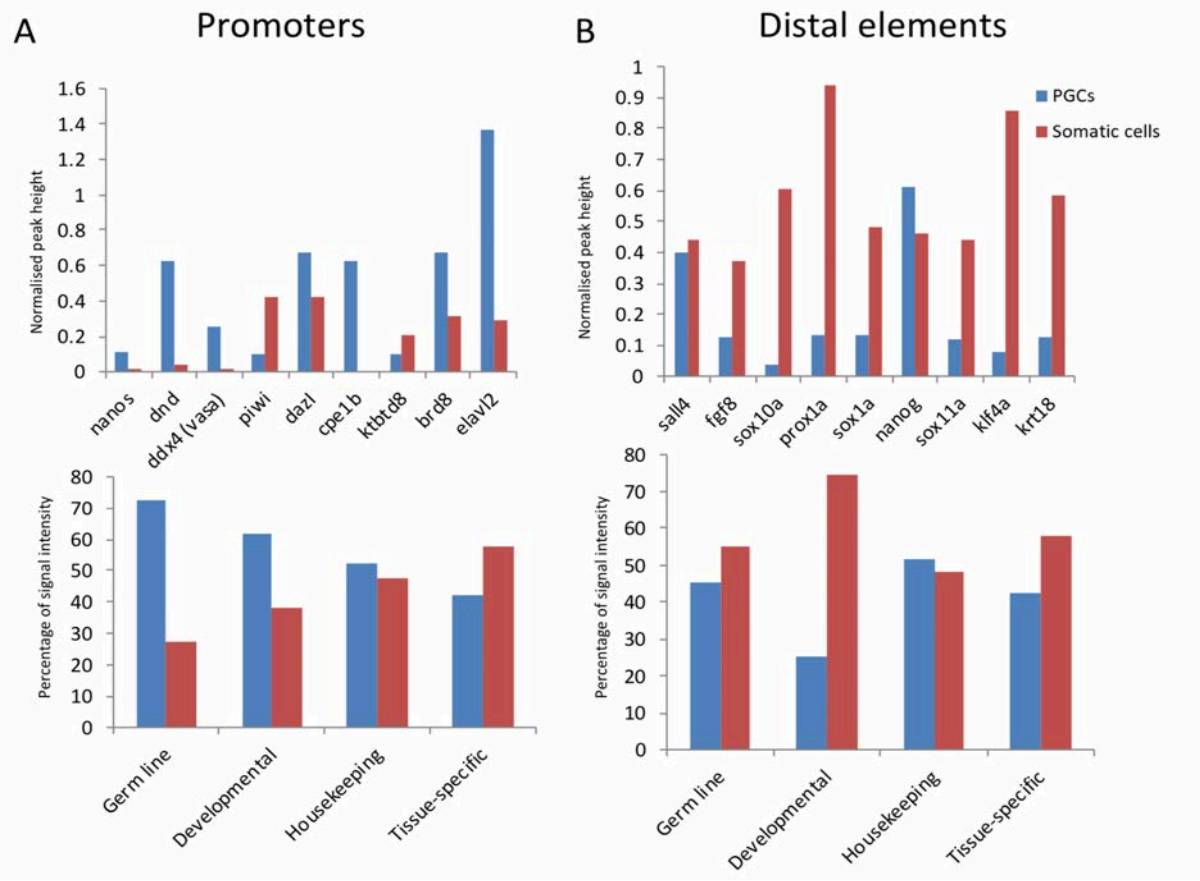

Figure 5. 17: **ATAC signal at promoters and distal elements shows diverse chromatin regulation for gene classes in PGCs and somatic cells.**
(A) Normalised ATAC-signal (fold change over the background) was plotted for sample subclasses of germ, developmental, housekeeping and tissue-specific genes in the PGCs (blue) and somatic cells (red) for promoters (A) and distal elements (B). The percentage of signal intensity in the subclasses of genes is shown at the bottom.

To verify that PGCs reduce chromatin accessibility on non-promoters cis-acting elements, the ATAC signal on developmental genes (GO term GO:0032502) was analysed. After having assigned each ATAC peak to the closed TSS and selected developmental genes only, the fold enrichment of chromatin accessibility was plotted against the background in PGCs and somatic cells. In order to filter out lowly enriched peaks, a threshold of 4 to the fold change was applied. The lower levels of chromatin accessibility observed in the PGCs compared to the somatic cells on developmental genes confirmed that PGCs tend to compact chromatin around developmental genes when compared to the rest of the embryo (Figure 5.18). In combination with the RNA-seq, showing downregulation of development- and differentiation-associated genes, it is tempting to suggest that PGCs inhibit expression of somatic/developmental genes by reducing chromatin accessibility on distal elements. To verify this hypothesis, candidate enhancers were selected and then linked to gene transcription.

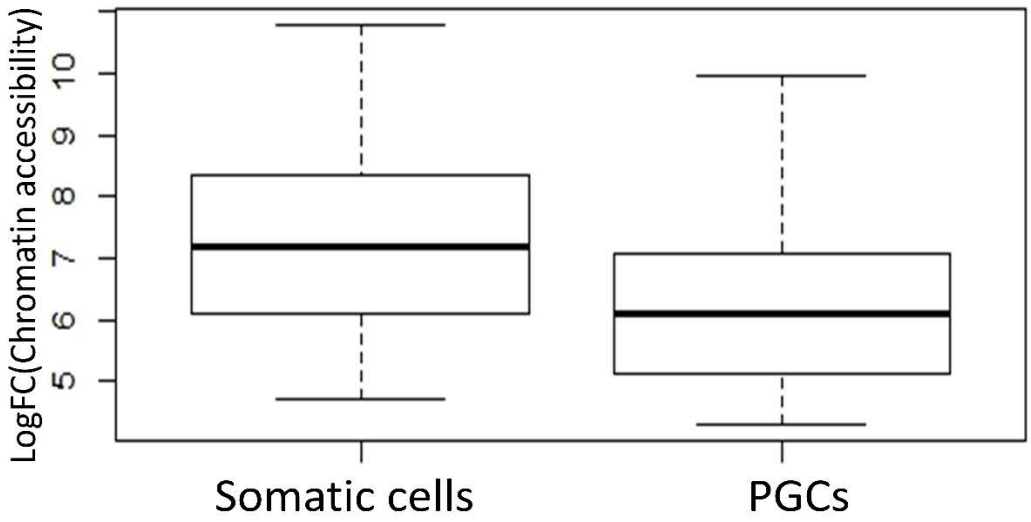

Figure 5. 18: Chromatin accessibility on distal elements associated with developmental genes in PGCs and somatic cells.
The boxplots show the $\log_2$(fold change) over the background of gene expression between PGCs and somatic cells at prim-5 stage for differentially regulated ATAC peaks. Median is drawn as black lines.

5.2.5.2 Prediction of enhancer position and correlation with gene expression

In order to predict position of active enhancers, their effect on gene expression should be determined. By definition, active enhancers promote transcription of nearby genes by helping and initiating the recruitment of the PIC and polymerases. For this approach, there are two main complications: first, embryonic cells often need primed enhancers (Zhang et al., 2018), regions of open chromatin that appear to be active by virtue of having increased accessibility but do not drive transcription yet. The second issue comes from the technical challenge of connecting an enhancer to their target gene. As explained, the complexity of enhancer-promoter interactions goes beyond simple genomic proximity. In fact, enhancers can act even Mega Bases (Mbs) away of their target and control multiple genes. Also, enhancers are tissue-specific, while most of the current data comes from pooled cell types. Currently, the best approximation of enhancer-promoter interactions comes from 3D genome sequencing. This field includes those techniques based on biotinylation and capture of fragments that are in physical proximity at a certain time. These are known as Chromosome Capture (C) and are developed in different variations such as Hi-C, 3C and 4C. At the time this project was designed, any available C technique was inapplicable to the small PGC numbers. However improvements were recently made (Díaz et al., 2018) and we predict that in the near future information about genomic interactions will become available for very small cell populations. Given that, it was decided to correlate gene expression with the local levels of chromatin openness for every gene. This was done by assigning Ensembl IDs to any predicted enhancer peak based on the closest gene. Although, this method can lead to incorrect assignments, it achieves a decent reliability when performed genome wide, in particular in gene-desert regions (Kumasaka et al., 2019).

First, reproducibility of ATAC peaks was assessed via irreproducible discovery rate (IDR), which filters out each open chromatin peak not detected in both the replicates (Zhang et al., 2017). After peak overlap was performed, reproducible peaks were assigned to each stage and cell type (Table 5.5).

Table 5.5: Number of reproducible peaks matching previously published datasets

	PGC prim-5	Soma prim-5	PGC high	Soma high
Number of peaks	61641	79494	14029	12862

To assign these peaks as potential enhancer candidates, existing ChIP-seq data were used. In fact, it was shown that putative enhancer regions are marked by specific histone modifications, such as H3K4me1 and H3K27ac (Heintzman et al., 2009; Melamed et al., 2016).

Therefore, differentially open regions between PGCs and somatic cells at prim-5 were overlapped with regions marked by H3K4me1 and H3K27ac from previously published datasets (Bogdanovic et al., 2012). As expected, PGC peaks showed minimal overlap with somatic putative enhancers, while over 30% of somatic detected ATAC peaks were having enhancer signatures (Figure 5.19). In total, 225 and 7079 regions overlapped putative enhancers in the PGCs and somatic cells respectively ($log_2FC > 2$, FDR < 0.01). Of these, 82 and 4367 were found within 50kb from a TSS.

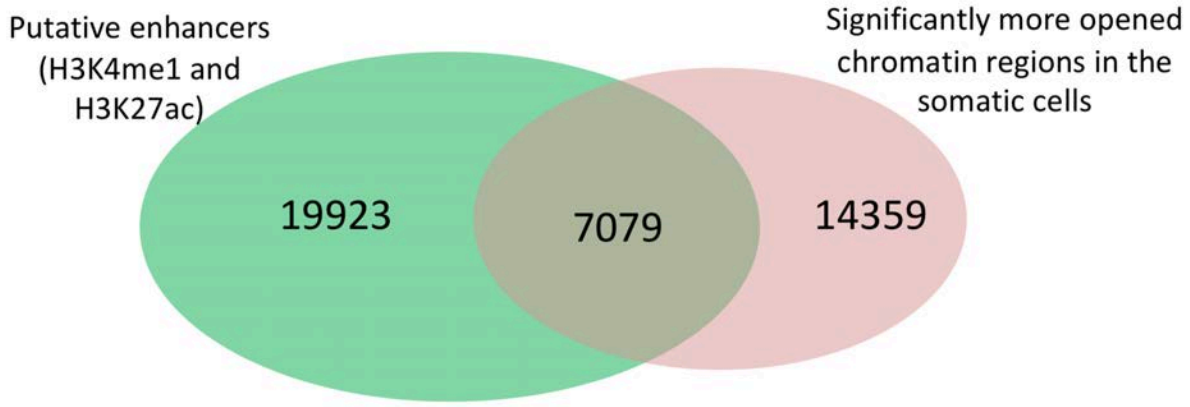

Figure 5. 19: ATAC-seq peaks with significantly more opened in the somatic cells overlapping previously published histone modification marking active enhancers.
Putative enhancers were assigned to genomic regions marked by both H3K27ac and H3K4me1. These were intersected with ATAC peaks significantly more opened in the somatic cells versus PGCs at prim-5 stage.

This result was consistent with the fact that published data were obtained from whole embryos, while PGCs-specific enhancers are undetectable in the whole embryo as heavily outnumbered by the other tissues (Bogdanović et al., 2016). Therefore, taking into account the weakness of this strategy in identifying PGC-specific putative enhancers, we could aim to verify whether somatic putative enhancer regions were differentially regulated in the two cell types.

In order to verify if the increase in chromatin accessibility correlates with increase in gene expression, genes regulated by each putative enhancer were computationally estimated. After assignment of each putative enhancer to the nearest TSS, the fold change of differential gene expression was plotted for those enhancers upregulated in the PGCs and those upregulated in the somatic cells. To avoid bias in fold change caused by lowly expressed genes, a threshold of 20 rpm (reads per million) was applied.

Surprisingly, we observed that both groups of enhancers (upregulated in PGCs and upregulated in somatic cells) were associated with increased gene expression in the somatic cells (Figure 5.20). Although a remarkable correlation was observed for putative enhancers and genes upregulated in the somatic cells, the downregulation of genes in PGCs associated with

'active' putative enhancers was unexpected. However, as discussed, the drawback of assigning peaks to TSSs based on distance is the likelihood of generating false positives. Also, we expected more false positives assigned to genes that are proximal regulated, while developmental genes, known to be regulated on distal enhancers, will be less affected. In order to increase the stringency of calling true enhancer-promoter associations, we lowered the distance threshold to 5kb. Interestingly, we noted that the differential gene expression was shifted towards genes upregulated in PGCs (Figure 5.20).

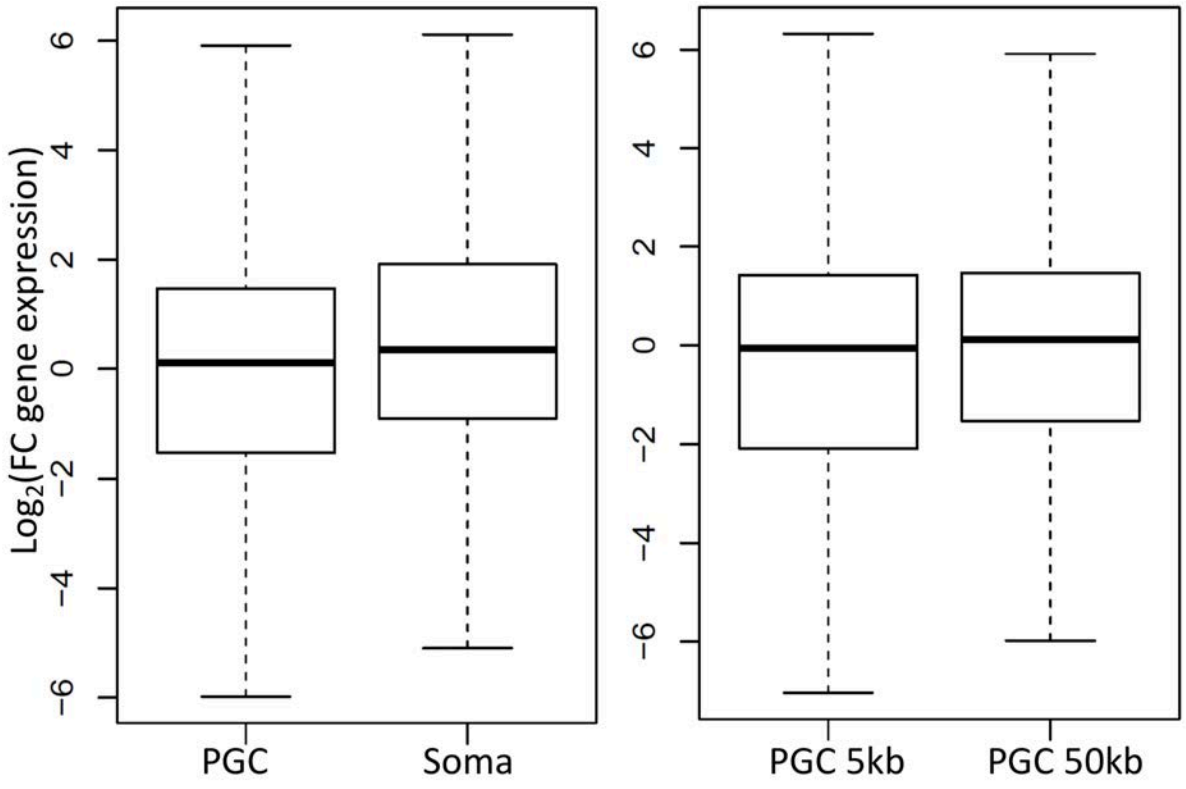

Figure 5. 20: Correlation between open chromatin regions non-overlapping with promoters and differential gene expression in PGCs and somatic cells.
The boxplots show the log_2(fold change) of gene expression between PGCs and somatic cells at prim-5 stage for differentially regulated ATAC peaks. Median values are reported under the boxplot and drawn as black lines. On the left panel, all distal peaks are accounted for PGCs and somatic cells. On the right panel, a distance filter was applied.

Given the challenge in correctly assign PGC enhancers to their target gene as broadly discussed, our results report that very few distal regions were upregulated in the PGCs. In fact, this observation stands regardless of which promoter each enhancer targets. Nevertheless, the

fact that there is no correlation between chromatin opening and transcription when distal elements are considered could suggest that PGCs regulate transcription on promoters and promoter-proximal regions, being less dependent on long-range promoter-enhancer interactions.

5.2.5.3 PGCs are mainly regulated by promoter-proximal elements

The comparative analysis of the open chromatin and transcriptome of the PGCs at prim-5 led us to build two hypotheses: 1) PGCs downregulate expression of developmental and somatic genes upon compaction of enhancers. 2) PGCs rely on promoter and promoter-proximal cis-acting element regulation to control gene expression.

To validate that cis-acting elements involved in transcriptional regulation in PGCs are found in proximity of the TSS, the cumulative distance in respect to TSSs was checked for each ATAC peak. To this end, the proportion of distal ATAC peaks in the PGCs and somatic cells was plotted in relation to the distance from the nearest TSS. This analysis indicated that 49% of the peaks upregulated in the PGCs was found within 1kb from the closest TSS and 74% within 10kb (Figure 5.21). In contrast, less than 10% of the peaks upregulated in the somatic cells are located within 1kb from the TSS with the vast majority being distributed between 5 and 50kb away from the gene promoters (Figure 5.21). Although a strong contribution is given by promoter-associated peaks in the PGCs (Figure 5.17), the majority of peaks is found between 1 and 10kb away from the promoters. This observation was confirmed by visual inspection of PGC-specific genes. In particular, germ cell genes, such as *dazl*, *dnd1*, *ddx4*, *gra* and *piwi* were all preceded by a PGC-specific ATAC peak within 10kb upstream the TSS (data not shown).

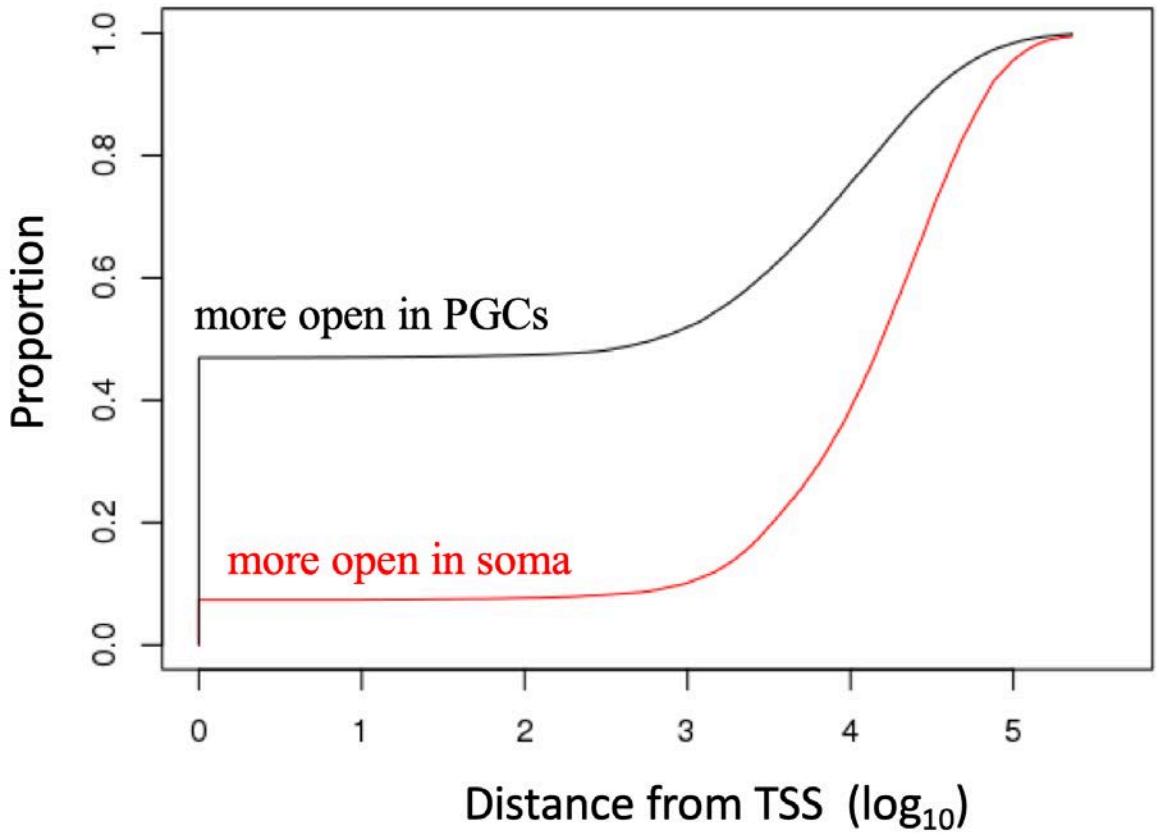

Figure 5. 21: **Cumulative proportion of ATAC-seq peaks in relation to the distance from Ensembl TSSs.**
Cumulative distribution of PGC- and somatic-specific ATAC peaks in relation to distance from TSSs. The black line shows the distribution of open chromatin regions upregulated in the PGCs at prim-5 stage. The red line shows the distribution of open chromatin regions upregulated in the somatic cells at prim-5. The threshold for ATAC peak fold change is 4.

5.2.5.4 Correlation between promoter opening and transcription in the PGCs

Analysis of ATAC peak distribution across genetic elements highlighted that the most of the PGC-specific open chromatin regions overlap gene promoters (Figure 5.10). The depletion of ATAC signal on distal regulatory regions in PGC samples led us to hypothesise that PGC transcription is regulated by promoter-proximal elements, with a greater contribution from promoter accessibility.

To verify if opening of chromatin on promoter regions was linked to differential gene expression between PGCs and somatic cells, we compared the expression fold changes of genes where promoter accessibility was significantly different between PGCs and somatic cells. We

noted a trend showing upregulation in the transcript PGCs for genes with significant more accessible promoter ($log_2FC < 2$, FDR < 0.05) in the PGCs and vice-versa. Interestingly, we observed a higher correlation between chromatin accessibility and gene expression in the somatic cells compared to the PGCs (Figure 5.22). However, the number of promoters included in the analysis was significantly lower in the somatic cells compared to the PGCs, potentially causing numerical bias.

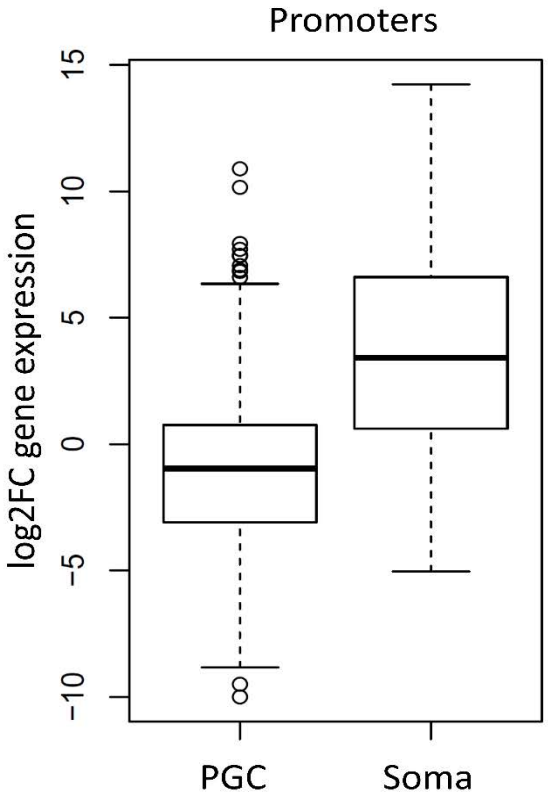

Figure 5. 22: Correlation between open chromatin regions overlapping with promoters and differential gene expression in PGCs and somatic cells.
The boxplot shows the log_2(fold change) of gene expression between PGCs and somatic cells at prim-5 stage for two groups of differentially regulated ATAC peaks. Median is drawn as black lines.

In summary, our data supports the existence of a PGC-specific gene expression regulation, where distal elements (putative enhancers) are organised in compacted regions of chromatin, while gene expression seems to be controlled by promoter-proximal elements. We speculate

that, in order to block transcription of enhancer-regulated developmental genes, compaction of distal regions induces promoter-proximal regulation.

5.3 Discussion

In this chapter, we investigated the open chromatin signatures of pre- and post-migrating PGCs in a germ plasm-dependent vertebrate. As mouse and human PGCs undergo epigenetic reprogramming during migration (Gkountela et al., 2015; Guo et al., 2017), we asked whether this phenomenon is conserved among vertebrates. In combination with transcriptome profiling, the differentiation pathways that the PGCs undergo during migration were studied in great detail and we reported that zebrafish PGCs experience epigenetic reprogramming accompanied by acquisition of specific transcriptional programmes.

Characterisation of the open chromatin profile of a cell type is a powerful tool that provides a vast amount of information about the gene regulation and DNA organisation in general. Currently, techniques such as ATAC-seq and DNase-seq are widely used in order to predict locations of tissue-specific cis-regulatory elements (Quillien et al., 2017). In this study, it was demonstrated that tracing of open chromatin profile by ATAC-seq is a suitable approach for detection of putative enhancer regions in small samples where ChIP-seq is inapplicable. Indeed, here we showed that RNA-seq and ATAC-seq data on zebrafish PGCs complement and support each other. Of note, we choose to apply ATAC-seq due to the limited number of retrievable PGCs from zebrafish embryos. In fact, an advantage of ATAC-seq in respect to the currently available techniques suitable to study open chromatin genome-wide is that transposition and addition of sequencing adaptors occur in one-step (Buenrostro et al., 2013). This generates tagged fragments from nucleosome-free sites ready to be amplified with high efficiently even from very limited cell numbers. In our hands, transposition of sequencing adapters combined with NGS produced mappable and peak-generating data from as few as 1500 cells.

The ability of handling small cell populations and perform genome-wide studies on them is a bottleneck for developmental biology. By achieving good results in isolating and sequencing DNA from very little material, we had the chance to investigate the epigenetic requirements needed to initiate the germ fate. In particular, our work demonstrated the existence of a PGC-specific epigenetic reprogramming for the first time in a germ plasm-dependent vertebrate. Both the transcriptome and the open chromatin profiles evidenced for epigenetic differentiation and germ fate acquisition initiated during migration. Remarkably, we found that different classes of genes are differently regulated in the PGCs and correlated transcriptional output with epigenetic profile. For example, developmental regulators and tissue-specific genes showed significant lower open chromatin in the PGCs when compared to the somatic cells as predicted by gene expression. This indicates that developmental and differentiation programmes are avoided in the migratory PGCs through pre-transcriptional regulation. In mouse, the repression of somatic genes and upregulation of germ line-specific genes was associated with the activity of Blimp-1 (Kobayashi et al., 2017; Tang et al., 2016). Blimp-1 was shown to bind promoters of somatic genes for their inhibition and to co-operate with Prmd14 and Ap2gamma to upregulate genes associated to pluripotency and germ cell-specific genes (Magnúsdóttir et al., 2013). Interestingly, Blimp-1 seems not fundamental for PGC development in zebrafish, indicating that other molecular pathways may be involved (Lee and Roy, 2006). Accordingly, one could hypothesise that transcription factors and/or chromatin remodelers may be contained within or activated by the germ plasm. Of note, as previously indicated, establishment of a PGC-specific chromatin profile occurs alongside the germ plasm re-localisation around the nuclear membrane, indicating for a role of the germ plasm in nuclear regulation and this phenomenon will be better studied in the next chapter.

Finally, the importance of our finding is evidenced by the discovery of unexplored mechanism of chromatin regulation. For example, we observe a tendency for PGC-regulated genes to repress cis-acting region downstream of the TSS (Figure 5.11). To our knowledge, there is no report in the literature of a similar type of regulation. One could speculate that the requirement of upstream regulation of the gene would ensure transcription directionality and strengthen enhancer-promoter interactions (Tan-Wong et al., 2012). Alternatively, lack of cis-regulatory elements downstream of the gene promoter may be a previously-unobserved intrinsic property of genes upregulated in the PGCs. Another possibility would see the existence of a diverse chromatin conformation in the nucleus of the PGCs and it has been shown that regions of open chromatin can coincide with anchors or hubs for chromatin folding (Song et al., 2011). Although proper conformation capture techniques should be applied in order to verify this hypothesis, it known that murine oocytes are characterised by a unique chromatin folding, with total loss of topological associated domains (TADs) (Flyamer et al., 2017; Ke et al., 2017), therefore indicating that similar topological regulation may exist in the embryonic germ cells.

In conclusion, although several questions about gene regulation of the developing germ line in zebrafish remain opened, the reported results led us to suggest that epigenetic profile establishment is achieved in the PGCs of a germ plasm-dependent organism during migration. Importantly, this suggests that the germ plasm does not replace epigenetic reprogramming, but rather cooperates or contributes to it. In fact, the temporal association between germ plasm re-localisation and the onset of epigenetic identity led us to hypothesis that the two events are linked.

In the next chapter, I discuss and explore the possibility that the germ plasm could trigger the acquisition of PGC-epigenetic identity.

6 Tdrd7 is crucial for maintaining stem-like epigenetic and transcriptional program in the PGCs

6.1 Introduction

As demonstrated in the previous chapter, the PGCs gain unique epigenetic features that suggest inhibition of terminal differentiation and retention of pluripotency retention. Also, our chromatin profile analysis over time has indicated that a link between the distribution of cellular germ plasm and the chromatin reprogramming may exist. In fact, the early stages of PGC development are characterised by little variations in total coding-transcriptome compared to the rest of the embryo, and the most dramatic increase in differential gene expression between germ and somatic lines is observed only after migration. We confirmed that the divergence between transcriptomes is not due exclusively to somatic differentiation as the PGC transcriptome also evolves when compared to the previous developmental stage.

Differential gene expression (DGE) analysis over time highlighted that PGCs undergo specific cell fate commitment after the onset of migration. In fact, the number of differentially expressed genes between the PGCs and the somatic cells increases from 327 at dome stage to 6632 at 10-somite stage (Figure 5.1C). In support, open chromatin profile of PGCs shows remarkable differences when compared to the somatic cells after somitogenesis, while no significant differences are detected earlier.

Interestingly, somitogenesis coincides with the time when the germ plasm approaches the nuclear membrane. Although its function is not completely understood, the perinuclear localisation of germ granules is observed in many organisms, such as *D. melanogaster*, *C. elegans* and *X. laevis* (Lehmann, 2016; Taguchi et al., 2012; Updike et al., 2011).

In this chapter, we hypothesised and that the germ plasm might be involved in the acquisition of germ fate through interaction with the nucleus and tested the epigenetic consequences of germ plasm mis-localisation.

6.1.1 Perinuclear localisation of germ granules is required during germ cell formation

As mentioned before, germ granules tend to switch from a cytoplasmic to a perinuclear state during the early phases of embryogenesis in several organisms and previous studies have investigated the nature of germ granules subcellular localisation.

In *D. melanogaster*, the germ granules are brought in proximity of the nuclear membrane after Embryonic Stage (ES5) by *Osk* and are found in perinuclear position by ES10 (post-gastrulation) (Lehmann, 2016). Also in *C. elegans* and *X. laevis*, the germ plasm transitions from cytoplasmic to perinuclear state at the onset of gastrulation (Brangwynne et al., 2009; Pitt et al., 2000; Taguchi et al., 2012). Interestingly, in these species, an association between germ granules and nuclear pores was described (Updike et al., 2011). In this study, injection of labelled dextrant into worm embryos showed that germ granules have the ability to size-select the molecular intake similarly to nuclear pore complexes (NPCs). This feature is highly dependent on hydrophobic interactions and on nucleation of the Vasa-related proteins GLH-1 and -3.

Of note, highly conserved germ granule-localised proteins, such as Vasa, Bucky-ball and Zili, contain phenylalanine-glycine (FG) repeat domains, which were shown to form hydrophobic interactions required for the formation of self-containing cellular structures (Adams et al., 2014). Interestingly, FG repeat domains are often found in NPCs, where they form a molecular 'mesh' filtering the nuclear in- and out-flow (Patel et al., 2007; Ribbeck and Görlich, 2002). Therefore, germ granules physically interact with nuclear pores during germ cells development and generate a hydrophobic microenvironment suitable for selective post-transcriptional regulation of RNAs.

To date, there is no reported association between chromatin reprogramming and perinuclear localisation of the germ granules, mostly because genome-wide experiments were limited to

cell populations of bigger sizes. Moreover, due to the requirement of tagged germ granules for germ cells isolation, it was impossible to remove the maternal germ plasm to verify whether germ granules were causing epigenetic reprogramming. Here, we have tackled this problem by inhibiting the perinuclear re-localisation of the germ granules. This approach allowed us to preserve cellular germ granules, therefore the PGC marker.

6.1.2 Tdrd proteins

Stable aggregation and perinuclear localisation of the germ plasm are also highly dependent on another family of proteins, known as Tudor-Domain (Tdrd) proteins. Tdrd was initially identified in *D. melanogaster* germ cells (Tud) where it was shown to contribute to germ cell development (Boswell and Mahowald, 1985; Golumbeski et al., 1991). Tdrd-related proteins are found in many organisms, where they accomplish similar functions (Pek et al., 2012). The family of Tdrd proteins includes proteins that carry Tudor domains. The Tudor domain consists of 11 repetitive amino acid sequences providing capacity of binding methyl groups on their targets (Sprangers et al., 2003). The Tdrd domain mainly acts through binding the methylated lysines and arginines and is involved in facilitating protein complexes formation by functioning as a protein scaffold (Brahms et al., 2001). The binding of Tdrd proteins to their target is often followed by recruitment of effectors (Cheng et al., 2007), making these proteins very versatile and involved in many different pathways.

Importantly, Tdrd proteins are involved in RNA metabolism (RNA splicing and small RNA pathways), regulation of histone modifications and DNA damage response (DDR) (Amikura et al., 2001; Botuyan et al., 2006; Huang et al., 2011).

6.1.2.1 Role of Tdrd proteins in splicing events

The ribonucleoprotein complex in charge of RNA splicing is known as the spliceosome and its assembly requires recruitment of small-nuclear ribonucleoproteins (snRNPs). These are named U1, U2, U4 and U5 and form the spliceosome core (SM core).

Some Tdrd proteins such as survival motoneuron (SMN) and SPF30 were shown to proactively induce formation of the spliceosome by recruitment of di-methylated subunits. In particular, SMN recognizes two methyl groups on SM ribonucleoproteins and triggers the formation of the SM core. SMN ensures that only correct SM proteins can bind the target snRNAs through sequence-specific recruitment (Fischer et al., 1997; Pellizzoni et al., 1998). Mutations of SMN leads to serious defects of the spinal cord motor neurons and fatal neuromuscular disorders (Gubitz et al., 2004).

6.1.2.2 Small RNA processing

6.1.2.2.1 micro-RNAs

Tdrd proteins are involved in small RNA pathway regulation due to their ability to bind short RNAs.

The so-called micro RNAs (miRNA) are ~22nt long, processed RNAs that take part to the post-transcriptional RNA processing by selectively targeting complementary sequences in the 3'UTR of other mRNAs. These molecules are generated by cleavage of long transcript precursors into 22 nucleotides long miRNAs by the nuclear RNase Drosha (Lee et al., 2003). The miRNAs are then transported to the cytoplasm where they can bind complementary mRNAs. The miRNA-mRNA complex is recognised by the RNA-induced silencing complex (RISC), which induces RNA degradation (Gregory et al., 2005).

The RISC includes a Tdrd-SN protein, containing a Tudor domain and four nuclease domains. Although it was demonstrated that the nuclease activity of the Tdrd-SN subunit is involved in RNA degradation and miRNA processing (Scadden, 2005), the role of the Tudor domain

remains unclear. Interestingly, it was shown that stress conditions induce Tdrd-SN localisation into stress granules, where it promotes their formation (Gao et al., 2010).

6.1.2.2.2 piRNAs

In contrast from the ubiquitous function of Tdrd proteins described before, a germ line-exclusive interaction between Tdrd and piRNAs was reported (Houwing et al., 2007; Vagin et al., 2006). piRNAs are generated from genomic regions harbouring transposons incapable of mobilization and they are often transcribed in antisense orientation. These antisense transcripts are then processed to form piRNAs, which are utilised by the cell as defence against deleterious transposons. piRNA maturation occurs also in the somatic cells, however in the germ line they undergo further levels of maturation. Two main mechanisms leading to piRNA formation are currently known: primary processing and ping-pong amplification (Brennecke et al., 2007; Zhang et al., 2012). While primary processing is observed in both germ and somatic lines, ping-pong amplification is a germ line-unique mechanism and seems to localise in the germ plasm. Both mechanisms of piRNAs biogenesis require Piwi proteins that belong to the Argonaute family; they can be arginine-methylated and targeted by Tdrd proteins (Huang et al., 2011; Kirino et al., 2009).

Primary processing starts with the Piwi-mediated recruitment of immature piRNAs in the nucleus. The Piwi-bound piRNAs are transported into a cytoplasmic structure called the Yb body where it is cleaved on the 5'end by Piwi. The mature piRNAs are translocated back in the nucleus where they promote heterochromatin formation of transposon-harbouring regions.

Ping-pong amplification generates mature piRNAs in the germ line of several organisms. In *D. melanogaster*, the two main players are the Aubergine (Aub) and Argonaute3 (Ago3) proteins, both belonging to the Piwi's family (Kirino et al., 2009; Tóth et al., 2016). Aub can interact with piRNAs and recognize target antisense transposon RNAs through sequence complementarity. The transposon is then cut and the 5' end of the newly generated short RNA

is in sense orientation to the transposon. The processed RNA is loaded onto Ago3, where, following further cleavage, it forms an antisense piRNA. Though this loop, antisense piRNAs are amplified and transposon RNAs targeted more efficiently (Figure 6.1).

Importantly, Piwi, Aub and Ago3 interact with Tdrd proteins (Kirino et al., 2009; Li et al., 2009). The zebrafish Tdrd1 was shown to physically interact with Ziwi (the zebrafish orthologue for Piwi) in a sequence-specific manner (Huang et al., 2011). Notably, Tdrd1 was found associated to a previously unidentified class of RNAs. These were longer than canonical piRNAs and lacked 2′-OMe. Although, it might be possible that the Tdrd1-associated RNAs represent an intermediate product of the ping-pong cycle, the mechanism by which these are processed needs to be further clarified.

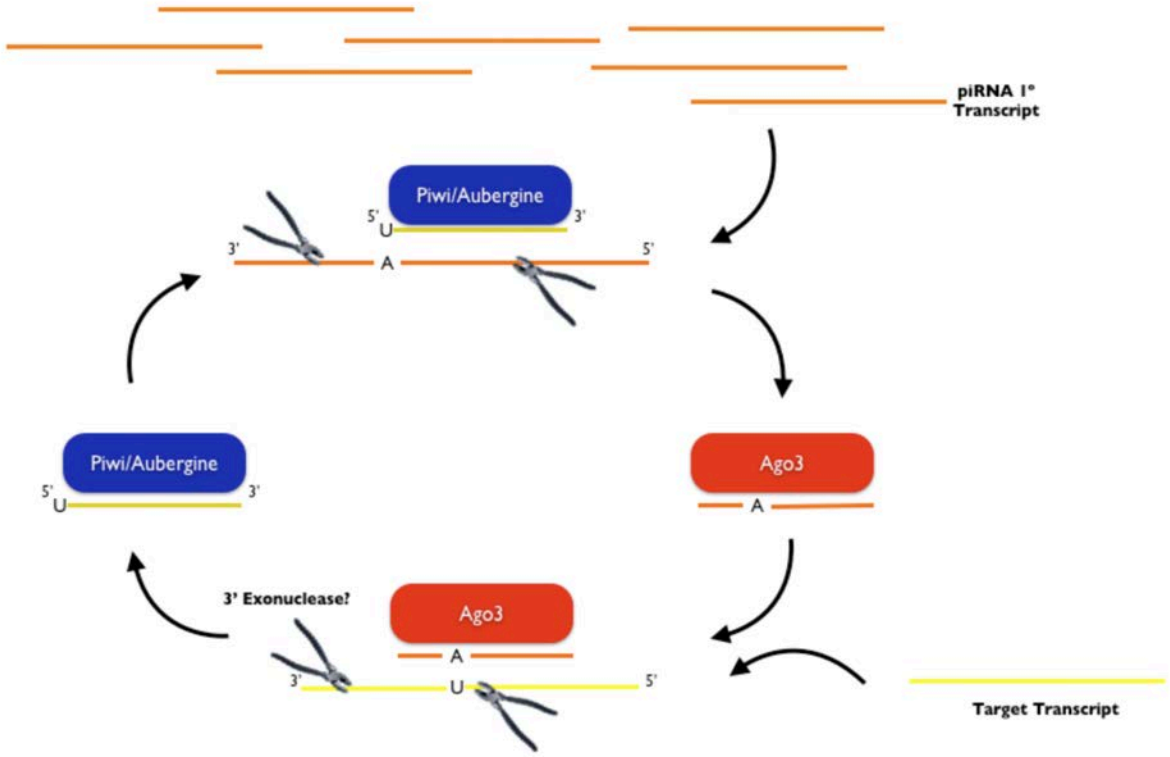

Figure 6. 1: Ping-pong mechanism of piRNA processing.
The piRNA cycle. piRNAs (orange) are bound by Piwi and Aub (blue) and cleaved. The resulting short RNA is transported into the cytoplasm and bound by Ago3 which directs the piRNA on the target RNA (yellow). After cleavage of the target RNA, a novel antisense piRNA is generated and re-processed upon binding of Piwi and Aub. The complex with the antisense piRNA, Piwi and Aub is able to target a newly synthetized sense piRNA. Adapted from https://mcmanuslab.ucsf.edu/node/266.

6.1.2.2.3 Tdrd function in germ plasm aggregation

Amongst the multiple functions that Tdrd proteins perform, a conserved role is to contribute to the aggregation of germ granules. In fact, Tdrd proteins display a high aggregating ability and are often involved where prompt formation of sub-cytoplasmic structures is required. In *D. melanogaster*, *tud* mutants fail to form pole cells and lack germ granules (Boswell and Mahowald, 1985). In zebrafish *tdrd1*, *tdrd6* and *tdrd7* mutant embryos show germ cell loss and inability to aggregate germ granules (Huang et al., 2011; Roovers et al., 2018; Strasser et al., 2008). Of note, when the *tdrd1* gene is knocked out in mature germ cells, the number of germ granules is reduced and they lose association with mitochondria. However, the perinuclear localisation seems to be preserved (Huang et al., 2011).

On the other hand, *tdrd7* transcripts are maternally provided to the embryo and our results suggest that *tdrd7* is also zygotically transcribed. Morpholino-mediated knock-down of *tdrd7* showed that the perinuclear localisation of germ granules in migrating and post-migrating PGCs is dependent on Tdrd7 (Figure 6.2). Interestingly, both the number of germ cells and amount of germ plasm is weakly affected by *tdrd7* KD; however, the germ granules mis-localise and lose perinuclear positioning (Strasser et al., 2008).

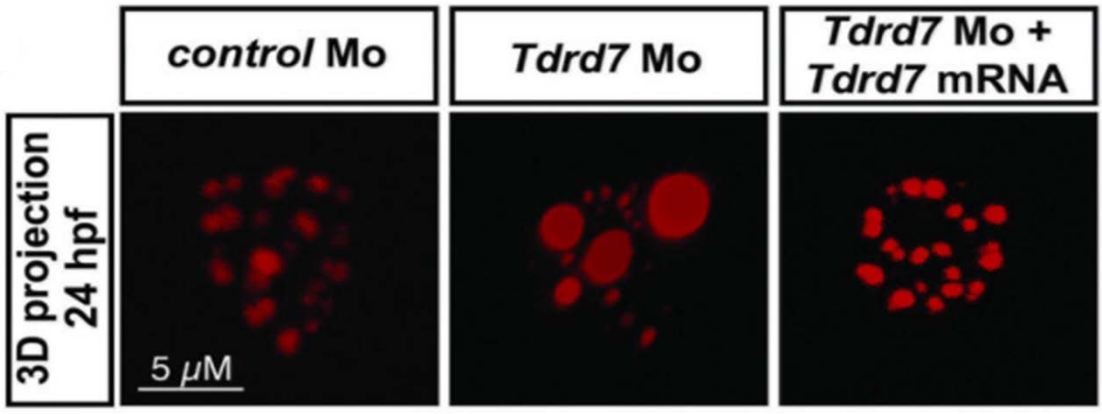

Figure 6. 2: Effect of Tdrd7 KD by morpholino injection and rescue via mRNA supplement. Confocal imaging of a single PGC at prim-5 stage injected with a control morpholino, a morpholino targeting tdrd7 and a combination of morpholino and an exogenous mutated tdrd7 mRNA. Germ plasm is labelled by the gra-RBP RNA injected at the 1-cell stage. From (Strasser et al., 2008).

In both zebrafish and mouse, Tdrd6 is known to aggregate the germ plasm in mature and embryonic germ cells (Roovers et al., 2018; Vasileva et al., 2009). A recent study showed that a direct interaction between Tdrd6 and Buc occurs via symmetrically di-methylated arginine residues. Buc is a germ plasm organiser and its function is dependent on Tdrd6. Lack of Tdrd6 leads to reduced number of germ cells, increased sterility and tendency towards male offspring. Interestingly, the germ plasm of *tdrd6* mutant females is smaller than in wild-type animals during oocyte maturation and showed defective aggregation (Roovers et al., 2018).

Overall, these studies highlight that Tdrd proteins are involved in germ plasm assembly and act in combination with germ plasm organizers (Buc and Osk) and Argonaute proteins (Aub, Ago3 and Piwi). Importantly, their capacity to bind methylated arginines could control aggregation and self-assembly of the germ granules.

In this study, we took advantage of the aggregating capacity of the Tdrd7 protein in order to interfere with the perinuclear localisation of the germ granules and to gain insights into the processes leading to germ cell formation.

6.2 Results

6.2.1 Selection of Tdrd7 as candidate target for germ plasm mis-localisation and experimental rational

During this study, we noted that the most dramatic epigenetic and transcriptional changes in the PGCs occur when the germ plasm becomes perinuclear and coincide with PGC-specific gene expression. In fact, we showed that minimal transcriptional and epigenetic differences are observed between germ and somatic cells before migration begins and that germ cell features are attributable to post-transcriptional regulation. Notably, the germ plasm reshapes within the cell during somitogenesis and distributes around the nuclear membrane. Hence, we hypothesised that the transcriptional and epigenetic signature of germ cells was dependent on the germ plasm localisation. In order to verify to what extend the germ plasm was contributing to the germ fate acquisition, we aimed to interfere with its subcellular localisation.

We took advantage of previously published results and evaluated the best candidate protein to knock down via in-silico analysis. Among the candidates, we selected Tdrd7 for three reasons. First, according to literature, Tdrd7 removal less affects PGC survival in respect to removal of other proteins. Second, within other candidates, *tdrd7* was the highest differentially expressed gene across stages in PGCs against the somatic cells, suggesting an important role of this transcript during early phases of embryogenesis (Figure 6.3A). Finally, analysis of known protein-protein networks showed that Tdrd7 interacts with Piwi proteins and is co-expressed (black line) with the nucleoporins Nup155 and Nupl1 (Figure 6.3B).

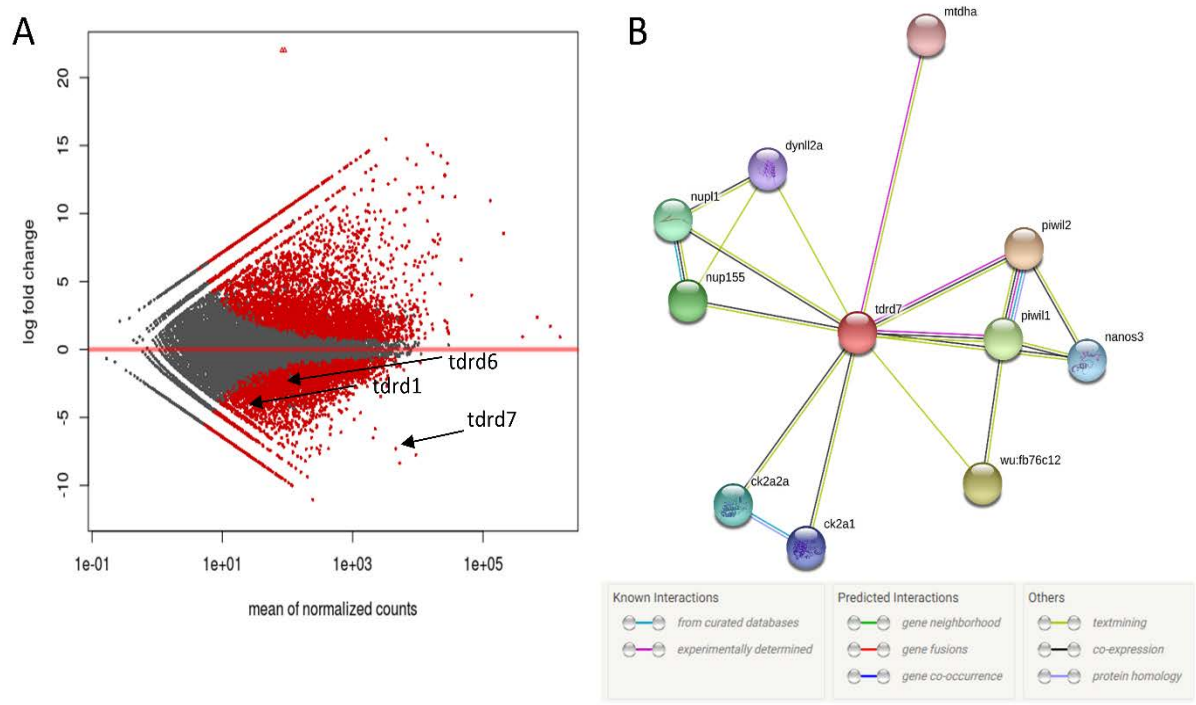

Figure 6. 3: Tdrd7 is a candidate player in triggering perinuclear localisation of the germ plasm. (A) MA plot for genes differentially regulated between PGCs and somatic cells at prim-5 stage. Black arrows highlight *tdrd* genes. p-adjusted < 0.1. (B) Protein interaction network for Tdrd7 in zebrafish visualised by the online tool String (see Methods). Proteins are represented as coloured balls and validated/predicted interactions are shown as connection lines in different colours.

6.2.2 Tdrd7 KD affects germ plasm aggregation but not PGC survival

We first attempted to confirm that lack of Tdrd7 translation results in mis-localisation of the germ plasm in post-migrating PGCs. To test this observation, I exploited published morpholinos (MOs) to target the translation start site of the *tdrd7* mRNA (Figure 6.4A, Table 6.6). As a control, a morpholino with 5 mismatches (5mm MOs) to the target mRNA was also used. About 0.3 pM of each MO per zebrafish embryo at one-cell stage were injected in order to reproduce previously published results (Strasser et al., 2008).

Table 6.6: Morpholinos used for Tdrd7 KD experiment

Name	Target sequence
Tdrd7 MO	AACCAACTCCACGTCACTCATCCTG
Tdrd7 5mm	ACCCAACTGCACGCCACTAATACTG

As expected, upon injections a clear effect on germ plasm localisation was detectable from 10hpf (2-somites stage) (Figure 6.4C). The canonical granular form and perinuclear distribution were replaced by cytoplasmic aggregates of germ plasm. The germ granules appeared more dispersed and disorganised in respect to controls. When the average area of the germ granules was measured, I observed a significant increase upon morpholino injection, which showed germ granules doubling their size on average (Figure 6.4B, left). Alongside, the number of germ granules per cell was reduced, probably due to the aggregation of granules in bigger structures (Figure 6.4C, right). Interestingly, Tdrd7 KD neither significantly affected PGC survival nor germ plasm behaviour at and before migration (data not shown).

This observation was in accord with previous reports, therefore I aimed study the causes and consequences of this phenotype.

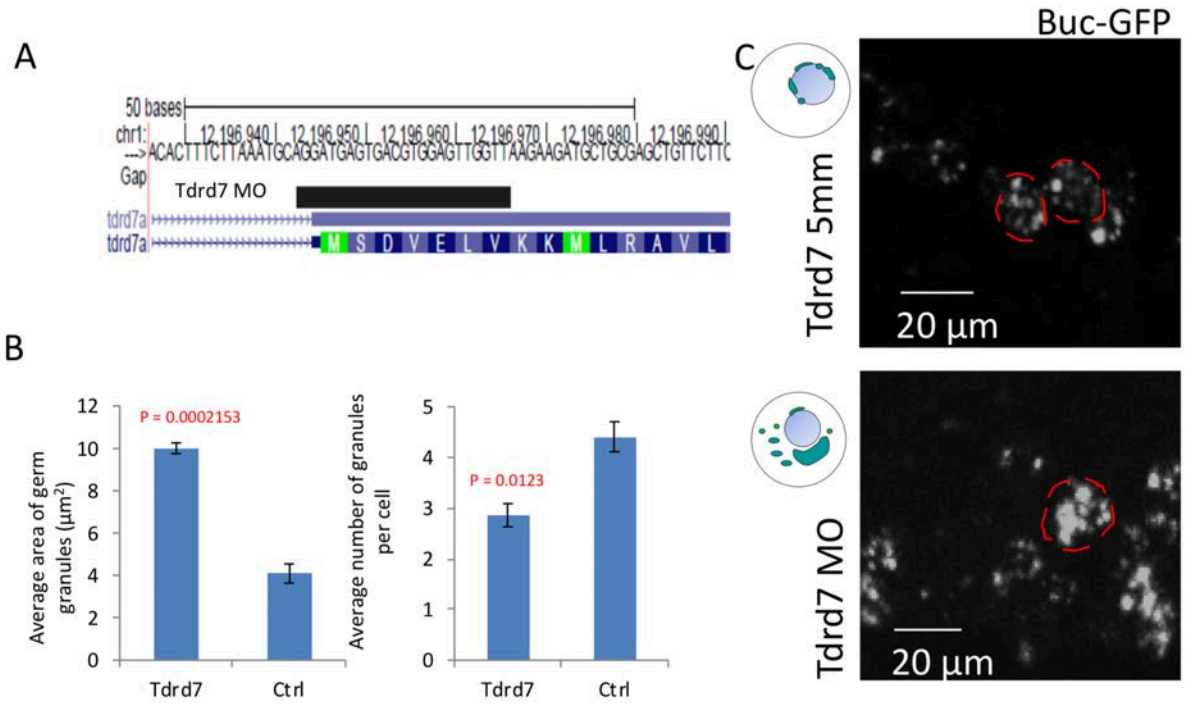

Figure 6. 4: Effect of Tdrd7 KD by morpholino injection.
(A) Genome browser view of *tdrd7* sequence targeted by the Tdrd7 MO (black). The tdrd7 gene is shown in blue. Thick and thin blue lines show exons and introns respectively. The gene underneath highlights the amino acid sequence of the protein. Green sequences indicate translation start sites (ATG). (B) Barcharts for measured average area of the germ granules and average number of germ granules per cells in embryos injected with Tdrd7 MO and Tdrd7 5mm. Bar indicates standard error. Significance was determined by Wilcoxon test and p value is reported in red. (C) Confocal microscopy of Tg(Buc-GFP) embryos at prim-5 injected with Tdrd7 MO and Tdrd7 5mm. Schematic of the phenotype is on the side. Red dashed lines highlight cell boundaries.

6.2.3 Tdrd7 KD affects the transcriptome of late PGCs

In order to verify whether the location of the germ plasm could influence the transcriptome of PGCs and their ability to gain germ fate, I performed RNA-seq on isolated PGCs upon injection of either Tdrd7 MOs or 5mm MOs to the 1-cell stage embryo. Embryos were grown in normal condition for 24 hours and then collected for the experiment. Prim-5 stage (24hpf) was chosen for the experiment due to the following reasons: 1) it provides the best compromise between PGC number and phenotypic effect according to previously published data (Strasser et al., 2008). 2) Comparative and integrative analyses are facilitated by the several datasets that are already available at this developmental stage. 3) The mis-localisation of the germ plasm was described as a late phenotype, therefore we expected less effect on the early stages.

For RNA-seq, we followed the same experimental procedure as throughout the reported work: two hundred GFP-positive cells were sorted alongside GFP-negative cells and cDNA libraries were generated. Illumina-compatible libraries were sequenced in a pool with other developmental stages and analysed together.

Analysis of sample variance by PCA plot showed a clear shift of the Tdrd7-lacking PGC transcriptome towards the somatic transcriptome, which clustered apart from the 5mm MO PGCs (Figure 6.5A). The 5mm MO PGCs appeared to naturally proceed towards germ cell differentiation as shown in Figure 4.5 and 5.1A, while the MO PGCs deviated from the canonical developmental trajectory. Noteworthy, the effect of the MOs on the somatic cells was minimal. This was expected, as there is no reported role for Tdrd7 in somatic cells.

Although the PCA alone already strongly suggested that Tdrd7 was crucial for correct PGC transcription, I looked more closely at the transcriptional changes between control and treated PGCs (5mm MOs vs MOs). Differential expression analysis revealed 4417 genes significantly differentially upregulated and 2786 downregulated (padj < 0.1) by Tdrd7 KD (Table 8.7, 8.8, Appendix). Overall, a total of 7203 genes were differentially regulated in the PGCs upon MO injections, although minimal germ plasm and germ cell loss were observed.

We then asked whether specific developmental pathways were affected and/or induced in MO PGCs. The two sets of differentially expressed genes (up- and down-regulated) were used as input for GO analysis with standard parameters. Interestingly, genes upregulated in the PGCs upon MO injections belonged to somatic developmental pathways (Figure 6.5C, right). Of note, most of these GO categories were associated to ectoderm (head and brain development). On the other hand, genes downregulated in the MO PGCs were mostly assigned to germ cell development and reproduction (Figure 6.5C, left).

In summary, a global analysis of the PGC transcriptome showed that the PGCs tend to lose germ fate-associated transcription, while progressing towards specific somatic development when translation of Tdrd7 is inhibited and the germ granules mis-localised.

At this point, it was tempting to speculate that a correct localisation of the germ plasm around the nuclear membrane could be necessary to ensure correct germ cell developmental program.

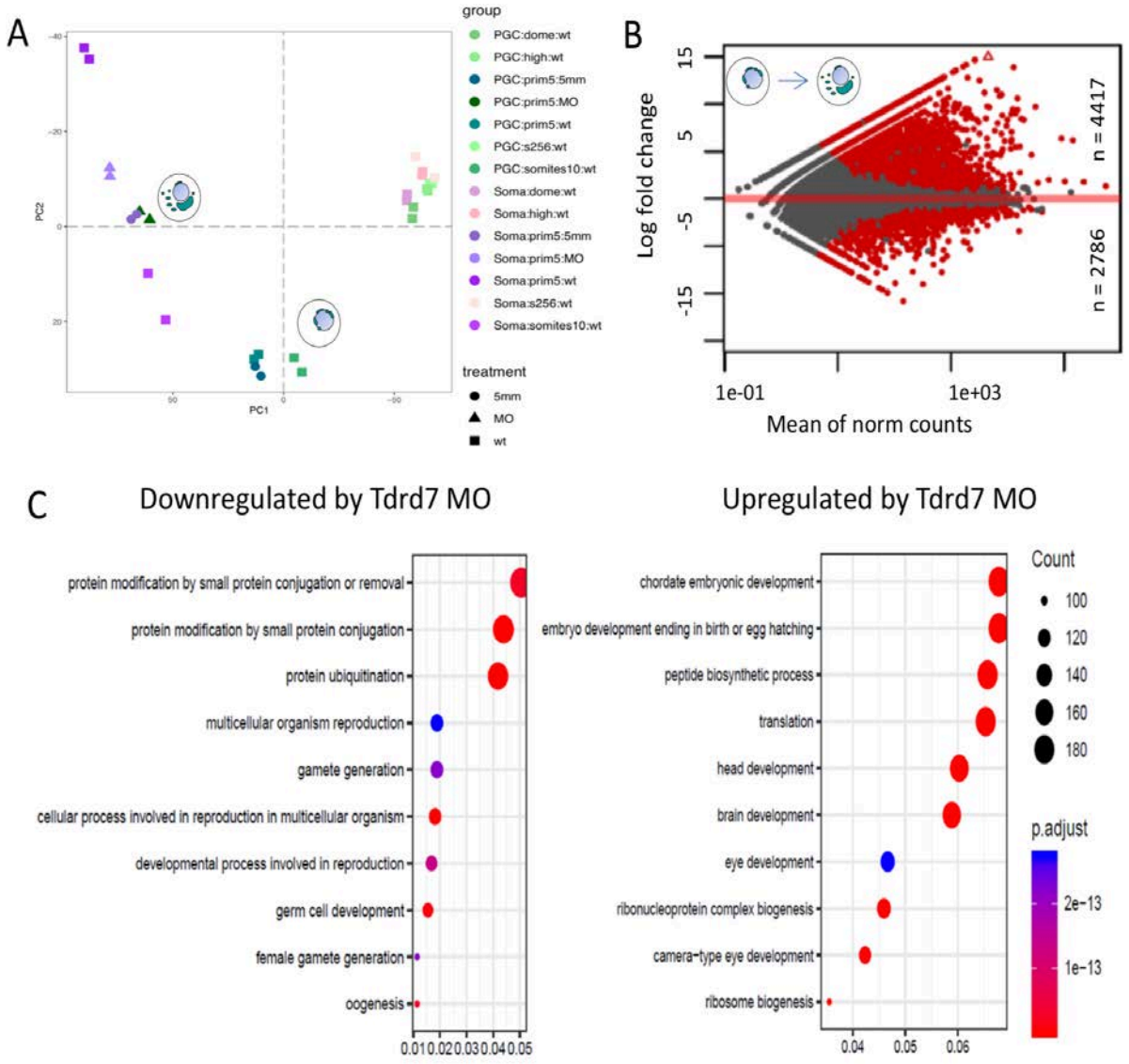

Figure 6. 5: Effect of Tdrd7 KD on the transcriptome of the PGCs.
(A) PCA plot of wt early stages (256-cell, high and dome) and morpholino-injected late stages (prim-5). PC1 variance is 75% and PC2 is 14%. PGC phenotype upon morpholino injection is drawn as a schematic. Circle, triangles and squares indicate samples injected with 5-mismatch MOs, *tdrd7*-tageting MOs and wild types. (B) Volcano plot of $\log_2$(fold change) versus normalised gene expression for genes significantly differentially expressed between PGCs injected with Tdrd7 MO and Tdrd7 5mm. Number of differentially expressed genes is reported on the side. p-adjusted < 0.1. (C) GO analysis for biological processes upregulated and downregulated in the PGCs by the injection of the Tdrd7 MO. p-adjusted < 0.005. Count indicates the number of genes contributing to the GO category.

6.2.4 PGCs retain germ plasm markers but lose germ line-specific transcriptome upon Tdrd7 loss

An initial overview of the transcriptome profile of Tdrd7-lacking PGCs suggested that Tdrd7 and germ plasm localisation are crucial in initiating and/or preserving a correct transcriptional profile. In order to better dissect the transcriptional dynamics of PGCs upon MO injections, I looked in more details at transcription of individual genes.

Initially, we were intrigued by the fact that these germ plasm-carrying cells expressing somatic genes preserve most of the germ plasm mass. Therefore, we asked to what extent Tdrd7 removal was affecting germ plasm-specific transcripts. To compare transcript levels, raw reads were first normalised by gene length and then divided by the library scaling factor (transcripts per million, tpm). Normalised reads of two replicates were merged and the median was compared among samples. Interestingly, we saw a weak, non-significant reduction in levels of germ plasm transcripts (GO:0007281) when MO and 5mm MO PGCs were compared. The difference was instead higher and statistically significant when compared to the two somatic groups (Figure 6.6A). We interpreted this result as a late effect of Tdrd7 loss, affecting zygotic transcription of germ cell genes, but not maternally-deposited RNAs, which are in great excess.

I then investigated the fate of pluripotency-associated transcripts retrieved under the gene ontology term GO:0019827. If PGCs experience somatic differentiation, one could expect that the pluripotency is lost. In accord to our expectation, the levels of pluripotency-associated transcripts were significantly diminished in comparison to the control, and comparable to the somatic cells. Importantly, no significant changes were detected for housekeeping transcripts involved in cellular homeostasis (GO:0030964 and GO:0044391), indicating for a germ line-specific effect of Tdrd7 (Figure 6.6A).

Preliminary analysis of the transcriptome of Tdrd7-lacking PGCs suggested that a somatic fate may have been triggered in the PGCs. As we previously linked germ fate with a PGC-specific epigenetic reprogramming, I looked at the effect of Tdrd7 MO on epigenetic regulators. In order to identify potential chromatin remodelling complexes acting upstream the general epigenetic reprogramming undertaken by the migrating PGCs, I performed an unbiased search for candidates based on literature. In fact, any analysis based on differentially regulated genes between wt PGCs and somatic cells would have highlighted downstream epigenetic events. Noteworthy, components of the Polycomb Repressive Complex 2 (PRC2) did not show significant variation of transcript levels (Figure 6.6, left), as well as genes involved in histone acetylation, DNA methylation and member of the SWI/SNF complex (data not shown). However, two genes involved in the assembly of the Nucleosome remodelling and deacetylase (NURD) complex, *mbd3a* and *mbd3d,* were significantly downregulated by the Tdrd7 MO (Figure 6.6B, right). The NURD complex is known to induce recruitment of transcription factors and nucleosome sliding (Basta and Rauchman, 2015). Intriguingly, inhibition of the subunit Mbd3 disallows induced pluripotent stem cell (iPSC) differentiation and facilitates reprogramming towards fibroblasts (Drozd et al., 2015; Luo et al., 2013), indicating that the complex is required to preserve pluripotent features. Whether the NURD complex is involved in PGC epigenetic reprogramming remains unclear and further studies will be needed in order to prove which role the NURD complex plays during PGC formation.

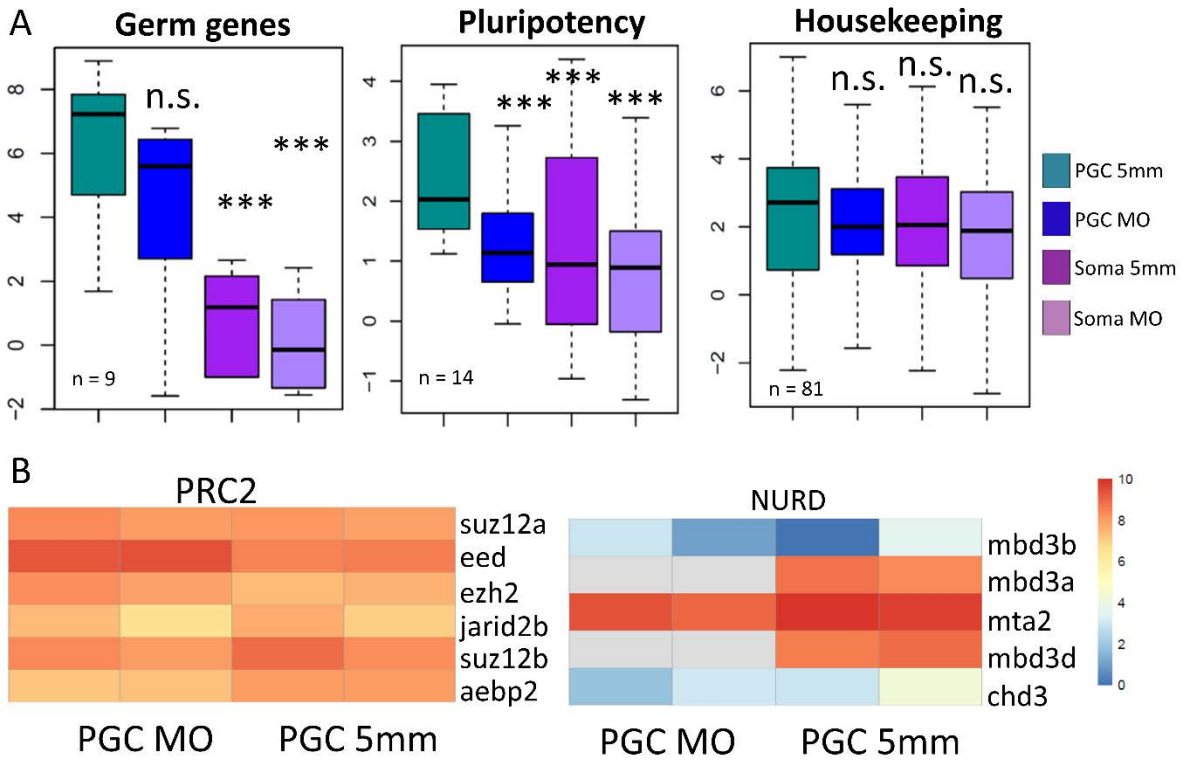

Figure 6. 6: Specific effect of Tdrd7 MO in PGCs.
(A) Boxplots show the median (black line) of normalised gene expression (log_2(tpm)) for selected subclasses of transcripts in PGCs and somatic cells injected either with Tdrd7 MO or Tdrd7 5mm. Statistical significance was determined by Wilcoxon test. p-value < 0.0005 is represented as ***. N.s. = not significant. (B) Heatmap of normalised gene expression (log_2(tpm+1)) for components of chromatin remodelling complexes.

6.2.5 Tdrd7-lacking cells undergo somatic-like epigenetic reprogramming

Having assessed that the lack of Tdrd7 was affecting the normal PGC transcriptional program, we isolated Buc-GFP-positive cells from prim-5 embryos after injection of 5mm MOs and MOs and performed ATAC-seq as described earlier. Our goal was to understand whether the transcriptional changes caused by MO injection were accompanied by somatic epigenetic reprogramming.

Fastqc analysis showed that sequencing generated high quality reads for all the samples, with low GC bias and PCR duplicates (data not shown). The fastq files were therefore aligned to the Zv9 and Zv10 reference genomes and peaks were called from bam files as described before.

To account for significantly enriched peaks over the background, a threshold was set to discard every peak with a fold enrichment lower than 4.

High correlation was observed between non-injected (wt) and 5mm MOs injected embryos for both PGCs and somatic cells (Figure 6.7A). This confirmed that the open chromatin landscape was minimally affected by 5mm MO injections. Interestingly, this analysis also showed a reduced correlation between the 5mm MO and the MO PGCs samples (data not shown).

When hierarchical clustering was performed on filtered peaks, we found that morpholino control PGCs (5mm MO PGCs) clustered with PGCs from non-injected embryos (wt PGCs). However, Tdrd7-lacking PGCs (MO PGCs) clustered apart from the other PGC samples and their global open chromatin status was shifted towards the somatic profile (Figure 6.7B).

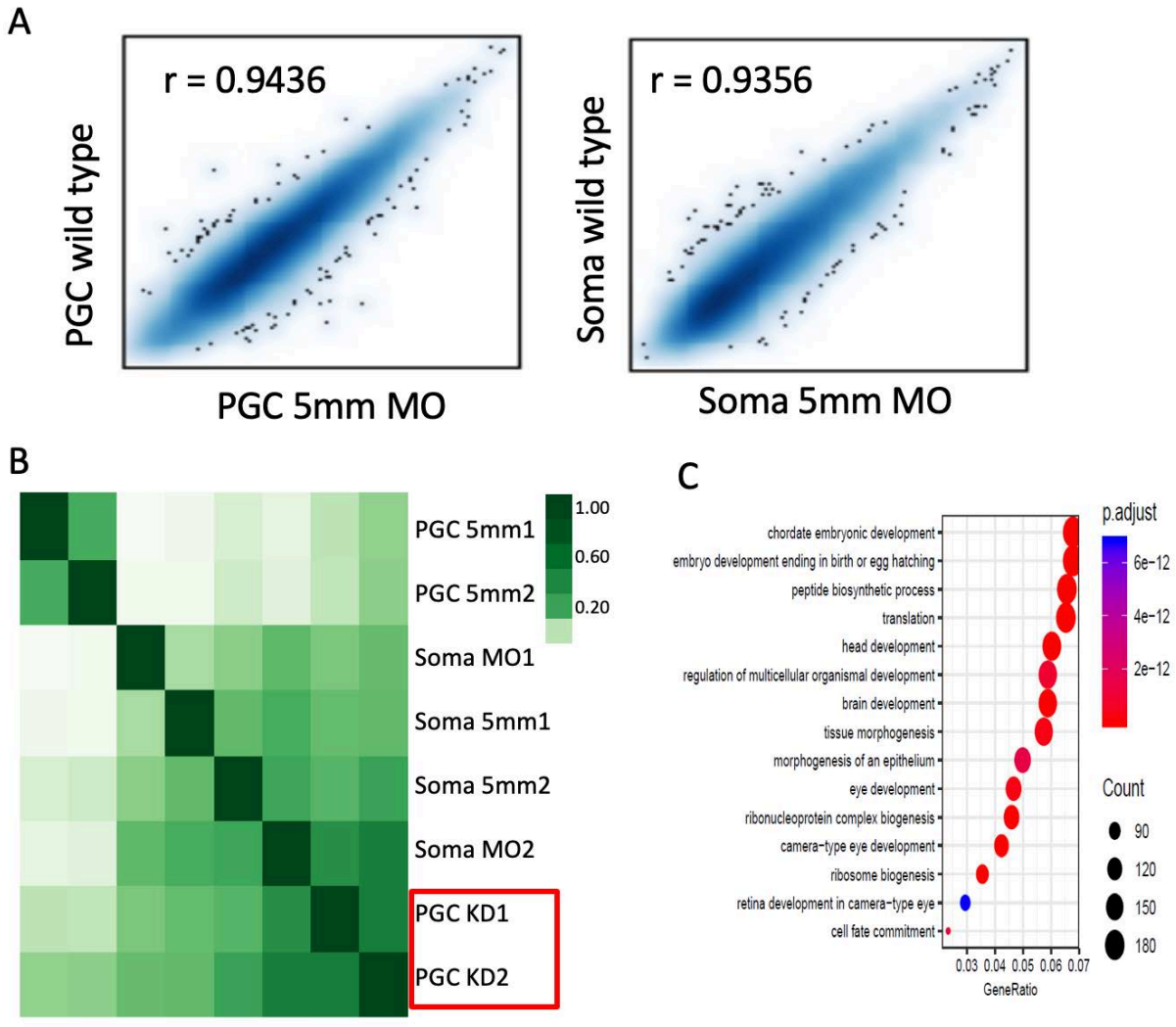

Figure 6. 7: Open chromatin profile of PGCs upon Tdrd7 KD by morpholino injection.
(A) Correlation scatterplots between ATAC-seq injected with Tdrd7 5mm and non-injected PGCs and somatic cells at prim-5 stage show high degree of correlation. (B) Hierarchical unsupervised heatmap of sample variance for morpholino injected PGCs and somatic cells at prim-5 stage. Shades of green indicate the correlation score. The red square highlights the position of the Tdrd7-lacking PGCs which cluster with the somatic samples. (C) GO analysis for genes associated with a more accessible chromatin landscape in the PGCs by Tdrd7 MO. p-adjusted < 0.005. Count indicates the number of Ensembl IDs contributing to the GO category.

When individual genes were inspected, PGCs showed a significant reduction of promoter accessibility on germ line-specific genes upon injection of *tdrd7*-targeting MOs.

I then performed differential openness analysis between MO PGCs and 5mm MO PGCs. When regions significantly more opened in the control were associated with the nearest gene, we saw enrichment for metabolic and cellular processes after GO analysis (data not shown). On the other hand, most of the chromatin accessible regions in the MO PGCs were associated with

developmental pathways (Figure 6.7C). This was confirmed by inspection of the genome browser and complementation with RNA-seq data (Figure 6.8). Importantly, we saw minimal epigenetic effect on housekeeping genes, indicating that the main effect of Tdrd7 removal was restricted to PGC fate (Figure 6.6A and 6.8). The effect of Tdrd7 removal on the PGCs was consistently reproducing somatic-like transcriptional and epigenetic states and we noted that most of the genes affected on the chromatin level resulted differentially transcribed between the two treatments. As expected, we found that MO PGCs gained accessible chromatin at putative enhancers around developmental genes whilst reduced chromatin accessibility of regulatory regions around canonical germ cell-specific genes (Figure 6.8).

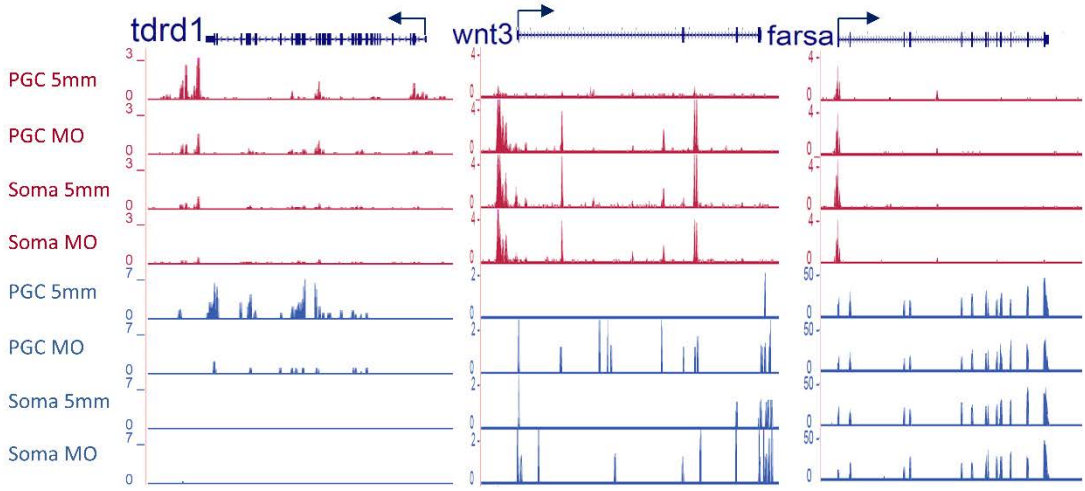

Figure 6. 8: Explicative selected genes and genome browser tracks for MO and 5mm PGCs and somatic cells.
Regions of open chromatin (magenta) and RNA levels (blue) are shown for MO and 5mm PGCs and somatic cells as ATAC- and RNA-seq tracks respectively. A zygotic PGC-specific gene (*tdrd1*), a developmental gene (*wnt3*) and a housekeeping gene (*farsa*) are shown. Arrows indicate position and directionality of transcriptional initiation.

Finally, we asked whether transcriptional output correlated with open chromatin profile. I then analysed the chromatin accessibility state of genes whose transcription is downregulated by the *tdrd7* MO. When the average fold enrichment over the background was plotted for the top 100 genes downregulated by *tdrd7*-targeting MOs, a significant decrease in chromatin accessibility

was detected in the MO PGCs in comparison to the 5mm MO PGCs (Figure 6.9). Importantly, the chromatin accessibility of the MO PGCs was decreased to somatic levels.

Taken together, these data suggest that germ fate is largely dependent on Tdrd7 and on perinuclear localisation of the germ plasm.

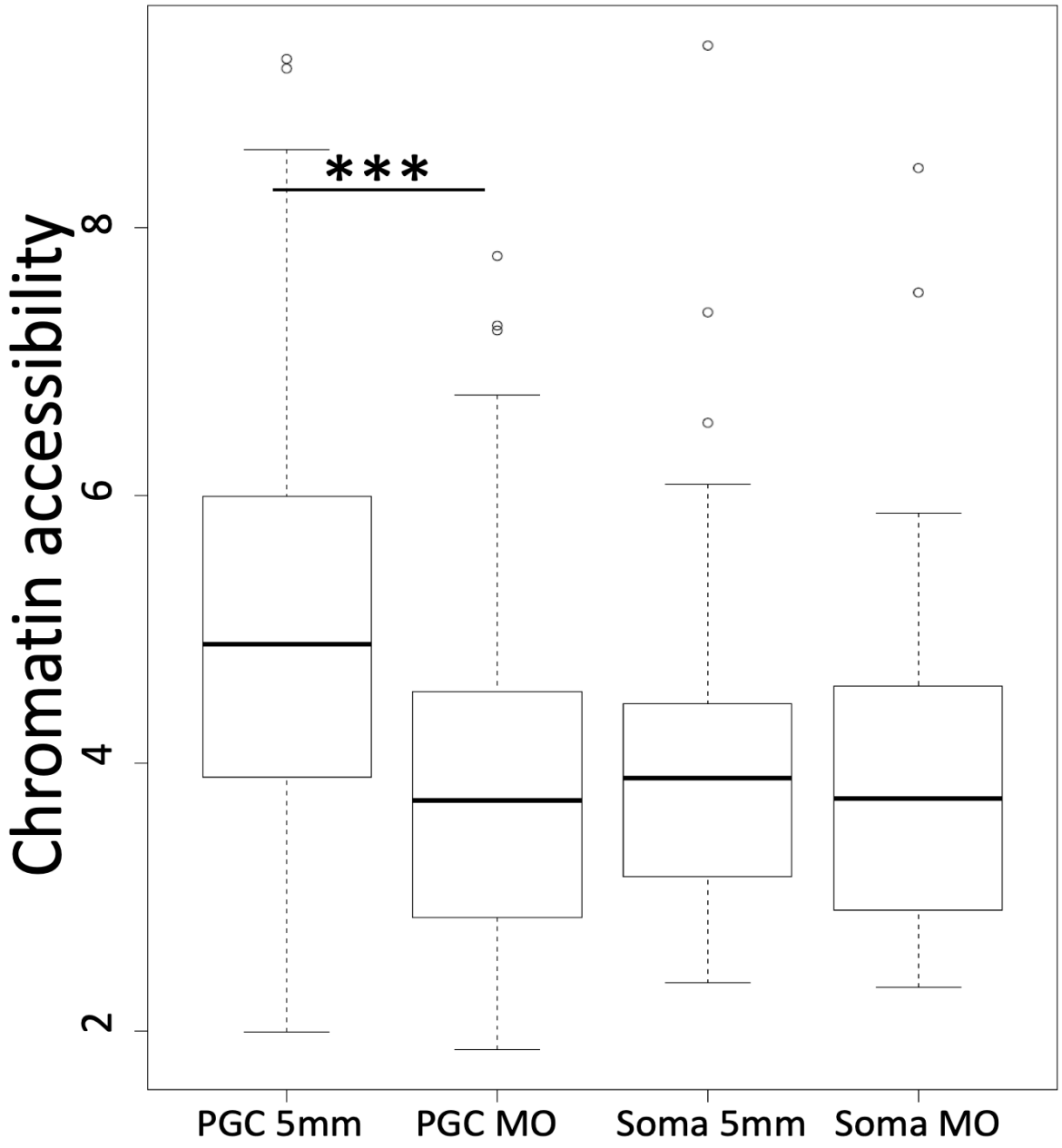

Figure 6. 9: Average chromatin accessibility for differentially transcribed genes upon Tdrd7 knock down.
After DEG, the top 100 genes downregulated by *tdrd7*-targeting MOs were selected and the chromatin accessibility score (fold change over the background) was plotted. p-value < 0.005 is shown as ***, based on Wilcoxon test.

6.3 Discussion

The composition and molecular features of the germ plasm have been intensively studied in the recent years, however the functions associated to subcellular localisations of the germ granules are unknown. Yet, germ plasm-dependent PGCs show re-localisation of the germ granules during early embryogenesis in several organisms, suggesting convergent evolution.

Here, we hypothesised and tested that the perinuclear localisation of the germ plasm and the epigenetic reprogramming of the PGCs are linked. This hypothesis was mainly driven by the observation that the first detected PGC-specific epigenetic signature was coinciding with the movement of the germ granules to the cytosolic side of the nuclear membrane. To verify this, I investigated the chromatin profile and the transcriptome of PGCs where the germ plasm localisation was disrupted by removal of Tdrd7, a germ plasm protein. Importantly, no other significant phenotype is observed in the morphant PGCs, indicating that *tdrd7* is mainly responsible for germ plasm aggregation and re-localisation. In fact, migration occurs normally in Tdrd7-lacking PGCs and PGCs set in the genital ridge at the end of somitogenesis. Moreover, our data suggest that the germ plasm amount is not reduced, but aggregated, perhaps with minimal loss of zygotic transcripts.

Intriguingly, it is important to note that lack of Tdrd7 causes a late phenotype (after migration onset), suggesting that the protein may be superfluous before gastrula stages. As our results and previously published in-situ hybridization studies show, *tdrd7* transcripts are maternally deposited (Strasser et al., 2008), however, it is unknown at what time translation occurs. As antibodies against Tdrd7 are not commercially available, the localisation of the protein in zebrafish has not been studies so far. Nevertheless, blocking *tdrd7* translation results in a germ cell-specific phenotype, indicating importance of the protein (Strasser et al., 2008). In support, many germ plasm-localised factors are maternally provided and translated into proteins at a later developmental stage (Knaut et al., 2000).

Interestingly, in spite of the mild phenotype, we detected remarkable transcriptional and epigenetic changes between Tdrd7-lacking PGCs and the controls. We were able to show that these changes are minimal in the somatic cells, where Tdrd7 is not expected to play a role, while a specific PGC-to-somatic reprogramming is observed.

The interpretation of reported results strongly indicates that Tdrd7 is important for the germ plasm localisation and for the acquisition of the epigenetic germ fate, however it remains unclear which mechanism is epistatic. In our opinion, these results could be caused by two scenarios: 1) Tdrd7 is responsible for the perinuclear localisation of the germ granules, which trigger epigenetic reprogramming. 2) Tdrd7 is involved in the epigenetic reprogramming of the germ line and the germ plasm localisation occurs downstream. For example, Tdrd7 is known to interact with Piwi, a piRNAs processor. As Piwi-mediated piRNAs processing was associated with epigenetic changes in both the somatic and the germ line (Aravin et al., 2008; Durdevic et al., 2018; Huang et al., 2013), it is tempting to speculate about a link between piRNA pathways and germ fate maintenance. Along this line, piRNAs have been associated to regulation of transposon silencing via H3K9me3 and, in general, to regulation of chromatin state on piRNA-target regions (Huang et al., 2013; Sienski et al., 2012). To date, we cannot conclude on whether the germ granules re-localisation is epistatic to the epigenetic germ fate acquisition or vice-versa. However, the multiple lines of evidence associating Piwi and piRNAs to epigenetic regulation strongly suggest a possible requirement of the germ plasm-nuclear interaction in order to trigger initiation of epigenetic germ fate.

In conclusion, the PGCs gain epigenetic identity that is lost when localisation of the germ plasm and Tdrd7 translation are disrupted. We therefore propose a role for Tdrd7 and the germ plasm in the epigenetic reprogramming of the PGCs during migration. Nevertheless, further experiments and validation will be required to decipher the causes and mechanisms of the germ plasm-mediated epigenetic reprogramming of the PGCs.

7 General discussion

This research project focuses on transcriptional and post-transcriptional mechanisms guiding PGC formation during the initial stages of zebrafish development. Generalising, the understanding of specific parental-driven differentiation dynamics in the germ line provides insights and explanations to broader biological interrogations. In fact, germ cell formation in zebrafish requires cooperation between localisation of cytoplasmic aggregates, RNA processing and chromatin rearrangements, therefore resulting an excellent system to study transcriptional and epigenetic regulation.

However, the characterisation of genome-wide mechanisms triggering specification and commitment of cell types has been a challenge until recently. In fact, two colliding factors make similar studies limited: the elevated requirement of biological material for performing NGS studies and the usually small amount of embryonic cells constituting a primordial tissue. In this work, the most advanced techniques in the field of imaging and NGS were applied to study PGC development of zebrafish embryos in great detail. This approach resulted in the identification of two phases of PGC development in which the germ plasm may play two different roles. The germ plasm-mediated protection of maternal RNAs against zygotic degradation was described in the past, however, we suggest that also zygotic transcripts avoid selective removal by translocating into the germ granules. Also, we reported a later, second germ plasm-driven mechanism of PGC specification dependent on Tdrd7. Although this observation opens several new scenarios about the effect of cytoplasmic self-aggregating structures on the epigenome, extra work and experiments will have to be undertaken towards these directions in order to fully understand the mechanisms guiding PGC fate acquisition.

7.1 Relevance of the study

The results presented in this thesis contribute to explain previously unknown biological problems. ZGA progression and triggering, for example, is a largely studied field nowadays; however, several aspects remain uncharacterised. The inheritance of the germ plasm from the oocyte and the generation of germ plasm-carrying cells before ZGA in zebrafish offers a unique system to decipher molecular functions and parental contribution when the embryo faces genome activation. We showed for the first time that zygotic transcriptional activation occurs in the zebrafish PGCs (512-cell stage) via NGS experiments and via a pioneering technique which allows detection of transcription in-vivo. We extensively characterised the transcriptome of pre- and post-ZGA PGCs, reporting potential new germ plasm markers and the very first transcribed genes in the zebrafish germ cells.

Additionally, we demonstrated that, although the presence of the germ plasm allows differential gene regulation, no epigenetic identity is gained prior to migration in the PGCs. This is of great relevance because it suggests that the germ plasm is the only requirement for early germ cell specification. Along this line, our results provide explanations to previous experiments in which germ cells were generated by transplantation of germ plasm in blastomeres. Of note, the observed transcriptional change of the coding transcriptome is minimal between early PGCs and somatic cells, however, germ plasm-carrying cells initiate independent and active migratory movements by dome stage (Blaser et al., 2005; Raz, 2003; Weidinger et al., 1999). As the early migratory movements are triggered by Dnd-mediated loss of cell adhesion (Blaser et al., 2005) and the interaction between the C-X-C chemokine receptor type 4b (Cxcr4b) on the PGC surface and the ligands SDF-1 and Cxcl12a (Doitsidou et al., 2002), one could imagine that epigenetic reprogramming may not be necessary at this point.

Moreover, we provided evidence for germ cell-specific genes transcribed in the somatic cells after ZGA. Generalising, this discovery may heavily impact the field of epigenetic

differentiation during the first phases of embryogenesis. In fact, our results suggest that when the embryonic genome is activated, there is only one default transcriptional programme and that cellular commitment and epigenetic reprogramming is achieved through cytoplasmic regulation exclusively. Along this line, the data presented here demonstrated that the default transcriptional programme includes transcription of germ line genes.

On the other hand, we show that epigenetic commitment occurs in the zebrafish germ line, although previous studies showed that the DNA methylation status remains constant during early embryogenesis (Skvortsova et al., 2019). Therefore, although no epigenetic requirement seems to play a role to trigger PGC migration, a PGC-specific epigenetic reprogramming is observed during migration. The combined use of ATAC- and RNA-seq allowed detection of putative cis-regulatory elements in the developing PGCs and identified potential novel mechanisms of epigenetic regulation, such as lack of downstream elements for genes upregulated. This study provides the first link between germ plasm and epigenetic germ identity, proving that germ factors are not replacing specialised chromatin state, but rather helping its establishment. In fact, we show that the germ plasm factor Tdrd7 plays a crucial role in the PGC epigenetic reprogramming and transcriptional output. As removal of Tdrd7 also leads to germ plasm mis-localisation, this opens interesting scenarios supporting links between perinuclear germ granules localisation and chromatin reprogramming. It remains unclear if the epigenetic phenotype observed in the Tdrd7-lacking PGCs is due to direct function of Tdrd7, detaching of germ granules from the nuclear membrane or both. However, in any of these cases, these data indicate the germ plasm as a main player in the establishment of a PGC chromatin profile.

7.2 Limitations of the work

During the work, many assumptions have been made and future work will be required in order to clarify several questions remaining opened.

As discussed through this thesis, ATAC-seq allows only in-silico enhancer prediction unless combined with ChIP-seq or Chromosome Conformation Capture (3C) techniques. Because none of these techniques was applicable to the limited number of PGCs at the time the project was started, the reported data only investigates candidate enhancer regions. Moreover, we have to acknowledge that predicting the correlation between regions of chromatin accessibility and transcriptional output exposes the analyses to a high degree of inaccuracy. However, currently it is challenging to assign a putative enhancer to its target gene when genomic contacts are not investigated. Overall, these assumptions have been made in the past by the scientific community and have proven to provide reliable results when performed genome-wide (Song et al., 2011; Tsompana and Buck, 2014). As we did use this inaccurate methodology, we carefully interpret these results. In the future, low-cell 3C could be performed on isolated PGCs in order to better assign enhancers to target genes.

Also, limitations in the demonstration that germ plasm localisation is required for epigenetic reprogramming needs to be acknowledged. The experimental design, in fact, lacks an independent validation to demonstrate that germ plasm re-localisation triggers somatic fate acquisition. Although we attempted to search the literature for similar phenotype, we failed in identifying agents causing detaching of germ granules from the nuclear membrane without affecting overall the number of PGCs. This would have enforced the assumption that germ plasm re-localisation is required for epigenetic reprogramming. Due to lack of independent validations, we highlight the need in the future to confirm this observation via alternative means.

Finally, throughout this thesis, it needs to be considered what is defined as somatic cells. The somatic epigenetic and transcriptional profiles are constituted by the combination of several cell types composing the embryo. At the early blastula stages, one could assume that most of the somatic cells share similar transcriptional programmes, as also proven in Chapter 4.2.4. However, after gastrulation cell commitment begins and several cell types are observed within the embryos. GO analysis of transcriptomic and epigenetic data suggested that the greatest contributors to what is called somatic cells in this dissertation is muscular tissue. This is due to the fact that these cells make up most of the mass in a prim-5 stage zebrafish embryo. Therefore, keeping in mind that epigenetic and transcriptional somatic profiles are probably reflecting mesoderm-specific pathways, we speculate that PGCs undergo mesodermal fate acquisition when Tdrd7 is removed.

7.3 Future perspective

This work aimed to provide a general overview of specific modes of germ fate acquisition and germ cell formation relying on germ plasm factors. Some of the reported results are pioneering and ground-breaking, therefore require extra experiments to completion.

Among these, the PGC-specific epigenetic regulation has been only superficially looked at. Beyond the presented data, one could hypothesis that PGC-specific cis-regulatory elements may supplement diverse functions. Moreover, it was described that putative enhancers are less represented downstream the gene promoter in PGCs compared to the somatic cells and that germ line-specific genes are less affected by distal cis-regulatory elements. Taken together, these observations suggest that the global chromatin organisation of PGCs is remarkably different from other cell types. One could propose to perform Hi-C on isolated PGCs in order to look at chromosomal interactions on a genome-wide scale. Although this was inapplicable when the project started, new optimised protocols allow nowadays to perform Hi-C on very limited amount of cells (Díaz et al., 2018).

Additionally, as Tdrd7 interacts with Piwi, a potential role of piRNA in the epigenetic reprogramming of the germ line could be expected. Accordingly, investigation of the effect on the epigenome upon Piwi mutation or co-depletion of Tdrd7 and Piwi might confirm the involvement of the piRNA processing in establishing PGC epigenetic landscape.

Finally, the use of morpholinos has been criticised by the research community for the numerous side effects reported and the non-penetrant phenotype (Gentsch et al., 2018). In order to clarify the function of *tdrd7* gene, conditional zebrafish KDs should be generated. In fact, KD of *tdrd* genes have shown problems during germ cell formation and the resulting mutant adults have

high degree of sterility (Huang et al., 2011; Roovers et al., 2018). For these reasons, it was implausible to generate mutants able to produce the offspring that a NGS experiment requires.

7.4 Conclusions

In summary, the presented results allowed us to predict a functional model of germ plasm-driven PGC formation. As in zebrafish the initial position of the germ plasm does not matter in terms of migration, one could hypothesise that the initial acquisition of the germ plasm is random. The germ masses are about four at the beginning of development because these are pulled apart by the initial orthogonal cleavages. These four main masses seem to remain stable if reach a certain size, while small granules are defused and cleared. At the time ZGA starts, early zygotic transcripts are translocated into the germ plasm and allow acquisition of migratory behaviour and onset of independent movements. At this stage, no evidence for differential transcription is detected.

As the germ cell migrates, the germ plasm approaches the nuclear membrane and specific germ cell epigenetic establishment is observed. This phenomenon could be either initiated by epigenetic regulators/transcription factors brought selectively into the nuclei of germ plasm-carrying cells or via alternative mechanisms of nuclear inflow and/or outflow regulation. The observed epigenetic reprogramming corresponds to acquisition of a specific transcriptional programme, where developmental and somatic genes are inhibited (Figure 7.1).

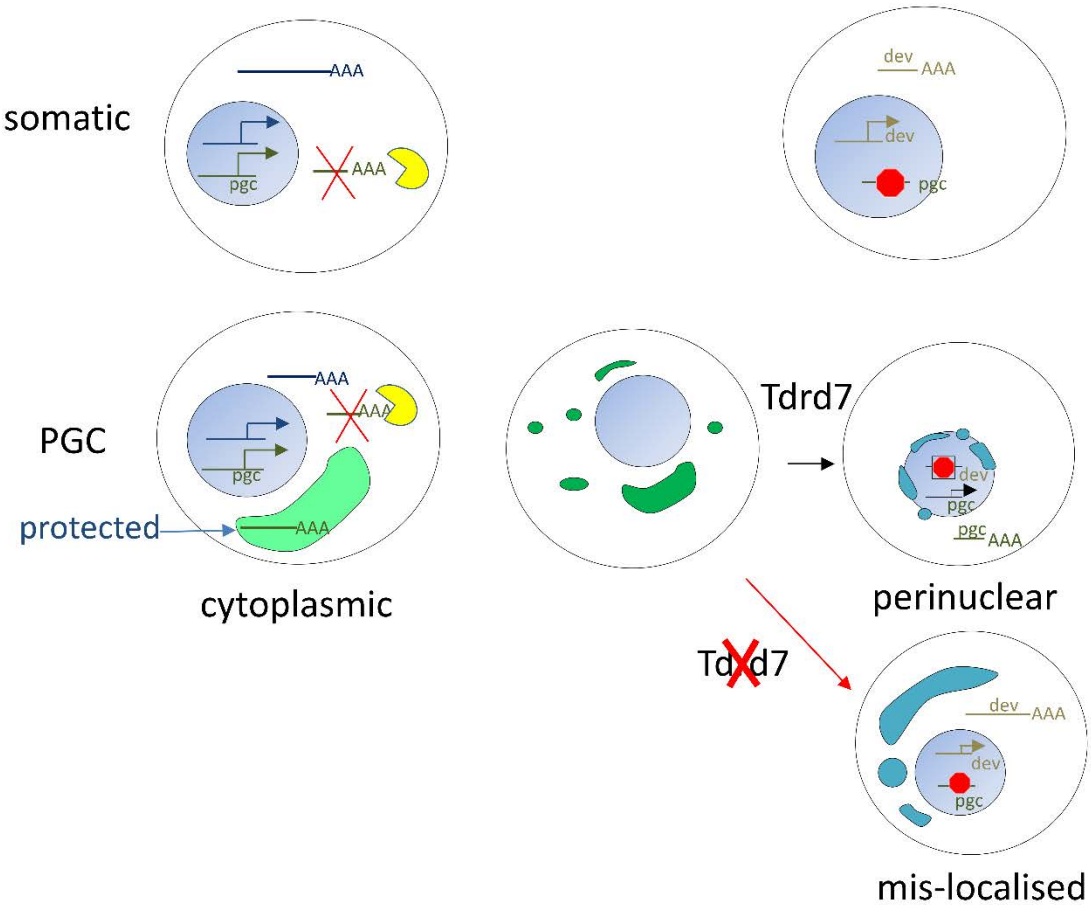

Figure 7. 1: Model of germ plasm-driven germ cell formation.
Prior to ZGA, the germ plasm is a single cytoplasmic aggregate, which asymmetrically segregates at cell division. When the germ plasm is cytoplasmic, zygotic and maternal RNAs are accumulated and protected by zygotic degradation driven by *miR-430* (yellow degradation machinery). The subsequent perinuclear localisation of the germ plasm allows chromatin rearrangement and consequent inhibition of developmental programmes. Developmental genes, transcribed in the somatic cells, are repressed in the PGCs. The somatic fate is re-established when the germ plasm is mis-localised by Tdrd7 removal.

8 Appendix

Table 8.7: Upregulated genes in PGCs vs somatic cells at 256-cell stage

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSDARG00000051768	387.189066	-2.0138388	0.38308538	-5.2568929	1.47E-07	0.00018061
ENSDARG00000037182	895.907204	-2.4730044	0.32477415	-7.6145358	2.65E-14	5.13E-11
ENSDARG00000062511	514.117945	-2.1934098	0.34824093	-6.2985412	3.00E-10	4.53E-07
ENSDARG00000099235	604.245335	-2.8370147	0.36164893	-7.8446651	4.34E-15	9.81E-12
ENSDARG00000104225	1106.3358	-1.7572373	0.34351541	-5.1154541	3.13E-07	0.00032647
ENSDARG00000053939	340.66871	-1.9121317	0.44675292	-4.2800652	1.87E-05	0.01490312
ENSDARG00000055900	292.31783	-1.7954258	0.42717034	-4.2030675	2.63E-05	0.01983693
ENSDARG00000032808	4198.46602	-3.9501965	0.30948087	-12.763944	2.61E-37	8.84E-34
ENSDARG00000078754	5981.75062	-3.5125349	0.92067589	-3.8151698	0.00013609	0.0838805
ENSDARG00000043963	6642.12397	-5.4739248	0.34546909	-15.844905	1.52E-56	2.07E-52
ENSDARG00000010301	1058.75521	-1.4720763	0.39191076	-3.7561518	0.00017255	0.0974885
ENSDARG00000053713	358.649471	-1.495763	0.39157862	-3.8198281	0.00013354	0.0838805
ENSDARG00000021521	1022.87067	-2.6349567	0.34937402	-7.5419366	4.63E-14	7.85E-11
ENSDARG00000022813	4322.15347	-4.2676495	0.32342142	-13.195321	9.34E-40	4.22E-36
ENSDARG00000079591	667.241714	-1.5429763	0.39955529	-3.8617342	0.00011259	0.08035014
ENSDARG00000040510	27326.1335	-4.042618	0.42986438	-9.4044033	5.23E-21	1.42E-17
ENSDARG00000071395	378.016118	-1.5249292	0.40372416	-3.7771561	0.00015863	0.09352232
ENSDARG00000036214	1537.0698	-2.2429398	0.42627903	-5.2616705	1.43E-07	0.00018061
ENSDARG00000102591	91.1954208	-3.2268119	0.61915554	-5.2116337	1.87E-07	0.00021152
ENSDARG00000062699	225.757507	-2.1878447	0.46260723	-4.7293784	2.25E-06	0.0021813
ENSDARG00000099967	106.771418	-3.0723036	0.6992227	-4.3938843	1.11E-05	0.00943633
ENSDARG00000014373	5188.99005	-4.1465014	0.30876773	-13.429193	4.08E-41	2.77E-37
ENSDARG00000038428	199.933501	-1.9870387	0.45152477	-4.4007301	1.08E-05	0.00943633

Table 8.8: tpm values for ERCC-spiked transcripts in PGCs and somatic cells

	PGC 256-cell	Soma 256-cell	PGC high1	PGC high2	Soma high1	Soma high2
ERCC-00002	97.8140965	133.511327	86.8513634	93.276232	62.9010886	165.250273
ERCC-00003	6.37938313	9.71181835	8.5829834	7.18507969	15.4821887	13.6389698
ERCC-00004	149.3767	28.3984645	153.051327	144.521003	36.5432896	26.3675019
ERCC-00009	11.2521857	8.68587865	11.825251	12.6650028	4.79656189	10.5569579
ERCC-00012	0	0	0	0	0	0
ERCC-00013	0.02231179	0.01075204	0.00264142	0.00822014	0.01447075	0
ERCC-00014	0.00460601	0.00710283	0	0	0.00149366	0.0048833
ERCC-00016	0	0	0.00505751	0	0	0

ERCC-00017	0	0	0	0	0	0
ERCC-00019	0.36858336	0.07824276	0.21541558	0.46066803	0.02572078	0.02967895
ERCC-00022	2.58056187	3.7619467	2.16269033	2.52939404	1.95910641	4.16113839
ERCC-00024	0	0	0	0	0.00181784	0
ERCC-00025	0.08890401	0.05663961	0.13700437	0.12990619	0.03127345	0.05271957
ERCC-00028	0.00797696	0	0.01510989	0.00783702	0.00086227	0
ERCC-00031	0.01056117	0.01526827	0.00937729	0	0.00342483	0.00839773
ERCC-00033	0	0	0.0042221	0.0032848	0.00144564	0
ERCC-00034	0.0678185	0.04774368	0.07330667	0.06518028	0.02772968	0.07971666
ERCC-00035	0.46000458	0.41669941	0.49862621	0.6191242	0.19745951	0.4355451
ERCC-00039	0.02030172	0.03756819	0.03749395	0.02393467	0.01975061	0.0258287
ERCC-00040	0	0.00233539	0	0.00892725	0	0
ERCC-00041	0	0	0.0019022	0.00197322	0	0.00425874
ERCC-00042	3.51277266	2.01437855	4.19552252	3.62933091	3.77839696	2.96600883
ERCC-00043	7.2840102	14.0174251	10.0955412	10.3361267	17.0470756	14.6151837

Table 8.9: Top 20 upregulated genes in the PGCs vs somatic cells at prim-5 stage

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSDARG00000008541	247.824762	-11.068457	1.51065777	-7.3269124	2.36E-13	6.08E-12
ENSDARG00000094037	119.206869	-10.012996	1.55021919	-6.4590841	1.05E-10	1.86E-09
ENSDARG00000087740	118.318932	-10.002198	1.55054814	-6.4507498	1.11E-10	1.96E-09
ENSDARG00000070668	117.961718	-9.9976889	1.54898828	-6.4543348	1.09E-10	1.92E-09
ENSDARG00000012718	114.686904	-9.9564676	1.55113211	-6.4188392	1.37E-10	2.38E-09
ENSDARG00000076685	104.483653	-9.8229465	1.56212126	-6.28821	3.21E-10	5.24E-09
ENSDARG00000091584	91.7900994	-9.6366468	1.59197744	-6.0532559	1.42E-09	2.10E-08
ENSDARG00000090817	83.6961539	-9.5027267	1.57857249	-6.0198228	1.75E-09	2.54E-08
ENSDARG00000001656	81.8032607	-9.4685032	1.58756792	-5.9641563	2.46E-09	3.48E-08
ENSDARG00000089237	79.7581649	-9.432221	1.58519576	-5.9501932	2.68E-09	3.76E-08
ENSDARG00000102588	76.9029315	-9.3793024	1.59570191	-5.8778537	4.16E-09	5.62E-08
ENSDARG00000025670	75.5988106	-9.3544024	1.60498643	-5.8283373	5.60E-09	7.37E-08
ENSDARG00000045306	204.481299	-9.3477905	1.4640244	-6.3849964	1.71E-10	2.92E-09
ENSDARG00000069133	72.1897035	-9.2891946	1.59271996	-5.8322837	5.47E-09	7.21E-08
ENSDARG00000089131	68.5277303	-9.2134333	1.59989491	-5.7587741	8.47E-09	1.08E-07
ENSDARG00000089359	67.7534124	-9.1980281	1.60344616	-5.7364122	9.67E-09	1.21E-07
ENSDARG00000093704	66.8654746	-9.1789721	1.604683	-5.7201155	1.06E-08	1.32E-07
ENSDARG00000097662	66.8176196	-9.1778675	1.60385805	-5.722369	1.05E-08	1.31E-07
ENSDARG00000068876	60.5029473	-9.0336933	1.61610833	-5.5897821	2.27E-08	2.68E-07

Table 8.10: Top 20 upregulated genes in the somatic cells vs PGCs at prim-5 stage

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSDARG00000032976	91.6712438	23.0405598	4.78549082	4.81467014	1.47E-06	1.21E-05
ENSDARG00000090615	83.1982574	22.9055232	4.78556789	4.78637516	1.70E-06	1.37E-05
ENSDARG00000100630	3304.54054	15.5057817	1.4935183	10.38205	2.99E-25	2.50E-23
ENSDARG00000079305	14294.2248	15.0835688	1.59337817	9.4664086	2.90E-21	1.73E-19
ENSDARG00000090689	1833.19145	14.6558573	1.93458949	7.57569364	3.57E-14	1.01E-12
ENSDARG00000089475	1680.8501	14.5309695	1.53593409	9.46067255	3.06E-21	1.81E-19
ENSDARG00000089124	17384.6902	14.4906839	0.89165152	16.2515103	2.18E-59	1.34E-56
ENSDARG00000019365	1617.53583	14.4749484	1.53800037	9.41153764	4.89E-21	2.82E-19
ENSDARG00000089963	22387.3385	14.2280953	0.75969994	18.7285722	2.90E-78	3.05E-75
ENSDARG00000089086	1321.80479	14.1839657	1.48304177	9.56410398	1.13E-21	7.08E-20
ENSDARG00000087390	7086.42444	14.0734119	1.57648051	8.92710803	4.37E-19	2.09E-17
ENSDARG00000101407	1147.17168	13.9794519	2.01919113	6.92329303	4.41E-12	9.53E-11
ENSDARG00000006588	1125.74754	13.9519558	1.5045362	9.27326025	1.81E-20	1.01E-18
ENSDARG00000004748	950.640569	13.7079251	1.53841562	8.91041723	5.08E-19	2.42E-17
ENSDARG00000038147	28981.508	13.6897505	0.72363554	18.9180183	8.10E-80	1.06E-76
ENSDARG00000028098	933.978785	13.6828134	1.97341978	6.93355442	4.10E-12	8.91E-11
ENSDARG00000060682	9554.63459	13.6066121	0.83642942	16.2674958	1.68E-59	1.06E-56
ENSDARG00000039499	837.03268	13.524313	1.50746005	8.97158966	2.92E-19	1.43E-17
ENSDARG00000105590	2308.73939	13.5132909	1.88421013	7.17185983	7.40E-13	1.78E-11

Table 8.11: Top 20 upregulated ATAC peaks in PGCs vs the somatic cells at prim-5 stage

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
chr24:29662139-29662274	25.5297132	-8.3220809	1.64891371	-5.0470081	4.49E-07	9.68E-06
chr7:34836272-34836447	23.9371114	-8.2182827	1.63306342	-5.0324333	4.84E-07	1.03E-05
chr7:33208089-33208306	35.1914151	-7.3218638	1.59151168	-4.6005718	4.21E-06	5.96E-05
chr23:30182192-30182275	18.8298732	-6.3997494	1.64806208	-3.8831968	0.00010309	0.00079641
chr8:29871729-29871963	54.1876916	-6.3717958	1.02582827	-6.211367	5.25E-10	4.01E-08
chr21:29715632-29716088	70.0522339	-6.3023359	0.93917561	-6.7104978	1.94E-11	2.73E-09
chr20:20416923-20417207	43.6794671	-6.0126374	1.02420584	-5.8705361	4.34E-09	2.19E-07
chr6:14034680-14034807	13.3961001	-5.928412	1.72152326	-3.4437013	0.00057381	0.00314117
chr19:28693170-28693303	35.0808594	-5.7235047	1.05436432	-5.4283938	5.69E-08	1.78E-06
chr7:42090900-42091104	22.5832755	-5.7020717	1.30207151	-4.3792309	1.19E-05	0.00013853
chr23:38284117-38284420	51.2396302	-5.5349146	0.85215521	-6.4951954	8.29E-11	9.12E-09
chr18:48947851-48948071	20.8602843	-5.5258351	1.30885522	-4.2218842	2.42E-05	0.00024662
chr4:35231303-35231396	20.5672551	-5.4981228	1.30165452	-4.2239493	2.40E-05	0.00024491
chr20:20186083-20186168	9.64452408	-5.4514259	1.81959068	-2.9959628	0.0027358	0.01083621

chr8:41001204-41001382	29.4972685	-5.4340688	1.15087198	-4.721697	2.34E-06	3.70E-05
chr1:23030455-23030550	9.11376132	-5.3644359	1.83709115	-2.9200706	0.00349952	0.01311229
chr17:32912276-32912373	8.8378539	-5.2812197	1.84664808	-2.8598951	0.00423781	0.01520869
chr5:42807889-42807964	8.36338341	-5.232854	1.87218394	-2.7950533	0.00518912	0.01777235
chr4:22033762-22033875	16.1031615	-5.1601668	1.35940065	-3.7959131	0.0001471	0.00106139

Table 8.12: Top 20 upregulated ATAC peaks in the somatic cells vs PGCs at prim-5 stage

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
chr4:2046860-2047115	50.6735966	8.9637288	1.58068723	5.67077954	1.42E-08	5.72E-07
chr18:17465555-17465834	41.6867864	8.68240411	1.6071233	5.40245052	6.57E-08	1.99E-06
chr2:45428178-45428395	40.8258923	8.65000046	2.48493659	3.48097433	0.00049959	0.00282092
chr16:31959006-31959429	31.2951608	8.26391526	1.60929212	5.13512443	2.82E-07	6.60E-06
chr25:16993769-16993943	28.7527936	8.14657814	1.6350504	4.98246301	6.28E-07	1.27E-05
chr8:49910277-49910492	26.9151355	8.050742	1.62927561	4.9413015	7.76E-07	1.50E-05
chr7:55483581-55483831	48.0285139	7.46618802	1.51923383	4.91444298	8.90E-07	1.68E-05
chr23:33544503-33544702	17.152947	7.39758423	1.69486991	4.36469147	1.27E-05	0.00014579
chr19:18703775-18704015	16.5387342	7.34595562	1.70039515	4.32014619	1.56E-05	0.00017241
chr4:8767301-8767620	41.4683896	7.25459315	1.53752876	4.71834631	2.38E-06	3.75E-05
chr5:68989099-68989290	14.6658359	7.17065014	1.73782285	4.12622618	3.69E-05	0.00034703
chr7:26109064-26109348	69.9525107	7.149881	1.87451992	3.81424648	0.0001366	0.00100006
chr16:25345595-25345930	39.4164676	7.14302778	1.53008039	4.66840031	3.04E-06	4.58E-05
chr7:39332063-39332302	37.6320024	7.11486556	1.55313838	4.58096049	4.63E-06	6.43E-05
chr9:11685772-11685956	13.5028159	7.04941555	1.80659052	3.90205499	9.54E-05	0.0007482
chr11:23423939-23424218	65.5270013	6.95804354	1.20682159	5.76559418	8.14E-09	3.65E-07
chr3:15716473-15716729	76.5875249	6.91537103	1.86478143	3.70840836	0.00020857	0.00140369
chr5:39971262-39971553	32.1442131	6.8844109	1.53757838	4.47743736	7.55E-06	9.59E-05
chr10:42236892-42237107	11.8263231	6.86149409	1.7876148	3.83835157	0.00012386	0.00092278

Table 8.13: Top 20 downregulated genes in the Tdrd7-lacking PGCs vs control at prim-5 stage

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSDARG00000005749	11.2020366	-7.0053228	1.8085522	-3.8734424	0.00010731	0.00071588
ENSDARG00000053630	7.18018338	-6.3628441	1.97967193	-3.2140902	0.00130859	0.00632
ENSDARG00000013787	7.14678567	-6.3566434	1.98259295	-3.2062272	0.00134488	0.00646923
ENSDARG00000041660	6.53133675	-6.2284818	2.03852039	-3.0553934	0.00224766	0.01005577
ENSDARG00000103179	6.32380542	-6.1811477	2.0471666	-3.0193672	0.00253303	0.0111327
ENSDARG00000090081	5.0857625	-5.8662327	2.16375762	-2.7111321	0.00670539	0.02487371
ENSDARG00000099957	5.05236479	-5.8574621	2.17224381	-2.6965031	0.00700718	0.02578388

ENSDARG00000053702	13.4833663	-5.824477	1.73050393	-3.3657693	0.00076331	0.00397669
ENSDARG00000010332	4.83768839	-5.7884414	2.26149867	-2.5595599	0.01048048	0.03597136
ENSDARG00000105564	4.71124262	-5.758568	2.24424232	-2.5659297	0.01028997	0.03550159
ENSDARG00000077687	4.537109	-5.7025221	2.24113735	-2.5444768	0.01094416	0.03725633
ENSDARG00000102785	4.50371129	-5.6926616	2.25546153	-2.5239453	0.0116046	0.03911698
ENSDARG00000063059	4.39637309	-5.6541837	2.25472182	-2.5077079	0.01215171	0.04054724
ENSDARG00000029039	4.32957767	-5.6338202	2.26475095	-2.4876114	0.01286042	0.04237088
ENSDARG00000014858	4.32957767	-5.6338202	2.26475095	-2.4876114	0.01286042	0.04237088
ENSDARG00000079373	4.32957767	-5.6338202	2.26475095	-2.4876114	0.01286042	0.04237088
ENSDARG00000057262	9.84617406	-5.3178192	1.80343979	-2.9487091	0.00319104	0.01346111
ENSDARG00000037286	9.81992142	-5.3167447	1.81411845	-2.9307594	0.00338135	0.01409298
ENSDARG00000068580	9.42811538	-5.3021788	1.84826289	-2.8687363	0.00412115	0.01660661

Table 8.14: Top 20 upregulated genes in the Tdrd7-lacking PGCs vs control

	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSDARG00000079376	4440.87867	15.5188676	1.45886997	10.6375948	1.99E-26	2.58E-24
ENSDARG00000094559	2734.61398	14.8193734	1.45739796	10.168378	2.74E-24	2.82E-22
ENSDARG00000100833	2385.49568	14.6223231	1.45750255	10.0324512	1.10E-23	1.03E-21
ENSDARG00000095254	2040.98961	14.3972711	1.45808606	9.87408867	5.39E-23	4.75E-21
ENSDARG00000089806	1763.23177	14.1861639	1.45922096	9.72173799	2.44E-22	2.04E-20
ENSDARG00000039173	1303.49972	13.7504315	1.45859021	9.42720676	4.21E-21	3.11E-19
ENSDARG00000095744	1193.51678	13.6233185	1.45872989	9.33916458	9.71E-21	6.88E-19
ENSDARG00000019122	605.846137	12.6450638	1.46225487	8.64764691	5.26E-18	2.83E-16
ENSDARG00000006008	573.384117	12.5656732	2.61736874	4.80088	1.58E-06	1.70E-05
ENSDARG00000089750	986.807142	11.9077242	1.43747079	8.28380256	1.19E-16	5.51E-15
ENSDARG00000023656	944.251263	11.844875	1.44143162	8.21743799	2.08E-16	9.26E-15
ENSDARG00000075527	261.503992	11.4328608	1.47155768	7.76922371	7.90E-15	2.96E-13
ENSDARG00000074306	1288.80144	11.2937285	1.04242719	10.8340694	2.37E-27	3.35E-25
ENSDARG00000056151	555.538272	11.0762258	1.50123287	7.37808643	1.61E-13	5.22E-12
ENSDARG00000087390	173.703159	10.8431757	1.75342685	6.18399094	6.25E-10	1.28E-08
ENSDARG00000093628	413.919072	10.6537985	1.44366312	7.37969845	1.59E-13	5.19E-12
ENSDARG00000033161	141.797279	10.5501661	2.46817936	4.27447301	1.92E-05	0.00015748
ENSDARG00000056938	122.714718	10.3412146	2.22203438	4.65393995	3.26E-06	3.26E-05
ENSDARG00000055926	119.531759	10.3055776	1.50224452	6.86011991	6.88E-12	1.85E-10

9 Literature

Abbassi, L., Malki, S., Cockburn, K., Macaulay, A., Robert, C., Rossant, J., and Clarke, H.J. (2016). Multiple Mechanisms Cooperate to Constitutively Exclude the Transcriptional Co-Activator YAP from the Nucleus During Murine Oogenesis. Biol. Reprod. *94*, 102.

Adams, R.L., Terry, L.J., and Wente, S.R. (2014). Nucleoporin FG domains facilitate mRNP remodeling at the cytoplasmic face of the nuclear pore complex. Genetics *197*, 1213–1224.

Adey, A., Morrison, H.G., Asan, null, Xun, X., Kitzman, J.O., Turner, E.H., Stackhouse, B., MacKenzie, A.P., Caruccio, N.C., Zhang, X., et al. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol. *11*, R119.

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

Amikura, R., Hanyu, K., Kashikawa, M., and Kobayashi, S. (2001). Tudor protein is essential for the localization of mitochondrial RNAs in polar granules of Drosophila embryos. Mech. Dev. *107*, 97–104.

Amodeo, A.A., Jukam, D., Straight, A.F., and Skotheim, J.M. (2015). Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. Proc. Natl. Acad. Sci. U. S. A. *112*, E1086-1095.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol. Cell *31*, 785–799.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Basta, J., and Rauchman, M. (2015). The nucleosome remodeling and deacetylase complex in development and disease. Transl. Res. J. Lab. Clin. Med. *165*, 36–47.

Batchelder, C., Dunn, M.A., Choy, B., Suh, Y., Cassie, C., Shim, E.Y., Shin, T.H., Mello, C., Seydoux, G., and Blackwell, T.K. (1999). Transcriptional repression by the Caenorhabditis elegans germ-line protein PIE-1. Genes Dev. *13*, 202–212.

Baxendale, S., Davison, C., Muxworthy, C., Wolff, C., Ingham, P.W., and Roy, S. (2004). The B-cell maturation factor Blimp-1 specifies vertebrate slow-twitch muscle fiber identity in response to Hedgehog signaling. Nat. Genet. *36*, 88–93.

Beer, R.L., and Draper, B.W. (2013). nanos3 maintains germline stem cells and expression of the conserved germline stem cell gene nanos2 in the zebrafish ovary. Dev. Biol. *374*, 308–318.

Bender, L.B., Cao, R., Zhang, Y., and Strome, S. (2004). The MES-2/MES-3/MES-6 complex and regulation of histone H3 methylation in C. elegans. Curr. Biol. CB *14*, 1639–1643.

Bender, L.B., Suh, J., Carroll, C.R., Fong, Y., Fingerman, I.M., Briggs, S.D., Cao, R., Zhang, Y., Reinke, V., and Strome, S. (2006). MES-4: an autosome-associated histone methyltransferase that participates in silencing the X chromosomes in the C. elegans germ line. Dev. Camb. Engl. *133*, 3907–3917.

Bentson, L.F., Agbor, V.A., Agbor, L.N., Lopez, A.C., Nfonsam, L.E., Bornstein, S.S., Handel, M.A., and Linder, C.C. (2013). New point mutation in Golga3 causes multiple defects in spermatogenesis. Andrology *1*, 440–450.

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

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396–398.

Blackler, A.W. (1962). Transfer of primordial germ-cells between two subspecies of Xenopus laevis. J. Embryol. Exp. Morphol. *10*, 641–651.

Blankenship, J.T., and Wieschaus, E. (2001). Two new roles for the Drosophila AP patterning system in early morphogenesis. Dev. Camb. Engl. *128*, 5129–5138.

Blaser, H., Eisenbeiss, S., Neumann, M., Reichman-Fried, M., Thisse, B., Thisse, C., and Raz, E. (2005). Transition from non-motile behaviour to directed migration during early PGC development in zebrafish. J. Cell Sci. *118*, 4027–4038.

Blaser, H., Reichman-Fried, M., Castanon, I., Dumstrei, K., Marlow, F.L., Kawakami, K., Solnica-Krezel, L., Heisenberg, C.-P., and Raz, E. (2006). Migration of zebrafish primordial germ cells: a role for myosin contraction and cytoplasmic flow. Dev. Cell *11*, 613–627.

Blythe, S.A., and Wieschaus, E.F. (2016). Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis. ELife *5*.

Bogdanovic, O., Fernandez-Miñán, A., Tena, J.J., de la Calle-Mustienes, E., Hidalgo, C., van Kruysbergen, I., van Heeringen, S.J., Veenstra, G.J.C., and Gómez-Skarmeta, J.L. (2012). Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. *22*, 2043–2053.

Bogdanović, O., Smits, A.H., de la Calle Mustienes, E., Tena, J.J., Ford, E., Williams, R., Senanayake, U., Schultz, M.D., Hontelez, S., van Kruijsbergen, I., et al. (2016). Active DNA demethylation at enhancers during the vertebrate phylotypic period. Nat. Genet. *48*, 417–426.

Bontems, F., Stein, A., Marlow, F., Lyautey, J., Gupta, T., Mullins, M.C., and Dosch, R. (2009). Bucky ball organizes germ plasm assembly in zebrafish. Curr. Biol. CB *19*, 414–422.

Boswell, R.E., and Mahowald, A.P. (1985). tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. Cell *43*, 97–104.

Botuyan, M.V., Lee, J., Ward, I.M., Kim, J.-E., Thompson, J.R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell *127*, 1361–1373.

Braat, A.K., Zandbergen, T., van de Water, S., Goos, H.J., and Zivkovic, D. (1999). Characterization of zebrafish primordial germ cells: morphology and early distribution of vasa RNA. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *216*, 153–167.

Braat, A.K., van de Water, S., Korving, J., and Zivkovic, D. (2001). A zebrafish vasa morphant abolishes vasa protein but does not affect the establishment of the germline. Genes. N. Y. N 2000 *30*, 183–185.

Brahms, H., Meheus, L., de Brabandere, V., Fischer, U., and Lührmann, R. (2001). Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. RNA N. Y. N *7*, 1531–1542.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729–1732.

Breiling, A., and Lyko, F. (2015). Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. Epigenetics Chromatin *8*, 24.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell *128*, 1089–1103.

Buehr, M.L., and Blackler, A.W. (1970). Sterility and partial sterility in the South African clawed toad following the pricking of the egg. J. Embryol. Exp. Morphol. *23*, 375–384.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes Dev. *10*, 711–724.

Burke, T.W., and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved from Drosophila to humans and is recognized by TAFII60 of Drosophila. Genes Dev. *11*, 3020–3031.

Bushati, N., Stark, A., Brennecke, J., and Cohen, S.M. (2008). Temporal reciprocity of miRNAs and their targets during the maternal-to-zygotic transition in Drosophila. Curr. Biol. CB *18*, 501–506.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Camacho, A., Rodriguez-Cuenca, S., Blount, M., Prieur, X., Barbarroja, N., Fuller, M., Hardingham, G.E., and Vidal-Puig, A. (2012). Ablation of PGC1 beta prevents mTOR dependent endoplasmic reticulum stress response. Exp. Neurol. *237*, 396–406.

Charlesworth, A., Wilczynska, A., Thampi, P., Cox, L.L., and MacNicol, A.M. (2006). Musashi regulates the temporal order of mRNA translation during Xenopus oocyte maturation. EMBO J. *25*, 2792–2801.

Chen, H.-H., Welling, M., Bloch, D.B., Muñoz, J., Mientjes, E., Chen, X., Tramp, C., Wu, J., Yabuuchi, A., Chou, Y.-F., et al. (2014). DAZL limits pluripotency, differentiation, and apoptosis in developing primordial germ cells. Stem Cell Rep. *3*, 892–904.

Cheng, X. (2014). Structural and functional coordination of DNA and histone methylation. Cold Spring Harb. Perspect. Biol. *6*.

Cheng, D., Côté, J., Shaaban, S., and Bedford, M.T. (2007). The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. Mol. Cell *25*, 71–83.

Collart, C., Allen, G.E., Bradshaw, C.R., Smith, J.C., and Zegerman, P. (2013). Titration of four replication factors is essential for the Xenopus laevis midblastula transition. Science *341*, 893–896.

Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev. *12*, 3715–3727.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., et al. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res. *16*, 123–131.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Dahl, J.A., Jung, I., Aanes, H., Greggains, G.D., Manaf, A., Lerdrup, M., Li, G., Kuan, S., Li, B., Lee, A.Y., et al. (2016). Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. Nature *537*, 548–552.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat. Genet. *49*, 941–945.

Dekens, M.P.S., Pelegri, F.J., Maischein, H.-M., and Nüsslein-Volhard, C. (2003). The maternal-effect gene futile cycle is essential for pronuclear congression and mitotic spindle assembly in the zebrafish zygote. Dev. Camb. Engl. *130*, 3907–3916.

Díaz, N., Kruse, K., Erdmann, T., Staiger, A.M., Ott, G., Lenz, G., and Vaquerizas, J.M. (2018). Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. Nat. Commun. *9*, 4938.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Doitsidou, M., Reichman-Fried, M., Stebler, J., Köprunner, M., Dörries, J., Meyer, D., Esguerra, C.V., Leung, T., and Raz, E. (2002). Guidance of primordial germ cell migration by the chemokine SDF-1. Cell *111*, 647–659.

Dosch, R., Wagner, D.S., Mintzer, K.A., Runke, G., Wiemelt, A.P., and Mullins, M.C. (2004). Maternal control of vertebrate development before the midblastula transition: mutants from the zebrafish I. Dev. Cell *6*, 771–780.

Doyle, H.J., Kraut, R., and Levine, M. (1989). Spatial regulation of zerknüllt: a dorsal-ventral patterning gene in Drosophila. Genes Dev. *3*, 1518–1533.

Draper, B.W., McCallum, C.M., and Moens, C.B. (2007). nanos1 is required to maintain oocyte production in adult zebrafish. Dev. Biol. *305*, 589–598.

Drozd, A.M., Walczak, M.P., Piaskowski, S., Stoczynska-Fidelus, E., Rieske, P., and Grzela, D.P. (2015). Generation of human iPSCs from cells of fibroblastic and epithelial origin by means of the oriP/EBNA-1 episomal reprogramming system. Stem Cell Res. Ther. *6*, 122.

Durdevic, Z., Pillai, R.S., and Ephrussi, A. (2018). Transposon silencing in the Drosophila female germline is essential for genome stability in progeny embryos. Life Sci. Alliance *1*, e201800179.

Eberharter, A., and Becker, P.B. (2002). Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. EMBO Rep. *3*, 224–229.

Eddy, E.M. (1975). Germ plasm and the differentiation of the germ cell line. Int. Rev. Cytol. *43*, 229–280.

Edgar, B.A., Kiehle, C.P., and Schubiger, G. (1986). Cell cycle control by the nucleo-cytoplasmic ratio in early Drosophila development. Cell *44*, 365–372.

Egerman, M.A., and Glass, D.J. (2014). Signaling pathways controlling skeletal muscle mass. Crit. Rev. Biochem. Mol. Biol. *49*, 59–68.

Eno, C., and Pelegri, F. (2013). Gradual recruitment and selective clearing generate germ plasm aggregates in the zebrafish embryo. BioArchitecture *3*, 125–132.

Eno, C., and Pelegri, F. (2016). Germ Cell Determinant Transmission, Segregation, and Function in the Zebrafish Embryo. In Insights from Animal Reproduction, R. Payan Carreira, ed. (InTech), p.

Eno, C., Gomez, T., Slusarski, D.C., and Pelegri, F. (2018). Slow calcium waves mediate furrow microtubule reorganization and germ plasm compaction in the early zebrafish embryo. Dev. Camb. Engl. *145*.

Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by oskar. Nature *358*, 387–392.

Evans, T., Wade, C.M., Chapman, F.A., Johnson, A.D., and Loose, M. (2014). Acquisition of Germ Plasm Accelerates Vertebrate Evolution. Science *344*, 200–203.

Extavour, C.G.M. (2007). Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. Integr. Comp. Biol. *47*, 770–785.

Filippov, V., Solovyev, V., Filippova, M., and Gill, S.S. (2000). A novel type of RNase III family proteins in eukaryotes. Gene *245*, 213–221.

Fischer, U., Liu, Q., and Dreyfuss, G. (1997). The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. Cell *90*, 1023–1029.

Flyamer, I.M., Gassler, J., Imakaev, M., Brandão, H.B., Ulianov, S.V., Abdennur, N., Razin, S.V., Mirny, L.A., and Tachibana-Konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature *544*, 110–114.

Forbes, A., and Lehmann, R. (1998). Nanos and Pumilio have critical roles in the development and function of Drosophila germline stem cells. Dev. Camb. Engl. *125*, 679–690.

Foygel, K., Choi, B., Jun, S., Leong, D.E., Lee, A., Wong, C.C., Zuo, E., Eckart, M., Reijo Pera, R.A., Wong, W.H., et al. (2008). A novel and critical role for Oct4 as a regulator of the maternal-embryonic transition. PloS One *3*, e4109.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574–1577.

Gao, X., Ge, L., Shao, J., Su, C., Zhao, H., Saarikettu, J., Yao, X., Yao, Z., Silvennoinen, O., and Yang, J. (2010). Tudor-SN interacts with and co-localizes with G3BP in stress granules under stress conditions. FEBS Lett. *584*, 3525–3532.

Gaskell, T.L., Esnal, A., Robinson, L.L.L., Anderson, R.A., and Saunders, P.T.K. (2004). Immunohistochemical profiling of germ cells within the human fetal testis: identification of three subpopulations. Biol. Reprod. *71*, 2012–2021.

Gazdag, E., Santenard, A., Ziegler-Birling, C., Altobelli, G., Poch, O., Tora, L., and Torres-Padilla, M.-E. (2009). TBP2 is essential for germ cell development by regulating transcription and chromatin condensation in the oocyte. Genes Dev. *23*, 2210–2223.

Gdula, D.A., Gerasimova, T.I., and Corces, V.G. (1996). Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. Proc. Natl. Acad. Sci. U. S. A. *93*, 9378–9383.

Gentsch, G.E., Spruce, T., Monteiro, R.S., Owens, N.D.L., Martin, S.R., and Smith, J.C. (2018). Innate Immune Response and Off-Target Mis-splicing Are Common Morpholino-Induced Side Effects in Xenopus. Dev. Cell *44*, 597-610.e10.

Gerovska, D., and Araúzo-Bravo, M.J. (2016). Does mouse embryo primordial germ cell activation start before implantation as suggested by single-cell transcriptomics dynamics? Mol. Hum. Reprod. *22*, 208–225.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79.

Gkountela, S., Zhang, K.X., Shafiq, T.A., Liao, W.-W., Hargan-Calvopiña, J., Chen, P.-Y., and Clark, A.T. (2015). DNA Demethylation Dynamics in the Human Prenatal Germline. Cell *161*, 1425–1436.

Godin, I., Deed, R., Cooke, J., Zsebo, K., Dexter, M., and Wylie, C.C. (1991). Effects of the steel gene product on mouse primordial germ cells in culture. Nature *352*, 807–809.

Golumbeski, G.S., Bardsley, A., Tax, F., and Boswell, R.E. (1991). tudor, a posterior-group gene of Drosophila melanogaster, encodes a novel protein and an mRNA localized during mid-oogenesis. Genes Dev. *5*, 2060–2070.

Gorokhova, S., Bibert, S., Geering, K., and Heintz, N. (2007). A novel family of transmembrane proteins interacting with beta subunits of the Na,K-ATPase. Hum. Mol. Genet. *16*, 2394–2410.

Gowher, H., and Jeltsch, A. (2018). Mammalian DNA methyltransferases: new discoveries and open questions. Biochem. Soc. Trans. *46*, 1191–1202.

Gregory, R.I., Chendrimada, T.P., Cooch, N., and Shiekhattar, R. (2005). Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell *123*, 631–640.

Gross-Thebing, T., Yigit, S., Pfeiffer, J., Reichman-Fried, M., Bandemer, J., Ruckert, C., Rathmer, C., Goudarzi, M., Stehling, M., Tarbashevich, K., et al. (2017). The Vertebrate Protein Dead End Maintains Primordial Germ Cell Fate by Inhibiting Somatic Differentiation. Dev. Cell *43*, 704–715.e5.

Gubitz, A.K., Feng, W., and Dreyfuss, G. (2004). The SMN complex. Exp. Cell Res. *296*, 51–56.

Guo, F., Yan, L., Guo, H., Li, L., Hu, B., Zhao, Y., Yong, J., Hu, Y., Wang, X., Wei, Y., et al. (2015). The Transcriptome and DNA Methylome Landscapes of Human Primordial Germ Cells. Cell *161*, 1437–1452.

Guo, H., Hu, B., Yan, L., Yong, J., Wu, Y., Gao, Y., Guo, F., Hou, Y., Fan, X., Dong, J., et al. (2017). DNA methylation and chromatin accessibility profiling of mouse and human fetal germ cells. Cell Res. *27*, 165–183.

Guven-Ozkan, T., Nishi, Y., Robertson, S.M., and Lin, R. (2008). Global transcriptional repression in C. elegans germline precursors by regulated sequestration of TAF-4. Cell *135*, 149–160.

Haberle, V., Li, N., Hadzhiev, Y., Plessy, C., Previti, C., Nepal, C., Gehrig, J., Dong, X., Akalin, A., Suzuki, A.M., et al. (2014). Two independent transcription initiation codes overlap on vertebrate core promoters. Nature *507*, 381–385.

Hadzhiev, Y., Qureshi, H.K., Wheatley, L., Cooper, L., Jasiulewicz, A., Van Nguyen, H., Wragg, J.W., Poovathumkadavil, D., Conic, S., Bajan, S., et al. (2019). A cell cycle-coordinated Polymerase II transcription compartment encompasses gene expression before global genome activation. Nat. Commun. *10*, 691.

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U.C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M.A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature *452*, 877–881.

Hammond, S.M., Crable, S.C., and Anderson, K.P. (2005). Negative regulatory elements are present in the human LMO2 oncogene and may contribute to its expression in leukemia. Leuk. Res. *29*, 89–97.

Han, R., Wang, R., Zhao, Q., Han, Y., Zong, S., Miao, S., Song, W., and Wang, L. (2016). Trim69 regulates zebrafish brain development by ap-1 pathway. Sci. Rep. *6*, 24034.

Hanyu-Nakamura, K., Sonobe-Nojima, H., Tanigawa, A., Lasko, P., and Nakamura, A. (2008). Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells. Nature *451*, 730–733.

Harrison, M.M., Li, X.-Y., Kaplan, T., Botchan, M.R., and Eisen, M.B. (2011). Zelda Binding in the Early Drosophila melanogaster Embryo Marks Regions Subsequently Activated at the Maternal-to-Zygotic Transition. PLoS Genet. *7*, e1002266.

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., and Cedar, H. (2003). The role of DNA methylation in setting up chromatin structure during development. Nat. Genet. *34*, 187–192.

Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S., and Saitou, M. (2011). Reconstitution of the Mouse Germ Cell Specification Pathway in Culture by Pluripotent Stem Cells. Cell *146*, 519–532.

Hayashi, T., Ozaki, H., Sasagawa, Y., Umeda, M., Danno, H., and Nikaido, I. (2018). Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat. Commun. *9*, 619.

He, H.H., Meyer, C.A., Chen, M.W., Jordan, V.C., Brown, M., and Liu, X.S. (2012). Differential DNase I hypersensitivity reveals factor-dependent chromatin dynamics. Genome Res. *22*, 1015–1025.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heyn, P., Kircher, M., Dahl, A., Kelso, J., Tomancak, P., Kalinka, A.T., and Neugebauer, K.M. (2014). The earliest transcribed zygotic genes are short, newly evolved, and different across species. Cell Rep. *6*, 285–292.

Hill, P.W.S., Leitch, H.G., Requena, C.E., Sun, Z., Amouroux, R., Roman-Trufero, M., Borkowska, M., Terragni, J., Vaisvila, R., Linnett, S., et al. (2018). Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. Nature *555*, 392–396.

Hopf, C., Viebahn, C., and Püschel, B. (2011). BMP signals and the transcriptional repressor BLIMP1 during germline segregation in the mammalian embryo. Dev. Genes Evol. *221*, 209–223.

Houston, D.W., and King, M.L. (2000). Germ plasm and molecular determinants of germ cell fate. Curr. Top. Dev. Biol. *50*, 155–181.

Houwing, S., Kamminga, L.M., Berezikov, E., Cronembold, D., Girard, A., van den Elst, H., Filippov, D.V., Blaser, H., Raz, E., Moens, C.B., et al. (2007). A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. Cell *129*, 69–82.

Huang, H.-Y., Houwing, S., Kaaij, L.J.T., Meppelink, A., Redl, S., Gauci, S., Vos, H., Draper, B.W., Moens, C.B., Burgering, B.M., et al. (2011). Tdrd1 acts as a molecular scaffold for Piwi proteins and piRNA targets in zebrafish. EMBO J. *30*, 3298–3308.

Huang, S., Shao, G., and Liu, L. (1998). The PR domain of the Rb-binding zinc finger protein RIZ1 is a protein binding interface and is related to the SET domain functioning in chromatin-mediated gene expression. J. Biol. Chem. *273*, 15933–15939.

Huang, X.A., Yin, H., Sweeney, S., Raha, D., Snyder, M., and Lin, H. (2013). A major epigenetic programming mechanism guided by piRNAs. Dev. Cell *24*, 502–516.

Hug, C.B., Grimaldi, A.G., Kruse, K., and Vaquerizas, J.M. (2017). Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell *169*, 216-228.e19.

Iguchi, N., Tobias, J.W., and Hecht, N.B. (2006). Expression profiling reveals meiotic male germ cell mRNAs that are translationally up- and down-regulated. Proc. Natl. Acad. Sci. U. S. A. *103*, 7712–7717.

Illmensee, K., and Mahowald, A.P. (1974). Transplantation of posterior polar plasm in Drosophila. Induction of germ cells at the anterior pole of the egg. Proc. Natl. Acad. Sci. U. S. A. *71*, 1016–1020.

Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H., and Surani, M.A. (2015). SOX17 is a critical specifier of human primordial germ cell fate. Cell *160*, 253–268.

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. Genome Res. *21*, 1543–1551.

Johnson, A.D., Richardson, E., Bachvarova, R.F., and Crother, B.I. (2011a). Evolution of the germ line-soma relationship in vertebrate embryos. Reproduction *141*, 291–300.

Johnson, A.D., Richardson, E., Bachvarova, R.F., and Crother, B.I. (2011b). Evolution of the germ line-soma relationship in vertebrate embryos. Reproduction *141*, 291–300.

Joseph, S.R., Pálfy, M., Hilbert, L., Kumar, M., Karschau, J., Zaburdaev, V., Shevchenko, A., and Vastenhouw, N.L. (2017). Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. ELife *6*.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H., and Razin, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev. *6*, 705–714.

Kane, D.A., Hammerschmidt, M., Mullins, M.C., Maischein, H.M., Brand, M., van Eeden, F.J., Furutani-Seiki, M., Granato, M., Haffter, P., Heisenberg, C.P., et al. (1996). The zebrafish epiboly mutants. Dev. Camb. Engl. *123*, 47–55.

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010). Nucleosome sequence preferences influence in vivo nucleosome organization. Nat. Struct. Mol. Biol. *17*, 918–920.

Kawahara, H., Imai, T., Imataka, H., Tsujimoto, M., Matsumoto, K., and Okano, H. (2008). Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. J. Cell Biol. *181*, 639–653.

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L., et al. (2017). 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell *170*, 367-381.e20.

Kedde, M., Strasser, M.J., Boldajipour, B., Oude Vrielink, J.A.F., Slanchev, K., le Sage, C., Nagel, R., Voorhoeve, P.M., van Duijse, J., Ørom, U.A., et al. (2007). RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. Cell *131*, 1273–1286.

Kim-Ha, J., Kerr, K., and Macdonald, P.M. (1995). Translational regulation of oskar mRNA by bruno, an ovarian RNA-binding protein, is essential. Cell *81*, 403–412.

Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B., and Schilling, T.F. (1995). Stages of embryonic development of the zebrafish. Dev. Dyn. Off. Publ. Am. Assoc. Anat. *203*, 253–310.

Kirino, Y., Kim, N., de Planell-Saguer, M., Khandros, E., Chiorean, S., Klein, P.S., Rigoutsos, I., Jongens, T.A., and Mourelatos, Z. (2009). Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat. Cell Biol. *11*, 652–658.

Knaut, H., Pelegri, F., Bohmann, K., Schwarz, H., and Nüsslein-Volhard, C. (2000). Zebrafish vasa RNA but not its protein is a component of the germ plasm and segregates asymmetrically before germline specification. J. Cell Biol. *149*, 875–888.

Knaut, H., Werz, C., Geisler, R., Nüsslein-Volhard, C., and Tübingen 2000 Screen Consortium (2003). A zebrafish homologue of the chemokine receptor Cxcr4 is a germ-cell guidance receptor. Nature *421*, 279–282.

Kobayashi, S., Mizuno, H., and Okada, M. (1988). Accumulation and Spatial Distribution of Poly(A)+RNA in Oocytes and Early Embryos of Drosophila melanogaster. (Poly(A)+RNA/in situ hybridization/oocytes and embryos/Drosophila melanogaster). Dev. Growth Differ. *30*, 251–260.

Kobayashi, S., Yamada, M., Asaoka, M., and Kitamura, T. (1996). Essential role of the posterior morphogen nanos for germline development in Drosophila. Nature *380*, 708–711.

Kobayashi, T., Zhang, H., Tang, W.W.C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann, S., Contreras, D.A., and Webb, R. (2017a). Principles of early human development and germ cell program from conserved model systems. Nature *546*, 416–420.

Kobayashi, T., Zhang, H., Tang, W.W.C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann, S., Contreras, D.A., Webb, R., et al. (2017b). Principles of early human development and germ cell program from conserved model systems. Nature *546*, 416–420.

Köprunner, M., Thisse, C., Thisse, B., and Raz, E. (2001). A zebrafish nanos-related gene is essential for the development of primordial germ cells. Genes Dev. *15*, 2877–2885.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Krishnakumar, P., Riemer, S., Perera, R., Lingner, T., Goloborodko, A., Khalifa, H., Bontems, F., Kaufholz, F., El-Brolosy, M.A., and Dosch, R. (2018). Functional equivalence of germ plasm organizers. PLoS Genet. *14*, e1007696.

Krøvel, A.V., and Olsen, L.C. (2002). Expression of a vas::EGFP transgene in primordial germ cells of the zebrafish. Mech. Dev. *116*, 141–150.

Kubik, S., Bruzzone, M.J., and Shore, D. (2017). Establishing nucleosome architecture and stability at promoters: Roles of pioneer transcription factors and the RSC chromatin remodeler. BioEssays News Rev. Mol. Cell. Dev. Biol. *39*.

Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2019). High-resolution genetic mapping of putative causal interactions between regions of open chromatin. Nat. Genet. *51*, 128–137.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. *22*, 1617–1635.

Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., Le Berre-Anton, V., Bouzayen, M., and Maza, E. (2018). Optimization of an RNA-Seq Differential Gene Expression Analysis Depending on Biological Replicate Number and Library Size. Front. Plant Sci. *9*, 108.

Lamb, M.M., and Laird, C.D. (1976). Increase in nuclear poly(A)-containing RNA at syncytial blastoderm in Drosophila melanogaster embryos. Dev. Biol. *52*, 31–42.

Lamond, A.I., and Earnshaw, W.C. (1998). Structure and function in the nucleus. Science *280*, 547–553.

Lange, U.C., Adams, D.J., Lee, C., Barton, S., Schneider, R., Bradley, A., and Surani, M.A. (2008). Normal germ line establishment in mice carrying a deletion of the Ifitm/Fragilis gene family cluster. Mol. Cell. Biol. *28*, 4688–4696.

Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C.V., Korving, J.P., and Hogan, B.L. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. Genes Dev. *13*, 424–436.

Lecuit, T., Samanta, R., and Wieschaus, E. (2002). slam encodes a developmental regulator of polarized membrane growth during cleavage of the Drosophila embryo. Dev. Cell *2*, 425–436.

Lee, B.C., and Roy, S. (2006). Blimp-1 is an essential component of the genetic program controlling development of the pectoral limb bud. Dev. Biol. *300*, 623–634.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat. Rev. Mol. Cell Biol. *8*, 284–295.

Lee, M.T., Bonneau, A.R., Takacs, C.M., Bazzini, A.A., DiVito, K.R., Fleming, E.S., and Giraldez, A.J. (2013). Nanog, Pou5f1 and SoxB1 activate zygotic gene expression during the maternal-to-zygotic transition. Nature *503*, 360–364.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.

Lehmann, R. (2016). Germ Plasm Biogenesis--An Oskar-Centric Perspective. Curr. Top. Dev. Biol. *116*, 679–707.

Leipe, D.D., and Landsman, D. (1997). Histone deacetylases, acetoin utilization proteins and acetylpolyamine amidohydrolases are members of an ancient protein superfamily. Nucleic Acids Res. *25*, 3693–3697.

Lesch, B.J., and Page, D.C. (2014). Poised chromatin in the mammalian germ line. Development *141*, 3619–3626.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. CB *20*, R754-763.

Levine, E., Cupp, A.S., Miyashiro, L., and Skinner, M.K. (2000). Role of transforming growth factor-alpha and the epidermal growth factor receptor in embryonic rat testis development. Biol. Reprod. *62*, 477–490.

Li, S., and Patel, D.J. (2016). Drosha and Dicer: Slicers cut from the same cloth. Cell Res. *26*, 511–512.

Li, C., Vagin, V.V., Lee, S., Xu, J., Ma, S., Xi, H., Seitz, H., Horwich, M.D., Syrzycka, M., Honda, B.M., et al. (2009). Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell *137*, 509–521.

Li, M., Zhu, F., Li, Z., Hong, N., and Hong, Y. (2016). Dazl is a critical player for primordial germ cell formation in medaka. Sci. Rep. *6*, 28317.

Liang, G., Lin, J.C.Y., Wei, V., Yoo, C., Cheng, J.C., Nguyen, C.T., Weisenberger, D.J., Egger, G., Takai, D., Gonzales, F.A., et al. (2004). Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. Proc. Natl. Acad. Sci. U. S. A. *101*, 7357–7362.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lifton, R.P., Goldberg, M.L., Karp, R.W., and Hogness, D.S. (1978). The organization of the histone genes in Drosophila melanogaster: functional and evolutionary implications. Cold Spring Harb. Symp. Quant. Biol. *42 Pt 2*, 1047–1051.

Lindeman, L.C., Winata, C.L., Aanes, H., Mathavan, S., Alestrom, P., and Collas, P. (2010). Chromatin states of developmentally-regulated genes revealed by DNA and histone methylation patterns in zebrafish embryos. Int. J. Dev. Biol. *54*, 803–813.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558–562.

Lolas, M., Valenzuela, P.D.T., Tjian, R., and Liu, Z. (2014). Charting Brachyury-mediated developmental pathways during early mouse embryogenesis. Proc. Natl. Acad. Sci. U. S. A. *111*, 4478–4483.

Looman, C., Abrink, M., Mark, C., and Hellman, L. (2002). KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. Mol. Biol. Evol. *19*, 2118–2130.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251–260.

Lund, E., Liu, M., Hartley, R.S., Sheets, M.D., and Dahlberg, J.E. (2009). Deadenylation of maternal mRNAs mediated by miR-427 in Xenopus laevis embryos. RNA N. Y. N *15*, 2351–2363.

Luo, M., Ling, T., Xie, W., Sun, H., Zhou, Y., Zhu, Q., Shen, M., Zong, L., Lyu, G., Zhao, Y., et al. (2013). NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. Stem Cells Dayt. Ohio *31*, 1278–1286.

Macrae, I.J., Zhou, K., Li, F., Repic, A., Brooks, A.N., Cande, W.Z., Adams, P.D., and Doudna, J.A. (2006). Structural basis for double-stranded RNA processing by Dicer. Science *311*, 195–198.

Maegawa, S., Yasuda, K., and Inoue, K. (1999). Maternal mRNA localization of zebrafish DAZ-like gene. Mech. Dev. *81*, 223–226.

Magnúsdóttir, E., Dietmann, S., Murakami, K., Günesdogan, U., Tang, F., Bao, S., Diamanti, E., Lao, K., Gottgens, B., and Azim Surani, M. (2013). A tripartite transcription factor network regulates primordial germ cell specification in mice. Nat. Cell Biol. *15*, 905–915.

Manoli, M., and Driever, W. (2012). Fluorescence-activated cell sorting (FACS) of fluorescently tagged cells from zebrafish larvae for RNA isolation. Cold Spring Harb. Protoc. *2012*.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nat. Rev. Genet. *11*, 285–296.

Margueron, R., Trojer, P., and Reinberg, D. (2005). The key to development: interpreting the histone code? Curr. Opin. Genet. Dev. *15*, 163–176.

Marlow, F.L., and Mullins, M.C. (2008). Bucky ball functions in Balbiani body assembly and animal-vegetal polarity in the oocyte and follicle cell layer in zebrafish. Dev. Biol. *321*, 40–50.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.

Meier, M., Grant, J., Dowdle, A., Thomas, A., Gerton, J., Collas, P., O'Sullivan, J.M., and Horsfield, J.A. (2018). Cohesin facilitates zygotic genome activation in zebrafish. Dev. Camb. Engl. *145*.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Melamed, P., Yosefzon, Y., Rudnizky, S., and Pnueli, L. (2016). Transcriptional enhancers: Transcription, function and flexibility. Transcription *7*, 26–31.

Mello, C.C., Schubert, C., Draper, B., Zhang, W., Lobel, R., and Priess, J.R. (1996). The PIE-1 protein and germline specification in C. elegans embryos. Nature *382*, 710–712.

Merrill, P.T., Sweeton, D., and Wieschaus, E. (1988). Requirements for autosomal gene activity during precellular stages of Drosophila melanogaster. Dev. Camb. Engl. *104*, 495–509.

Meyen, D., Tarbashevich, K., Banisch, T.U., Wittwer, C., Reichman-Fried, M., Maugis, B., Grimaldi, C., Messerschmidt, E.-M., and Raz, E. (2015). Dynamic filopodia are required for chemokine-dependent intracellular polarization during guided cell migration in vivo. ELife *4*.

von Meyenn, F., and Reik, W. (2015). Forget the Parents: Epigenetic Reprogramming in Human Germ Cells. Cell *161*, 1248–1251.

Mishima, Y., Giraldez, A.J., Takeda, Y., Fujiwara, T., Sakamoto, H., Schier, A.F., and Inoue, K. (2006). Differential regulation of germline mRNAs in soma and germ cells by zebrafish miR-430. Curr. Biol. CB *16*, 2135–2142.

Molyneaux, K.A., Stallock, J., Schaible, K., and Wylie, C. (2001). Time-lapse analysis of living mouse germ cell migration. Dev. Biol. *240*, 488–498.

Molyneaux, K.A., Zinszner, H., Kunwar, P.S., Schaible, K., Stebler, J., Sunshine, M.J., O'Brien, W., Raz, E., Littman, D., Wylie, C., et al. (2003). The chemokine SDF1/CXCL12 and its receptor CXCR4 regulate mouse germ cell migration and survival. Dev. Camb. Engl. *130*, 4279–4286.

Murphy, P.J., Wu, S.F., James, C.R., Wike, C.L., and Cairns, B.R. (2018). Placeholder Nucleosomes Underlie Germline-to-Embryo DNA Methylation Reprogramming. Cell *172*, 993-1006.e13.

Nair, S., Marlow, F., Abrams, E., Kapp, L., Mullins, M.C., and Pelegri, F. (2013). The chromosomal passenger protein birc5b organizes microfilaments and germ plasm in the zebrafish embryo. PLoS Genet. *9*, e1003448.

Nakamura, A., Sato, K., and Hanyu-Nakamura, K. (2004). Drosophila cup is an eIF4E binding protein that associates with Bruno and regulates oskar mRNA translation in oogenesis. Dev. Cell *6*, 69–78.

Newport, J., and Kirschner, M. (1982). A major developmental transition in early Xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell *30*, 675–686.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell *11*, 709–719.

Ogbourne, S., and Antalis, T.M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. Biochem. J. *331 (Pt 1)*, 1–14.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S.C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., et al. (2005). Blimp1 is a critical determinant of the germ cell lineage in mice. Nature *436*, 207–213.

Ohinata, Y., Ohta, H., Shigeta, M., Yamanaka, K., Wakayama, T., and Saitou, M. (2009). A Signaling Principle for the Specification of the Germ Cell Lineage in Mice. Cell *137*, 571–584.

Patel, S.S., Belmont, B.J., Sante, J.M., and Rexach, M.F. (2007). Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. Cell *129*, 83–96.

Payer, B., Saitou, M., Barton, S.C., Thresher, R., Dixon, J.P.C., Zahn, D., Colledge, W.H., Carlton, M.B.L., Nakano, T., and Surani, M.A. (2003). Stella is a maternal effect gene required for normal early development in mice. Curr. Biol. CB *13*, 2110–2117.

Pek, J.W., Anand, A., and Kai, T. (2012). Tudor domain proteins in development. Dev. Camb. Engl. *139*, 2255–2266.

Pellizzoni, L., Kataoka, N., Charroux, B., and Dreyfuss, G. (1998). A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. Cell *95*, 615–624.

Percharde, M., Wong, P., and Ramalho-Santos, M. (2017). Global Hypertranscription in the Mouse Embryonic Germline. Cell Rep. *19*, 1987–1996.

Pitt, J.N., Schisa, J.A., and Priess, J.R. (2000). P granules in the germ cells of Caenorhabditis elegans adults are associated with clusters of nuclear pores and contain RNA. Dev. Biol. *219*, 315–333.

Quillien, A., Abdalla, M., Yu, J., Ou, J., Zhu, L.J., and Lawson, N.D. (2017). Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq. Cell Rep. *20*, 709–720.

Raz, E. (2003). Primordial germ-cell development: the zebrafish perspective. Nat. Rev. Genet. *4*, 690–700.

Razin, A., and Cedar, H. (1991). DNA methylation and gene expression. Microbiol. Rev. *55*, 451–458.

Rechtsteiner, A., Ercan, S., Takasaki, T., Phippen, T.M., Egelhofer, T.A., Wang, W., Kimura, H., Lieb, J.D., and Strome, S. (2010). The Histone H3K36 Methyltransferase MES-4 Acts Epigenetically to Transmit the Memory of Germline Gene Expression to Progeny. PLoS Genet. *6*, e1001091.

Ribbeck, K., and Görlich, D. (2002). The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion. EMBO J. *21*, 2664–2671.

Richardson, B.E., and Lehmann, R. (2010). Mechanisms guiding primordial germ cell migration: strategies from different organisms. Nat. Rev. Mol. Cell Biol. *11*, 37–49.

Riemer, S., Bontems, F., Krishnakumar, P., Gömann, J., and Dosch, R. (2015). A functional Bucky ball-GFP transgene visualizes germ plasm in living zebrafish. Gene Expr. Patterns GEP *18*, 44–52.

Rippe, K., von Hippel, P.H., and Langowski, J. (1995). Action at a distance: DNA-looping and initiation of transcription. Trends Biochem. Sci. *20*, 500–506.

Roovers, E.F., Kaaij, L.J.T., Redl, S., Bronkhorst, A.W., Wiebrands, K., de Jesus Domingues, A.M., Huang, H.-Y., Han, C.-T., Riemer, S., Dosch, R., et al. (2018). Tdrd6a Regulates the Aggregation of Buc into Functional Subcellular Compartments that Drive Germ Cell Specification. Dev. Cell *46*, 285-301.e9.

Rose, N.R., and Klose, R.J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. Biochim. Biophys. Acta *1839*, 1362–1372.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.

Rothschild, S.C., Lahvic, J., Francescatto, L., McLeod, J.J.A., Burgess, S.M., and Tombes, R.M. (2013). CaMK-II activation is essential for zebrafish inner ear development and acts through Delta-Notch signaling. Dev. Biol. *381*, 179–188.

Sabour, D., Araúzo-Bravo, M.J., Hübner, K., Ko, K., Greber, B., Gentile, L., Stehling, M., and Schöler, H.R. (2011). Identification of genes specific to mouse primordial germ cells through dynamic global gene expression. Hum. Mol. Genet. *20*, 115–125.

Saitou, M., Barton, S.C., and Surani, M.A. (2002). A molecular programme for the specification of germ cell fate in mice. Nature *418*, 293–300.

Scadden, A.D.J. (2005). The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. Nat. Struct. Mol. Biol. *12*, 489–496.

Schejter, E.D., and Wieschaus, E. (1993). bottleneck acts as a regulator of the microfilament network governing cellularization of the Drosophila embryo. Cell *75*, 373–385.

Schug, J., Schuller, W.-P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol. *6*, R33.

Seifert, J.R.K., and Lehmann, R. (2012). Drosophila primordial germ cell migration requires epithelial remodeling of the endoderm. Dev. Camb. Engl. *139*, 2101–2106.

Semotok, J.L., Cooperstock, R.L., Pinder, B.D., Vari, H.K., Lipshitz, H.D., and Smibert, C.A. (2005). Smaug Recruits the CCR4/POP2/NOT Deadenylase Complex to Trigger Maternal Transcript Localization in the Early Drosophila Embryo. Curr. Biol. *15*, 284–294.

Seydoux, G., and Braun, R.E. (2006). Pathway to totipotency: lessons from germ cells. Cell *127*, 891–904.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

Siddall, N.A., McLaughlin, E.A., Marriner, N.L., and Hime, G.R. (2006). The RNA-binding protein Musashi is required intrinsically to maintain stem cell identity. Proc. Natl. Acad. Sci. U. S. A. *103*, 8402–8407.

Siddiqui, N.U., Li, X., Luo, H., Karaiskakis, A., Hou, H., Kislinger, T., Westwood, J.T., Morris, Q., and Lipshitz, H.D. (2012). Genome-wide analysis of the maternal-to-zygotic transition in Drosophila primordial germ cells. Genome Biol. *13*, R11.

Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell *151*, 964–980.

Skvortsova, K., Tarbashevich, K., Stehling, M., Lister, R., Irimia, M., Raz, E., and Bogdanovic, O. (2019). Retention of paternal DNA methylome in the developing zebrafish germline. Nat. Commun. *10*, 3054.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.-K., Sheffield, N.C., Gräf, S., Huss, M., Keefe, D., et al. (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. *21*, 1757–1767.

Sprangers, R., Groves, M.R., Sinning, I., and Sattler, M. (2003). High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. J. Mol. Biol. *327*, 507–520.

Strasser, M.J., Mackenzie, N.C., Dumstrei, K., Nakkrasae, L.-I., Stebler, J., and Raz, E. (2008). Control over the morphology and segregation of Zebrafish germ cell granules during embryonic development. BMC Dev. Biol. *8*, 58.

Strome, S., and Lehmann, R. (2007). Germ versus soma decisions: lessons from flies and worms. Science *316*, 392–393.

Strome, S., and Updike, D. (2015). Specifying and protecting germ cell fate. Nat. Rev. Mol. Cell Biol. *16*, 406–416.

Strome, S., and Wood, W.B. (1983). Generation of asymmetry and segregation of germ-line granules in early C. elegans embryos. Cell *35*, 15–25.

Strome, S., Martin, P., Schierenberg, E., and Paulsen, J. (1995). Transformation of the germ line into muscle in mes-1 mutant embryos of C. elegans. Dev. Camb. Engl. *121*, 2961–2972.

Suzuki, H., Maegawa, S., Nishibu, T., Sugiyama, T., Yasuda, K., and Inoue, K. (2000). Vegetal localization of the maternal mRNA encoding an EDEN-BP/Bruno-like protein in zebrafish. Mech. Dev. *93*, 205–209.

Suzuki, H., Saba, R., Sada, A., and Saga, Y. (2010). The Nanos3-3'UTR is required for germ cell specific NANOS3 expression in mouse embryos. PloS One *5*, e9300.

Tada, H., Mochii, M., Orii, H., and Watanabe, K. (2012). Ectopic formation of primordial germ cells by transplantation of the germ plasm: direct evidence for germ cell determinant in Xenopus. Dev. Biol. *371*, 86–93.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Dev. Camb. Engl. *136*, 3033–3042.

Tadros, W., Goldman, A.L., Babak, T., Menzies, F., Vardy, L., Orr-Weaver, T., Hughes, T.R., Westwood, J.T., Smibert, C.A., and Lipshitz, H.D. (2007). SMAUG is a major regulator of maternal mRNA destabilization in Drosophila and its translation is activated by the PAN GU kinase. Dev. Cell *12*, 143–155.

Taguchi, A., Takii, M., Motoishi, M., Orii, H., Mochii, M., and Watanabe, K. (2012). Analysis of localization and reorganization of germ plasm in Xenopus transgenic line with fluorescence-labeled mitochondria. Dev. Growth Differ. *54*, 767–776.

Takeuchi, Y., Molyneaux, K., Runyan, C., Schaible, K., and Wylie, C. (2005). The roles of FGF signaling in germ cell migration in the mouse. Dev. Camb. Engl. *132*, 5399–5409.

Tam, P.P.L., and Zhou, S.X. (1996). The Allocation of Epiblast Cells to Ectodermal and Germ-Line Lineages Is Influenced by the Position of the Cells in the Gastrulating Mouse Embryo. Dev. Biol. *178*, 124–132.

Tang, W.W.C., Kobayashi, T., Irie, N., Dietmann, S., and Surani, M.A. (2016). Specification and epigenetic programming of the human germ line. Nat. Rev. Genet. *17*, 585–600.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. Science *338*, 671–675.

Theusch, E.V., Brown, K.J., and Pelegri, F. (2006). Separate pathways of RNA recruitment lead to the compartmentalization of the zebrafish germ plasm. Dev. Biol. *292*, 129–141.

Thomas, M.C., and Chiang, C.-M. (2006). The general transcription machinery and general cofactors. Crit. Rev. Biochem. Mol. Biol. *41*, 105–178.

Tiwari, B., Kurtz, P., Jones, A.E., Wylie, A., Amatruda, J.F., Boggupalli, D.P., Gonsalvez, G.B., and Abrams, J.M. (2017). Retrotransposons Mimic Germ Plasm Determinants to Promote Transgenerational Inheritance. Curr. Biol. CB *27*, 3010-3016.e3.

Torres-Padilla, M.-E., Bannister, A.J., Hurd, P.J., Kouzarides, T., and Zernicka-Goetz, M. (2006). Dynamic distribution of the replacement histone variant H3.3 in the mouse oocyte and preimplantation embryos. Int. J. Dev. Biol. *50*, 455–461.

Tóth, K.F., Pezic, D., Stuwe, E., and Webster, A. (2016). The piRNA Pathway Guards the Germline Genome Against Transposable Elements. Adv. Exp. Med. Biol. *886*, 51–77.

Tsompana, M., and Buck, M.J. (2014). Chromatin accessibility: a window into the genome. Epigenetics Chromatin *7*, 33.

Updike, D.L., Hachey, S.J., Kreher, J., and Strome, S. (2011). P granules extend the nuclear pore complex environment in the C. elegans germ line. J. Cell Biol. *192*, 939–948.

Vagin, V.V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P.D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. Science *313*, 320–324.

Van Doren, M., Broihier, H.T., Moore, L.A., and Lehmann, R. (1998). HMG-CoA reductase guides migrating primordial germ cells. Nature *396*, 466–469.

Varga, Z.M., Lawrence, C., Ekker, S.C., and Eisen, J.S. (2016). Universal Healthcare for Zebrafish. Zebrafish *13 Suppl 1*, S1-4.

Vasileva, A., Tiedau, D., Firooznia, A., Müller-Reichert, T., and Jessberger, R. (2009). Tdrd6 is required for spermiogenesis, chromatoid body architecture, and regulation of miRNA expression. Curr. Biol. CB *19*, 630–639.

Vastenhouw, N.L., Zhang, Y., Woods, I.G., Imam, F., Regev, A., Liu, X.S., Rinn, J., and Schier, A.F. (2010). Chromatin signature of embryonic pluripotency is established during genome activation. Nature *464*, 922–926.

Venkatarama, T., Lai, F., Luo, X., Zhou, Y., Newman, K., and King, M.L. (2010). Repression of zygotic gene expression in the Xenopus germline. Development *137*, 651–660.

Voigt, P., Tee, W.-W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev. *27*, 1318–1338.

von Meyenn, F., and Reik, W. (2015). Forget the Parents: Epigenetic Reprogramming in Human Germ Cells. Cell *161*, 1248–1251.

Wang, H., Liu, Y., Ye, D., Li, J., Liu, J., and Deng, F. (2016). Knockdown of zebrafish Nanog increases primordial germ cells during early embryonic development. Dev. Growth Differ. *58*, 355–366.

Weber, S., Eckert, D., Nettersheim, D., Gillis, A.J.M., Schäfer, S., Kuckenberg, P., Ehlermann, J., Werling, U., Biermann, K., Looijenga, L.H.J., et al. (2010). Critical function of AP-2 gamma/TCFAP2C in mouse embryonic germ cell maintenance. Biol. Reprod. *82*, 214–223.

Weidinger, G., Wolke, U., Köprunner, M., Klinger, M., and Raz, E. (1999). Identification of tissues and patterning events required for distinct steps in early migration of zebrafish primordial germ cells. Dev. Camb. Engl. *126*, 5295–5307.

Weidinger, G., Stebler, J., Slanchev, K., Dumstrei, K., Wise, C., Lovell-Badge, R., Thisse, C., Thisse, B., and Raz, E. (2003). dead end, a novel vertebrate germ plasm component, is required for zebrafish primordial germ cell migration and survival. Curr. Biol. CB *13*, 1429–1434.

White, R.J., Collins, J.E., Sealy, I.M., Wali, N., Dooley, C.M., Digby, Z., Stemple, D.L., Murphy, D.N., Billis, K., Hourlier, T., et al. (2017). A high-resolution mRNA expression time course of embryonic development in zebrafish. ELife *6*.

Wilhelm, J.E., Hilton, M., Amos, Q., and Henzel, W.J. (2003). Cup is an eIF4E binding protein required for both the translational repression of oskar and the recruitment of Barentsz. J. Cell Biol. *163*, 1197–1204.

Williamson, A., and Lehmann, R. (1996). Germ cell development in Drosophila. Annu. Rev. Cell Dev. Biol. *12*, 365–391.

Wolffe, A.P., and Guschin, D. (2000). Review: chromatin structural features and targets that regulate transcription. J. Struct. Biol. *129*, 102–122.

Wu, G., Han, D., Gong, Y., Sebastiano, V., Gentile, L., Singhal, N., Adachi, K., Fischedick, G., Ortmeier, C., Sinn, M., et al. (2013). Establishment of totipotency does not depend on Oct4A. Nat. Cell Biol. *15*, 1089–1097.

Wu, S.-F., Zhang, H., and Cairns, B.R. (2011). Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Res. *21*, 578–589.

Wylie, C.C., Holwill, S., O'Driscoll, M., Snape, A., and Heasman, J. (1985). Germ plasm and germ cell determination in Xenopus laevis as studied by cell transplantation analysis. Cold Spring Harb. Symp. Quant. Biol. *50*, 37–43.

Yamaguchi, S., Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2005). Nanog expression in mouse germ cell development. Gene Expr. Patterns *5*, 639–646.

Yamaguchi, S., Kurimoto, K., Yabuta, Y., Sasaki, H., Nakatsuji, N., Saitou, M., and Tada, T. (2009). Conditional knockdown of Nanog induces apoptotic cell death in mouse migrating primordial germ cells. Dev. Camb. Engl. *136*, 4011–4020.

Yamaji, M., Seki, Y., Kurimoto, K., Yabuta, Y., Yuasa, M., Shigeta, M., Yamanaka, K., Ohinata, Y., and Saitou, M. (2008). Critical function of Prdm14 for the establishment of the germ cell lineage in mice. Nat. Genet. *40*, 1016–1022.

Ying, Y., and Zhao, G.Q. (2001). Cooperation of endoderm-derived BMP2 and extraembryonic ectoderm-derived BMP4 in primordial germ cell generation in the mouse. Dev. Biol. *232*, 484–492.

Yoon, C., Kawakami, K., and Hopkins, N. (1997). Zebrafish vasa homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells. Dev. Camb. Engl. *124*, 3157–3165.

Yu, J., Angelin-Duclos, C., Greenwood, J., Liao, J., and Calame, K. (2000). Transcriptional repression by blimp-1 (PRDI-BF1) involves recruitment of histone deacetylase. Mol. Cell. Biol. *20*, 2592–2603.

Yuan, G.-C., Liu, Y.-J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626–630.

Zalokar, M. (1976). Autoradiographic study of protein and RNA formation during early development of Drosophila eggs. Dev. Biol. *49*, 425–437.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016). Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature *537*, 553–557.

Zhang, B., Wu, X., Zhang, W., Shen, W., Sun, Q., Liu, K., Zhang, Y., Wang, Q., Li, Y., Meng, A., et al. (2018). Widespread Enhancer Dememorization and Promoter Priming during Parental-to-Zygotic Transition. Mol. Cell *72*, 673-686.e6.

Zhang, F., Barboric, M., Blackwell, T.K., and Peterlin, B.M. (2003). A model of repression: CTD analogs and PIE-1 inhibit transcriptional elongation by P-TEFb. Genes Dev. *17*, 748–758.

Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B.S., Schultz, N., Vreven, T., Meignin, C., Davis, I., Zamore, P.D., et al. (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell *151*, 871–884.

Zhang, W., Liu, Y., Zhang, M., Zhu, C., and Lu, Y. (2017). Quantitative reproducibility analysis for identifying reproducible targets from high-throughput experiments. BMC Syst. Biol. *11*, 73.

Zylka, M.J., Simon, J.M., and Philpot, B.D. (2015). Gene length matters in neurons. Neuron *86*, 353–355.