



**André Figueiredo
Rendeiro**

**Regulation of *Oikopleura dioica*'s alternative cell
cycle modes**

**Regulação dos ciclos celulares alternativos de
*Oikopleura dioica***



Declaração

Declaro que este relatório é integralmente da minha autoria, estando devidamente referenciadas as fontes e obras consultadas, bem como identificadas de modo claro as citações dessas obras. Não contém, por isso, qualquer tipo de plágio quer de textos publicados, qualquer que seja o meio dessa publicação, incluindo meios eletrónicos, quer de trabalhos académicos.



**André Figueiredo
Rendeiro**

**Regulation of *Oikopleura dioica*'s alternative cell
cycle modes**

**Regulação dos ciclos celulares alternativos de
*Oikopleura dioica***

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica de Manuel António da Silva Santos, Professor Associado do Departamento de Biologia da Universidade de Aveiro e de Eric Malcolm Thompson, Professor do Departamento de Biologia da Universidade de Bergen e do Sars International Centre for Marine Molecular Biology.

o júri

presidente

Doutora Maria de Lourdes Gomes Pereira
Professor Associado da Universidade de Aveiro

arguente

Doutor Rui Miguel Pinheiro Vitorino
Investigador Auxiliar da Universidade de Aveiro

orientador

Doutor Manuel António da Silva Santos
Professor Associado da Universidade de Aveiro

acknowledgements

I thank Prof. Thompson for allowing me to work in his lab and for all the help with my stay in Norway. Pavla and Gemma, for support and guidance during this period, and everyone in the lab for help when needed.

I'd like to thank my family for all the support throughout these years, allowing me to follow my dreams.

This thesis was made entirely with free open source software. I'd like to acknowledge the free open source software community for its marvellous efforts in building an open world and Tomás Oliveira e Silva for the University of Aveiro \LaTeX thesis template.

palavras-chave

ciclo celular, endociclos, E2F, *Oikopleura dioica*, H3K79me, Dot1

resumo

O ciclo celular eucariótico é um dos processos biológicos mais estudados. No entanto, variações ao ciclo celular canónico têm sido descobertas e são mais predominantes do que anteriormente previsto. Endoreplicação e endociclos - duas variantes ao ciclo celular canónico - produzem células poliplóides, o que confere vantagens em crescimento e em resistência a stress genotóxico. Apesar de alguns dos princípios regulatórios de endociclos estarem a ser descobertos, ainda pouco é sabido sobre como podem células transitar do ciclo mitótico para eles. Recentemente foi proposto que a metilação da lisina 79 na histona H3 (H3K79me) possa ter uma função distinta da maior parte das modificações de histonas - ao ser um relógio para a idade celular - pois todos os estados de metilação de H3K79 são depositados numa maneira distributiva (ao contrário de todas as outras lisinas metiladas) pela metiltransferase Dot1, quando nenhuma demetilase é conhecida. Estudos funcionais da Dot1 mostraram padrões aberrantes de replicação, ligando-a directamente à capacidade de regulação dos ciclos celulares. Este trabalho dedica-se ao estudo dos mecanismos de transição entre os diferentes modos de ciclo celular em *Oikopleura dioica* - um organismo marinho do filo *Chordata*, classe *Appendicularia* - que usa endociclos extensivamente ao longo do seu ciclo de vida. Realizámos ChIP-seq em dois factores de transcrição (E2F1 e E2F7) antagonistas em ciclos mitóticos e endociclos. Sabe-se que estes factores estão envolvidos no controlo da progressão do ciclo celular e que se sabe têm diferentes padrões de regulação nos diferentes ciclos. Os resultados revelam um grande proporção de co-localização spatio-temporal dos dois factores de transcrição em proximidade a sítios de iniciação de transcrição, regulando genes com funções no controlo de replicação e do ciclo celular. Proposemo-nos a um estudo inicial de caracterização funcional de H3K79me em *Oikopleura* através da inibição de Dot1 em estados de desenvolvimento que usam diferentes modos de ciclo celular. Isto provocou desregulação dos padrões de proliferação celular nos diferentes tipos de ciclo celular, mostrando uma contribuição de H3K79me para a sua regulação em *Oikopleura*.

keywords

cell cycle, endocycles, E2F, *Oikopleura dioica*, H3K79me, Dot1

abstract

The eukaryotic cell cycle is one of the most studied biological processes. However, variations of the canonical cell cycle have been discovered and found to be more predominant than previously expected. Endoreplication and endocycling - two such variants - produce polyploid cells, conferring advantages in growth and genotoxic stress. Although some of the regulatory principles of these cycles are being discovered, little is known about how cells can transition from a mitotic cell cycle to them. Recently, it has been proposed that methylation of histone 3 lysine 79 (H3K79me) might have a function distinct from most histone modifications - a timer for a cell's age -, for methylation of all H3K79 states is performed in a distributive fashion (unlike all other methylated lysines in histones) by Dot1 and no demethylase is known. Functional studies on Dot1 have shown aberrant patterns of replication, linking it directly to a cell's ability to regulate its cell cycle. This project aims to study the mechanisms of transition between different cell cycle modes in *Oikopleura dioica* - a marine chordate employing endocycling extensively through its life cycle. We performed ChIP-seq on two antagonist transcription factors (E2F1 and E2F7) involved in the control of cell cycle progression that have been shown to have differential patterns of regulation in mitotic and endocycling cell cycle modes. Results reveal a high degree of spatial and temporal occupancy of both transcription factors in close proximity to transcription start sites, regulating genes with a function highly enriched for the control of replication and cell cycle. We set ourselves to perform an initial functional study of H3K79me in *Oikopleura* through Dot1 inhibition in developmental stages employing different cell cycle modes. This caused deregulation of proliferation patterns in different ways in the different cell cycle modes, highlighting a contribution of H3K79me to their regulation in *Oikopleura*.

Contents

Contents	i
List of Figures	vii
List of Tables	ix
Abbreviations	xi
1 Introduction	1
1.1 Polyploid cell cycles - endoreplication and endocycles	1
1.1.1 An overview of the eukaryotic canonical cell cycle	2
1.1.1.1 Regulation of cell cycle progression by Cyclin-CDK complexes	2
1.1.2 Endocycle regulation	4
1.1.2.1 Abolishment of mitosis and cell division	4
1.1.2.2 Alternative regulation of cell cycle effectors	5
1.1.2.2.1 Endoreplication by CDK activity regulation	5
1.1.2.2.2 E2F factors in CDK activity regulation	5
1.2 Histone modifications in cell cycle regulation	7
1.2.1 The role of H2Bub in DNA replication	8
1.2.2 Dot1 and H3K79 methylation in the cell cycle	9
1.3 <i>Oikopleura dioica</i> , a marine chordate extensively employing endocycles	13
1.3.1 <i>Oikopleura</i> 's life cycle	13
1.3.2 The genome of <i>Oikopleura dioica</i>	14
1.3.3 Histone variants and modifications of <i>Oikopleura dioica</i>	15
1.3.4 The cell cycle modes of <i>Oikopleura dioica</i>	16
1.3.5 Growth arrest	16
1.4 Project Goals	18
2 Materials and Methods	19
2.1 Materials	19
2.1.1 Antibodies	19
2.1.2 Chemicals and reagents	20
2.1.3 Consumables	20
2.1.4 Instruments and equipment	20
2.1.5 Buffers and solutions	21
2.1.5.1 ChIP	21
2.1.5.2 Western blot	21

Contents

2.1.5.3	Immunohistochemistry, TUNEL assay and EdU incorporation and detection	21
2.1.5.4	TUNEL assay	21
2.2	Animal culture and collection	22
2.2.1	Culture of <i>O. dioica</i>	22
2.2.2	Collection of <i>O. dioica</i>	22
2.2.2.1	Tailbud stage	22
2.2.2.2	Day two stage	22
2.2.2.3	Day six, immature stage	23
2.3	ChIP-seq	23
2.3.1	Animal fixation	23
2.3.2	Cell lysis and chromatin sonication	23
2.3.3	Chromatin quality assessment	24
2.3.4	Immunoprecipitation	24
2.3.5	Illumina library construction and high-throughput sequencing	25
2.4	Western blot	25
2.4.1	Sample preparation	25
2.4.2	SDS-PAGE	25
2.4.3	Protein transference	25
2.4.4	Blotting	25
2.4.5	Detection	26
2.5	Whole mount Immunohistochemistry	26
2.6	Dot1 inhibition	26
2.7	TUNEL assay	27
2.8	EdU incorporation and detection	27
2.9	Data analysis	28
2.9.1	Data preprocessing	28
2.9.2	Finding of ChIP-enriched regions	28
2.9.3	Genomic characterization of ChIP-enriched regions	28
2.9.4	Transcription factor motif finding in ChIP-enriched regions	28
2.9.5	Local epigenetic landscape of ChIP-enriched regions	28
2.9.6	Assignment of ChIP-enriched regions to genes	28
2.9.7	Testing of significant Gene Ontology terms	28
2.9.8	Expression levels of genes	29
3	Results	31
3.1	E2F factors in <i>Oikopleura</i> 's late developmental stages	31
3.1.1	Detection of E2F binding through ChIP-chip	32
3.1.1.1	Validation of E2F antibodies	32
3.1.1.2	Genomic context of E2F binding sites	32
3.1.1.3	Insights into the mechanism of E2F binding	35
3.1.1.4	Spatial and temporal co-localization of E2F factors	37
3.1.2	Putative E2F regulated genes (target genes)	38
3.1.2.1	Gene ontology of target genes	38
3.1.2.2	Differences of expression in E2F target genes	40
3.1.2.3	E2F regulation of key cell cycle effectors	40
3.2	Dot1 and H3K79 methylation in <i>Oikopleura dioica</i> 's cell cycle modes	42

Contents

3.2.1	Use of H3K79me along <i>Oikopleura</i> 's life cycle	42
3.2.2	Functional studies on Dot1 in two cell cycle modes of <i>Oikopleura</i>	43
4	Discussion	49
4.1	E2F factors in <i>Oikopleura</i> 's late developmental stages	49
4.2	Dot1 and H3K79 methylation in <i>Oikopleura dioica</i> 's cell cycle modes	51
5	Conclusion and Future Perspectives	53
	Bibliography	55
	Appendix A Gene Ontology terms associated with genes putatively regulated by E2F factors	61
A.1	Genes bound only by E2F1 in ovary tissue	61
A.2	Genes bound only by E2F7 in ovary tissue	62
A.3	Genes bound only by E2F1 in testis tissue	63
A.4	Genes bound only by E2F7 in testis tissue	63
	Appendix B List of key cell cycle effectors putatively regulated by E2F factors	65

List of Figures

1.1	Variants of cell cycles employed by eukaryotic organisms	2
1.2	Cyclin-CDK complexes during progression of cell cycle phases	3
1.3	Cyclin-CDK phosphorylation of pRb and release of the E2F TF during G1 phase of the cell cycle	3
1.4	The phases of the canonical eukaryotic cell cycle	4
1.5	The E2F family of transcription factors in mammals	5
1.6	Oscillation of opposing networks of cell cycle genes during endocycles	6
1.7	Schema of the dynamics of H3K79 methylation during a cell's cycle.	11
1.8	Phylogenetic position of the Tunicate group and <i>Oikopleura</i>	13
1.9	Development of <i>Oikopleura</i> and the predominantly employed modes of cell cycle	14
1.10	Goals	18
3.1	Expression levels of E2F factors during <i>Oikopleura</i> 's development	32
3.2	Detection of E2F proteins in two developmental stages of <i>Oikopleura</i>	32
3.3	Comparison of the distribution of genomic features in <i>Oikopleura</i> and other animal models	33
3.4	Genomic context of E2F ChIP-enriched regions	34
3.5	Influence of sequence specificity and local chromatin environment in E2F TF binding	36
3.6	Portion of E2F ChIP-ERs shared within and between developmental stages .	37
3.7	Gene Ontology of genes putatively regulated by both E2F transcription factors	39
3.8	Expression values of all genes and genes exclusively targeted by one E2F TF in ovary and testis tissue	41
3.9	Expression values of Dot1 and Bre1 enzymes during optimal <i>Oikopleura</i> development	43
3.10	H3K79 methylation states during selected developmental stages of <i>Oikopleura</i> and in Dot1 ⁱⁿ -treated animals	44
3.11	Effects of the treatment with Dot1 ⁱⁿ in Tadpole animals	45
3.12	Assessment of the proliferative state of animals during a predominant endocycle mode of by detection of EdU incorporation. Green: EdU; Blue: DNA;	47

List of Tables

2.1	Antibodies	19
2.2	Chemicals and reagents of particular relevance	20
2.3	Consumables of particular relevance	20
2.4	Instruments of particular relevance	20
2.5	Feeding regime of <i>Oikopleura</i>	22
2.6	Fixation conditions for each developmental stage of <i>Oikopleura</i>	23
2.7	Settings used on the sonicator for <i>Oikopleura</i> chromatin	24
2.8	Reaction mix used in detection of EdU	27
3.1	N° of ChIP-enriched regions detected for all samples	33
3.2	N° of target genes from each TF	38

Abbreviations

aa Amino acid

APC/C Anaphase promoting complex

bp Nucleotide base pair

CDK Cyclin-dependent kinase

CDS Coding sequence of genes

ChIP Chromatin immunoprecipitation

ChIP-chip Chromatin immunoprecipitation followed by DNA microarray hybridization

ChIP-ER ChIP-enriched region

ChIP-seq Chromatin immunoprecipitation followed by high-throughput DNA sequencing

CKI Cyclin-dependent kinase inhibitor

DNA Deoxyribonucleic acid

GSCs Germline stem cells

GA Growth arrest

kD Kilo Dalton

O/N Over night

R/T Room temperature

RC pre-Replication Complex

TF Transcription factor

UTR Untranslated regions of genes

Units of the International System of Units (SI) are assumed.

Chapter 1

Introduction

1.1 Polyploid cell cycles - endoreplication and endocycles

Polyploid cells possess more than two pairs of their set of chromosomes. This polyploid state is common to be predominant within fungi, plants and some vertebrates like fish and amphibians, but also within restricted cell types in many other organisms [1].

Documented advantages of polyploid organisms compared to diploids include heterosis (an evolutionary situation where the progeny has fitness advantages compared to its progenitors), gene redundancy, and gain of asexual reproduction by loss of self-incompatibility [2]. From a strict cellular perspective, gene redundancy is the most relevant - it provides a way of masking the effect of deleterious alleles, and shields the possible effect of mutagens in the DNA. On the other hand, the increase of the gene-space available for mutations to act can also be seen as an evolutionary advantage [2].

Changing ploidy within a cell has tremendous consequences for its physiology. Additional genetic material, when active, will produce more products which will contribute to increase the cell volume and this inevitably changes the general cellular architecture. Higher number of genomic loci will bring new challenges in nuclear organization which can alter the epigenetic stability and consequently, gene expression. Cell division can be seriously compromised due to the inadequacy of the cell division machinery in dealing with an abnormal number of chromosomes. Like anything exposed to selection, only the sum of all these attributes will predict the outcome of the success of the organism, and while some of these events may be seriously restrictive to an organism survival in a specific environment, in others it might confer advantages [2] [3].

Polyploidy as a cellular state is known for more than a century, but studies of it under the light of endoreplication, a cell cycle mode which relies on cycles of genome replication without cell division, only recently began to gain relevance. Two main ways exist for the occurrence of endoreplication: endocycling, where genome replication in the S phase of the cell cycle alternates with a G phase preparing for the next replication; endomitosis, where an abortive mitotic cell division attempt after genome replication returns to a new G phase (Figure 1.1) [3]. Since the outcome of both is the same - the creation of a single polyploid cell - the distinction between the two can sometimes be unclear [1]. Most remarkably, the core protein machinery that drives and regulates the canonical mitotic cell cycle performs as well in the alternative cycles.

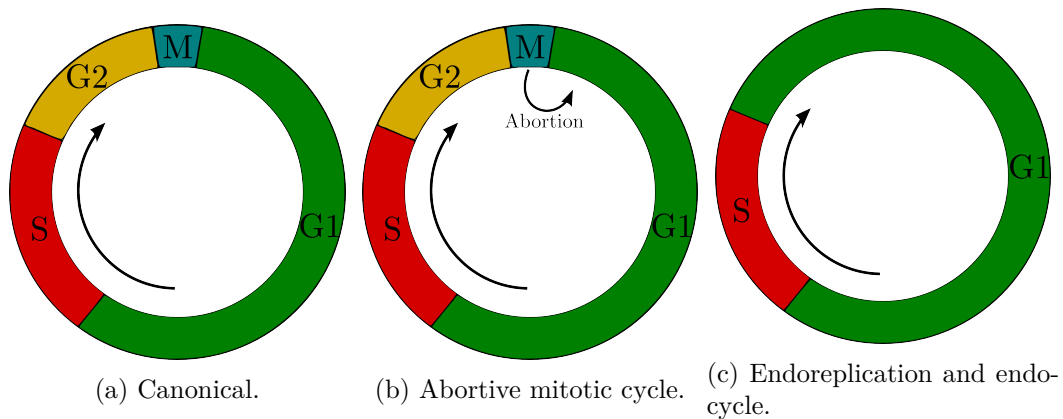


Figure 1.1: Variants of cell cycles employed by eukaryotic organisms.

1.1.1 An overview of the eukaryotic canonical cell cycle

Cell division is a mechanism common to all organisms, being necessary for reproduction and also used for growth and proliferation in multicellular organisms. Regulation of the whole cell cycle and cell division is intricately complex, but is necessary to assure progression into the several phases in a controlled manner. Multicellular eukaryotes in particular, require precise timing and sensing of internal and environmental clues, which is regulated by complex cellular machinery, to achieve correct tissue and organ development.

The canonical cell cycle consists of four major phases: gap phase one (G1) - cell growth and synthesis of required substrates for DNA synthesis; synthesis (S) - genome replication; gap phase two (G2) - continued cell growth; mitotic phase (M) - cell division through mitosis. Progression through the cell cycle phases is controlled by a number of checkpoints where genomic and cellular integrity are attested and various internal and external cues are considered, allowing progression, or arresting cell cycle [4].

Numerous protein classes interact during regulation of the canonical cell cycle, with the complexes formed by Cyclins and Cyclin-dependent kinases (CDK) very prominent. Different CDK-Cyclin complexes are activated throughout the cell cycle, and their specific temporal regulation allows modulation of the activity of specific downstream effectors of cycle progression. While CDKs remain relatively constant through the cycle, the level of the Cyclin subunit is regulated thus restricting the complexes' temporal activity [5].

1.1.1.1 Regulation of cell cycle progression by Cyclin-CDK complexes

In general, the overall activity of CDK-Cyclin complexes increases from G1 to M, with several thresholds of their activity used by the cell to identify the phase it is in. An overview of the major CDK-Cyclin complexes during the phases of the cell cycle can be seen in Figure 1.2.

The G1 phase is the one with lowest CDK-Cyclin activity. This allows the cell to start replication with the assembly of the replication-specific proteins into a pre-replication complex in DNA sites that will become origins of replication. The assembly of pre-replication complexes during the G1 phase licenses replication and when CDK activity is detected above a threshold, the S phase begins with DNA replication. High CDK activity activates replication but also prevents replication licensing during other phases, for it requires low activity.

1.1. Polyploid cell cycles - endoreplication and endocycles

A unique cell division per cycle is thus assured, for only when CDK activity is abolished at the end of mitosis can the genome be licensed for replication again [4].

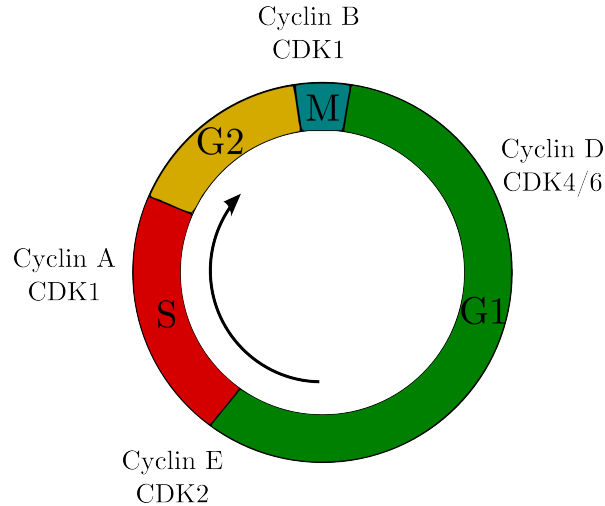


Figure 1.2: Cyclin-CDK complexes during progression of cell cycle phases.

In the G1 phase, the D cyclins are expressed in response to external cues such as growth factors. These associate with both CDK 4 and 6 and phosphorylate downstream targets which most importantly include the retinoblastoma protein (pRb). pRb natively binds to a complex formed by a member of the E2F family of transcriptions factors (TF) and DP (also a TF), inhibiting it. When phosphorylated, pRb releases the E2F-DP TF complex and this is free to bind DNA and activate the expression of downstream genes that are responsible for entry into the S phase (Figure 1.3) [6] [4].

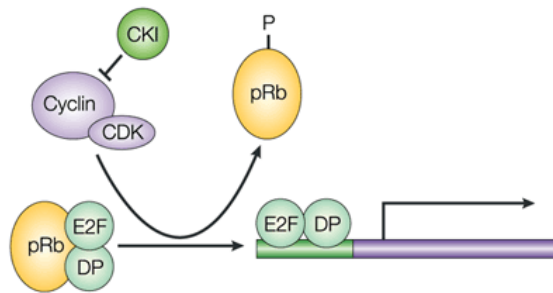


Figure 1.3: Cyclin-CDK phosphorylation of pRb and release of the E2F TF during G1 phase of the cell cycle. Adapted from [7]

Among the genes regulated by the E2F TF upon S phase entry is Cyclin A. This Cyclin first interacts with CDK2 and this complex is required for successful completion of S phase. Later, Cyclin A complexes with CDK1 having a function in G2/M phases transition.

In the mitotic phase the predominant Cyclin-CDK complex is the Cyclin B-CDK1, which has important functions in phosphorylating cytoskeletal proteins involved in mitosis.

Towards the end of cell division, cyclins are degraded, therefore greatly reducing the overall level of CDK activity. Both A and B cyclins are ubiquitinated by the ubiquitin ligase anaphase-promoting complex/cyclosome (APC/C), which includes the Fzr/Cdh1 subunit.

This subunit confers Cyclin A and B specificity, marking them for proteolytic degradation starting at the end of mitosis but continuing through the G1 phase, which helps to ensure unidirectional cell cycle progression [4].

Cyclin-CDK activity can also be modulated by Cyclin-dependent kinase inhibitors (CKI). Among the most prominent are p27 which inhibits CDK4 in the pre-S phase Cyclin D-CDK4 complexes - therefore preventing phosphorylation of pRb and E2F release - and CDK 2 in the Cyclin E-CDK2 complex causing the cell to stay in S phase [8].

Downstream effects of Cyclin-CDK activity involve activation of various cell cycle effectors including cyclins necessary for the subsequent phase, and this is fundamental for all cell cycle events to occur, especially in the beginning of the S phase, and during mitosis. This assures that the latter cannot start without the first and vice-versa, and that once they have started they cannot be reversed [6]. These concepts can be visualized in Figure 1.4 and will prove useful when considering the implications of an endoreplicative cell cycle - a cycle of G and S phases.

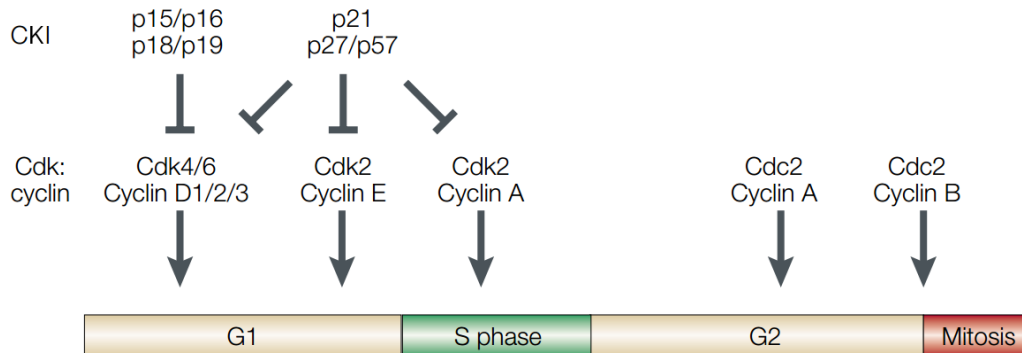


Figure 1.4: The phases of the canonical eukaryotic cell cycle. Adapted from [6]

1.1.2 Endocycle regulation

Much less is known about the regulation of alternative modes of cell cycles, endocycling in particular. Compared with the canonical cell cycle regulation, two prominent events must occur for an endocycle to take place:

1. abolishment of mitosis and cell division;
2. alternative regulation of cell cycle effectors to continue to allow genome replication despite that there has been no cell division.

1.1.2.1 Abolishment of mitosis and cell division

Bypassing mitosis and cell division in endocycles can be accomplished in several ways. One such way involves the exploitation of the regular regulation mechanism in the canonical cell cycle through modulation of the Cyclin-CDK regulators. The APC/C subunit, Frz/Cdh1 is used in *Drosophila* and *Arabidopsis* as the way to achieve this, and its continued expression is in fact sufficient to induce endoreplication [9]. Its expression throughout the endocycles is required to continue suppression of pro-mitotic effectors, in particular Cyclin A and B.

1.1. Polyploid cell cycles - endoreplication and endocycles

Evidence from mammalian cells, specifically placental trophoblast giant cells (TGC) and megakaryocytes in the bone marrow - both employ endocycling - suggest that Cyclin kinase inhibitors (CKI), proteins that act by direct binding to CDKs, also promote endocycling. This happens, with the activation of the p57 protein (a CKI), which inhibits CDK1 and prevents it from actively promoting entry into mitosis [8].

Blocking cytokinesis is a way to inhibit cell division and thus leads to endoreplication, a mechanism known to be normal in plants, where it has even been used in horticulture to improve crops. RhoA, a GTPase with a key role in regulating cell division is not activated due to the downregulation of two of its activators (GEF-H1 and ECT2) during cytokinesis in megakaryocytes as well, which causes failure of cell division. Endoreplication through an incomplete mitosis, is also known as endomitosis [1].

1.1.2.2 Alternative regulation of cell cycle effectors

1.1.2.2.1 Endoreplication by CDK activity regulation

The Cyclin E-CDK2 complex is the main driver for S phase entry in the canonical cell cycle, although in the absence of CDK2, CDK1 can act as a substitute in mammalian cells [10]. Evidence suggests that modulation of Cyclin E expression during endoreplication in *Drosophila* is crucial and possibly sufficient for its maintenance [11], and megakaryocytes with Cyclin E overexpression show increased ploidy [12], reinforcing the role of Cyclin E complexed with a CDK in endoreplication. The Cyclin E-CDK complex has also been shown to be responsible for the inhibition of the pre-replication complex (RC) protein Orc1 indirectly, by inhibiting the APC/C component Fzr/Cdh1 [13].

All these observations put the Cyclin E-CDK2 complex in the center of regulation of endoreplicative cell cycles: it seems to generally contribute to increased replication, but also inhibits elements necessary for it, indicating that a temporally cyclic regulation of expression is necessary. Cyclin E is present at both transcript and protein level just before and during S phase, but not G, in endoreplicating *Drosophila* cells [14] and if expressed continually, endoreplication is suppressed [15].

1.1.2.2.2 E2F factors in CDK activity regulation

The E2F family of TFs is comprised of nine TFs in mammalian organisms, which can be grouped in a set of seven canonical and two atypical TFs (Figure 1.5) [16]. In this set there is a recent gene duplication of E2F3 into E2F3a and E2F3b. Canonical E2F TFs (E2F1 to 6) dimerize with the DP co-transcriptional factor and are the most studied. Atypical E2Fs (E2F7 and 8) lack the domain of interaction with DP and therefore act independently of it. They have however two DNA-binding domains, and have been shown to homo- and heterodimerize. In both canonical and atypical E2Fs, there is functional specialization: E2F1 to 3a are activators, while E2F3b to 8 are repressors of gene expression [16].



Figure 1.5: The E2F family of transcription factors in mammals. Adapted from [16]

The E2F family of TFs is well known to have important functions in the regulation of cell cycle progression in the canonical and endoreplicative cell cycles [17] [18] [19]. This is mostly understood due to the function of activator E2Fs in gene expression activation in the transition to S and M phases, which is also regulated by Rb 1.3. Repressor E2F TFs have been recently implying particularly in the regulation of alternative cell cycles. They do so by antagonizing the function of activator E2Fs by negatively regulating a highly overlapping set of genes that are being activated by the canonical activator E2Fs [18]. Cellular outcomes of this antagonism between the two groups of opposing E2Fs can be seen by functional experiments on atypical E2F: ablation of canonical activators in the two endocycling tissues of mammals, trophoblast giant cells (TGCs) in the placenta and hepatocytes in the liver, caused increased ploidy, whereas ablation of atypical repressors diminished ploidy [19]. Genes under regulation of both antagonistic groups of TFs are involved in the coordination of G2 to M transition, critical for mitosis, karyokinesis and cytokinesis. This antagonism is also present to some extent in human cells employing a canonical cell cycle, where E2F7 has been shown to repress a network of cell cycle genes to control S-phase progression. [20]

Since E2Fs are sequence-specific DNA binding proteins, the outcome of their activity is the activation or repression of their targets. These are the genes which will perform the molecular functions involved in progression of cell cycle phases.

E2F1 is most prominently known to target Cyclin E, activating its expression and inducing S phase entry when complexed with CDK2. Cyclin E-CDK2 in turn activate the ubiquitin ligase CRL4/Cdt2 which acts during DNA replication and ubiquitinates E2F1, leading to its degradation [9] [21] [22]. This negative feedback mechanism triggered by E2F1 itself creates an oscillation during endocycles that both promotes and is the consequence of the oscillation of other key effectors in cell cycle regulation. Other effectors are other ubiquitin ligases such as APC/C which is inhibited by the Cyclin E-CDK2 complex. It is important for APC/C to have its activity reduced, for it targets the Gemini protein, which is in its turn an inhibitor of Cdt1, an essential replication licensing protein. Cdt1 is degraded upon APC/C activity, and with its inhibition, is allowed to promote replication again. This has been confirmed by the observation that the combination of mitotic CDK1 inhibition and Geminin proteolysis are sufficient to trigger endoreplication in human cells [23].

Oscillation of these two groups of factors occurring between G and S phases is required for the establishment and maintenance of endocycles (Figure 1.6) [16] [24] [1], and provides a mechanism through which other important aspects of endoreplication can be controlled.

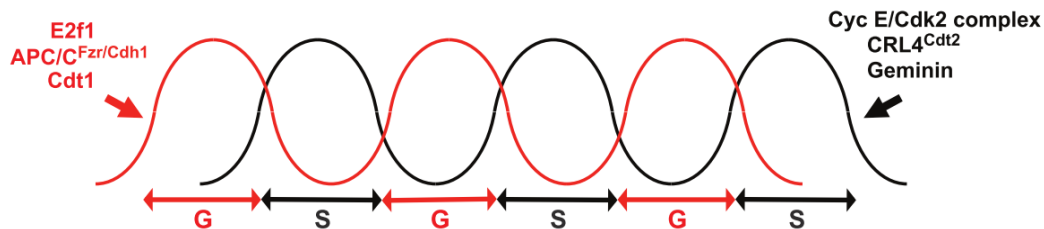


Figure 1.6: Oscillation of opposing networks of cell cycle genes during endocycles. Adapted from [1]

In summary, the oscillation of key cell cycle effectors in endocycles promoted by E2F factors, ensures the cycling of CDK activity and thus, the cycling of replication licensing that is at the core of endocycles. Nevertheless, the mechanism by which the oscillation of CDK activity is achieved may differ among species.

1.2 Histone modifications in cell cycle regulation

In the post-genomic era, characterizing the genetic repertoire of a model species is solely the first step to understanding of its biology from a molecular point of view. Gene regulation is currently the limelight of studies in Biology and Biomedicine and the characterization of gene expression and the mechanisms of its regulation pre- and post-transcription and translation is the current forefront of research. An important contribution to gene regulation comes from epigenetics. Long-non coding RNAs (lncRNAs), histone variants, histone modifications and modification of DNA bases (prominently methylation) are four great classes of mechanisms by which gene expression is altered epigenetically.

The fundamental structural unit of chromatin is the nucleosome - a structure formed by histone proteins with DNA wrapped around them. A core octamer of histones comprised of two copies of H2A, H2B, H3 and H4 histones each, forms the canonical nucleosome structure to which 147 base pairs of DNA are wrapped around twice. This basic structure repeats itself throughout the genome and provides the basis for higher-order structure in chromatin. The disposition of the histone proteins creates a disc-shaped structure resulting from the left-handed superhelical DNA structure. Since the first structural evidence of the nucleosome, more *in-vivo* nucleosome structures presenting different stoichiometries have been discovered and although in some cases their functional implications have yet to be shown, they have changed the way we look at nucleosomes - it can no longer be viewed as a static identity.

Genes of histone proteins involved in the formation of the canonical nucleosome have several specific conserved features. They lack introns and are organized in clusters. Due to the occurrence of a major cellular event requiring a high number of histone proteins - replication in S phase - histone genes are known to be replication dependent (RD). Their expression is coupled with the cellular event of replication by the use of a conserved stem-loop (SL) in the 3'UTR of the mRNAs, which promotes translation in this phase. Genes of variant histone proteins are often transcribed from orphan genes (not in clusters), can contain introns and lack the SL motif that couples translation to replication. Although considerable structural divergence exists, in most cases histone variants (also called replication-independent (RI) histones) are still fairly conserved due to being subject to the structural constraints promoted by the assembly into nucleosomes.

Post-translational modifications of histone proteins were discovered in the mid-1960s and are long known to have a functional role. Due to the fact that most post-translational modifications of histones were found on the protruding histone tails, initial theories suggested they affected the interactions between DNA and histone tails electrostatically. This less stringent DNA-histone link would loosen chromatin structure and make attachment and detachment possible, explaining changes in accessibility of DNA sequence to regulatory proteins and differential position of nucleosomes along the DNA sequence. Later, with better knowledge of the diversity of post-translational modifications of histones and the discovery of more functions associated with them, post-translational modifications of histones are now thought to function as epitopes for regulatory proteins to act on, marking discrete parts of the histone-bound DNA for a specific function. At the same time, discovery of post-translational modifications of histones nearer to the histone core still supported the idea that they also contribute to alter DNA-histone and histone-histone interactions.

Some post-translational modifications of histones are known to be correlated with gene silencing, others with gene activation, and some correlate with presence or activity of regulatory elements, transcriptional units or replication origins, but although many have been identified

and associated with these functions, the exact mechanism by which they achieve this is not entirely known. Common post-translational modifications in histone proteins are methylation, acetylation or ubiquitination of lysines, serine phosphorylation or arginine methylation, but others exist as well [25] [26] [27].

This array of chemical varieties in histones is therefore heavily loaded with information that has been shown to condition gene expression and contribute to triggering of major cellular events (*e.g.* cell division). This information is stably maintained and passed to daughter cells upon cell division which is important for the maintenance of regulatory status along cell division and differentiation. These processes are therefore considered epigenetic because they influence gene expression and are inheritable by daughter cells.

The cell cycle is currently one of the most studied biological processes, due to its significance in growth and development, but increasingly due to its deregulation in many human disorders. Studies using a diverse set of model organisms have greatly expanded knowledge of cell cycle regulation and have contributed to a uniform view of how the basic cell cycle machinery is regulated [28].

Although studied for long, only recently have histone modifications been shown to contribute to it. This chapter of the review highlights the function of some post-translational histone modifications in cell cycle regulation.

1.2.1 The role of H2Bub in DNA replication

Monoubiquitylation of histone H2B on lysine 123 (H2Bub1) is known to be an important histone modification during transcription. H2Bub1 is promoted by the Bre1 ubiquitin ligase and this happens mainly in coding regions, particularly the ones that are being expressed. It is also a highly dynamic histone modification, with cycles of ubiquitylation and deubiquitylation naturally occurring during the transcriptional process. It has been shown that this oscillation is required for proper transcriptional elongation through the control of the RNA Polymerase II (Pol II) and inhibition of recruitment of the CTK1 kinase, which has a function in later phases of transcription. H2Bub1 also controls the methylation state of both lysine 4 and 79 residues on histone H3 - two other important modifications in the transcriptional and replicative process [26].

The mechanistic similarities between transcription and replication suggest that H2Bub1 may have a role in replication as well. In both processes, nucleosomes must be displaced ahead of the replication fork and assembled again after (in replication in more strands than in the original template).

In recent studies performed in the budding yeast [29], H2Bub1 has been shown to be enriched at replication origins, while the writer of this modification, Bre1, is also present. H2Bub1 continued to be present after replication, during G2 and M phases, suggesting that it is maintained on daughter strands of chromatin and after origins are fired for replication start.

Hydroxyurea, an inhibitor of a ribonucleotide reductase enzyme required for deoxyribonucleotides production, has the effect of causing S phase arrest due to nucleotide depletion. When G1 arrested synchronized cells were released into hydroxyurea, Bre1 and H2Bub1 were detected at regions away from replication origins, posing the hypothesis that Bre1 could be travelling with the replisome as happens in transcription due to Pol II complex association [29].

Using a mutant strain for H2BK123, it was possible to detect that H2Bub1 was shown not to be required for the initiation of DNA synthesis as mutant cells initiated DNA replication.

1.2. Histone modifications in cell cycle regulation

Nevertheless, when assessed for their capability to complete replication in the same time as wild-type cells, H2BK123 mutants were slower, as assessed by the levels of H3K56ac and the G2/M marker Clb2. This confirmed the involvement of H2Bub1 in normal cell cycle progression, particularly in the elongation of DNA replication in S phase [29].

Dissecting the reason of the inefficiency of H2BK123 mutant cells in replication elongation, it was shown that in these cell only a fraction of the factors required for replication elongation (Pol α , Pol ϵ Mcm4, Cdc45 Psf2) did not accumulate significantly at positions downstream of replication origins when compared with wild-type cells. Additionally, reduced intermediate single-stranded DNA was detected in mutants. Together this reveals the important role of H2Bub1 in the progression of replication forks, possibly through the interaction with other effectors necessary for replication elongation [29].

In these mutant cells, it was also noted that global histone H3 levels were reduced in replication origins after replication compared with wild-type cells, pointing to either a defect in nucleosome assembly on newly synthesized strands or reduced stabilization of assembled nucleosomes [29].

Spt16, a subunit of the FACT complex that is responsible for nucleosome displacement ahead of the transcription fork is known to interact with H2Bub1 to restore nucleosome occupancy during elongation of transcription. In replication the same was also detected, with Spt16 present at replication origins and more distal positions with the replication fork advancement. However, in H2BK123 mutant cells, Spt16 was significantly reduced in distal replicating positions, suggesting that H2Bub1 stabilizes Spt15 on replicating chromatin [29].

Taken together, evidence collected in this study [29] showed an unprecedented role for a histone modification in replication. Ubiquitylation of H2BK123 is present in origins of replication, promotes efficient replication elongation and the stability of the replisome as well as nucleosome assembly or stability.

1.2.2 Dot1 and H3K79 methylation in the cell cycle

All known histone lysine methyltransferases contain a SET domain, with the exception of Dot1, which has a conserved catalytic core more similar to arginine methyltransferases. Dot1 is conserved from trypanosomes to humans and catalyses mono-, di- and trimethylation of histone H3K79, a residue located in the nucleosome core [26]. Until now, no protein has been identified that unambiguously binds a specific methylated state of H3K79 or that can remove any methyl group from it [30].

It has been shown that Dot1's mode of action is distributive *in vivo* and not processive as all SET domain-containing methyltransferases [30]. This means Dot1 acts in a 'methylate-and-run' fashion rather than using a multistep processive approach - this makes it the first known distributive protein lysine methyltransferase. This also implies that the initial methylation states of H3K79 are obligatory intermediates in the synthesis of higher methylation states by Dot1.

The earliest known function of H3K79 methylation (H3K79me) was in euchromatin, where it restricts Sir-mediated silencing of chromatin to regions of silent chromatin [30]. In this function there is no correlation between the level of any of the specific methylation states and the degree of silencing, but the sum of all methylation states is indeed correlated, which implies that the different methylation states are likely to have redundant functions.

It is also known that di- and trimethylation of H3K79 in yeast requires ubiquitination of histone H2B on lysine 123 (H2BK123ub1) which is deposited by the Bre1 ubiquitin ligase.

Bre1 does not specifically regulate H3K79me but instead enhances its overall catalysis [26] [30]. This has been proposed to be due to the possibility that the ubiquitin moiety directly affects the active site of Dot1 or that H2BK123ub1 affects the *in vivo* chromatin substrate such that H3K79 on the nucleosome core becomes more accessible for interaction with Dot1 [30]. The second option seems more supported due to the fact that *in vitro*, H2BK123ub1 is not required for multiple methylation of H3K79 by Dot1.

Dot1's distributive mode of action has prompted the mathematical modelling of H3K79me states in yeast, using quantitative proteomics data, growth rates and estimates of nuclear Dot1 and histone abundance *in vivo* [31]. This model predicts that increased states of H3K79 methylation accumulate with time until a H3K79me3-rich steady state is reached. Another important prediction is that each subsequent methylation reaction is slower than the one before - in agreement with a non-processive mechanism of methylation [31].

Since all methylation states of H3K79 accumulate during a cell's life cycle, one prediction of the model is that cell-cycle length can affect the average pattern of methylation throughout the cell cycle phases. This was confirmed experimentally with the observation that in slowly growing or arrested cells, H3K79me is indeed higher, probably due to the fact that Dot1 has more time to introduce methyl groups. Over all cell cycle stages, it was noted a temporary drop in methylation during S phase, when new, unmodified histones are deposited on chromatin [31]. The level of H3K79me0 was highest in the S phase but progressive accumulation of methyl groups until a new S phase. This would implicate that H3K79 methylation is at least not fully maintained on new histones during replication. Nevertheless, only when the cell cycle time was doubled, a steady state of high H3K79me3 was reached, indicating that in a normal yeast cell cycle a steady state is not reached and histone renewal might play a role on the balance of H3K79me [31].

The model thus accurately predicted the dynamics of H3K79me through cell cycle and highlighted the importance of the residence time of histones in chromatin. It was shown experimentally that the level of H3K79me correlates with the age of the histones, regardless of the cell's age and that nucleosomes binding genes known for high histone turnover had significantly less H3K79me. These observations confirm the temporal dependency between H3K79 methylation and the age of histones. The same dependency was not detected in H3K4 methylation, suggesting that the presence of a demethylase can counteract the accumulation of methylation on ageing histones [31].

This temporal dependence of H3K79me has additional implications: epigenetic inheritance of H3K79me states is hindered by the fact that at least one of the daughter cells will not receive the same information as the parent cell. This is incompatible with the current model of histone modification as marks of epigenetic memory. However it shouldn't be excluded that not the information of single nucleosomes, but the global state of H3K79me over larger chromatin domains might bare epigenetic function.

A schema of the cyclic model of H3K79me through cell cycle progression can be seen in Figure 1.7.

In *Trypanosoma* - a unicellular eukaryotic parasite with considerable divergence of key cell cycle regulators - the view of H3K79 methylation (H3K76 in *Trypanosoma*) as a carrier of epigenetic information is also challenged by the depletion of H3K79me on newly incorporated histones during S phase [32].

In this organism there are two Dot1 homologue proteins of Dot1 (Dot1A and Dot1B) and there seems to be some division of labour between the two regarding the different methylation states of H3K79. Although both enzymes can mono- and dimethylate H3K79, Dot1A seems

1.2. Histone modifications in cell cycle regulation

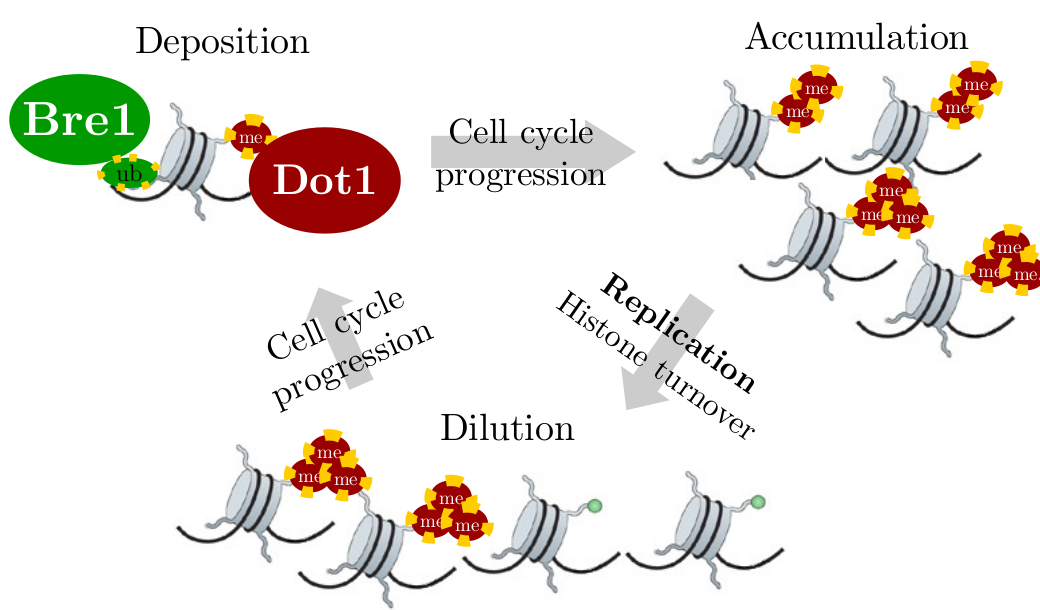


Figure 1.7: Schema of the dynamics of H3K79 methylation during a cell's cycle.

incapable of trimethylating H3K79, whereas Dot1B does, but it is not clear if both enzymes have the same affinity for the two shared methylation forms and therefore contribute equally [32]. Only Dot1A is required for survival.

Depletion of Dot1A decreased H3K79 mono- and di-methylation and generated cells with reduced DNA content, which suggests a role for H3K79me in accurate cell-cycle progression. These cells suffered complete replication inhibition which nevertheless didn't prevent cell division [32].

The inverse disturbance in Dot1A expression (overexpression) resulted in premature H3K79me (during S phase) opposed to generalised increase of methylation levels. Nevertheless, this temporal deregulation caused continuous DNA replication which created cells with increased levels DNA content - the opposite effect of Dot1A depletion. This suggests that not only the global methylation states of H3K79 are implied in accurate cell cycle regulation, but that the timing of addition of this histone modification also plays a role at least in *Trypanosoma* [32].

In human cells, depletion of Dot1, and consequently H3K79me, did not seem to affect replication mechanism itself since the frequency of replication initiation events, replication fork velocity and the proportion of cell cycle phases were not different when compared with normal cells [33]. However, an increased fraction of cells exhibiting DNA content greater than 4N and apoptosis were detected.

Although cells were still able to proliferate and replicate DNA in the absence of H3K79me, its absence affected the regulatory processes modulating the timing of DNA replication, for the cause of a higher fraction of cells with higher ploidy than normally were identified as cells skipping mitosis after S phase and re-replicating their DNA without completing cell division [33].

This deregulation of replication timing led to the hypothesis that Dot1 mediated H3K79me marks origins of replication that have started replication, preventing initiation a second time before the cell completes division. In the case of cells lacking Dot1, the cause of re-replication

might then be new replication from origins or replication that had already been used previously in the same cell cycle. This hypothesis is supported by evidence that H3K79me2 was shown to be associated with functional replication origins, but not with a mutant replication origin that resembled a functional one [33].

These studies highlight the non-canonical use of a histone modification (H3K79 methylation) and show its importance in the regulation of replication, a major cellular event.

1.3. *Oikopleura dioica*, a marine chordate extensively employing endocycles

1.3 *Oikopleura dioica*, a marine chordate extensively employing endocycles

Oikopleura dioica is a marine chordate organism belonging to the class Appendicularia, which is a member of the Tunicate subphylum along with the classes Thaliacea (salps) and Ascidiacea (sea squirts). Tunicates are the most closely extant group related to the vertebrates (see Figure 1.8) [34]. *Oikopleura* shares some biological traits with most tunicates (e.g. filter feeding through a secreted house), but unlike them has some peculiarities which make it a very interesting model for the study of many biological features, such as an accelerated development and life cycle, a miniature genome and an extensive use of endocycles through its life. *Oikopleura dioica* owes its name to the fact that it is the only dioecious appendicularian known, being most other species hermaphrodites.

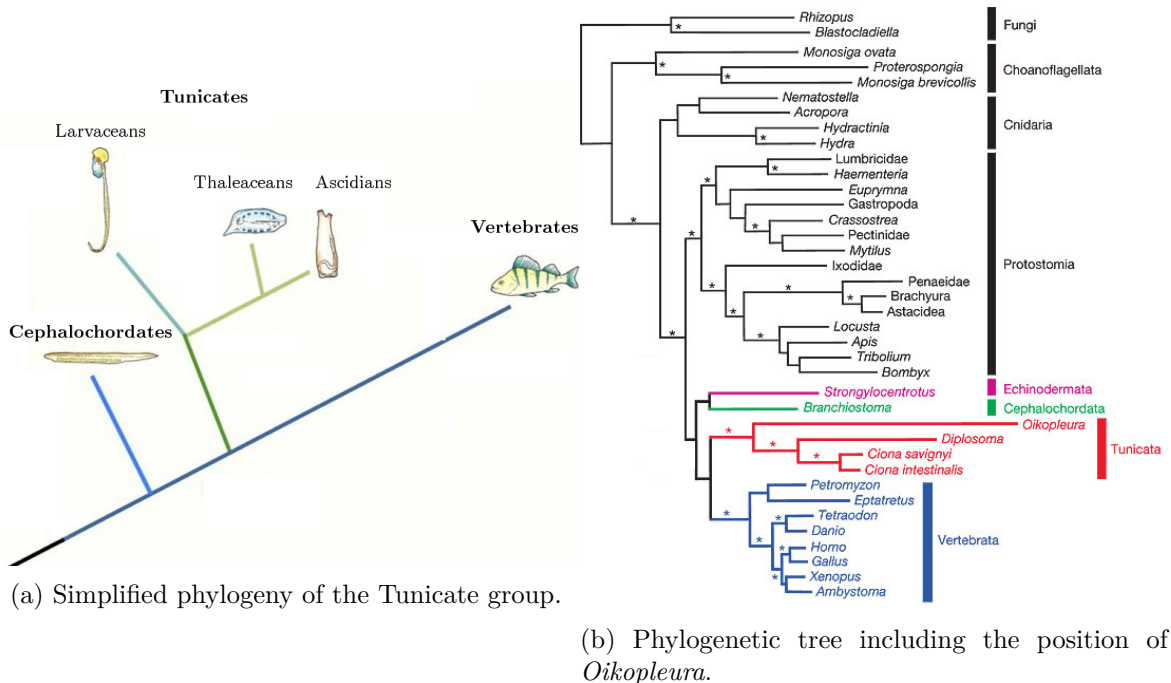


Figure 1.8: Phylogenetic position of the Tunicate group and *Oikopleura dioica*. Adapted from a figure by Charlotte Konikoff at the University of Washington and [34].

1.3.1 *Oikopleura*'s life cycle

Oikopleura reproduces through external fertilization after the rupture of both female and male gonads or sperm release via the spermiduct [35]. Remarkably, the first cell division occurs only 35 min after fertilization proceeds with a basically bilateral pattern with minor left-right asymmetries. Gastrulation takes place starting at the 32-cell stage (roughly two hpf at 20°C) and progresses until all of the vegetal cells have ingressed and are covered by the animal cells. By the 64-cell stage eight cells disposed in two rows of four align themselves along the anterior-posterior axis and are internalized giving rise to the neural tube later. This simple but accelerated early development of *Oikopleura*, gives rise to a conserved chordate

body plan with the animal bent ventrally inside the chorion (Tailbud stage) soon hatching and becoming a free-swimming larva [36].

Larval development takes place until fifteen hours after fertilization, where a metamorphosis known as Tailshift occurs: the tails changes orientation towards the ventral side of the animal. At this point, most cells have stopped mitotic division and start performing endocycling, a endoreduplicative cell division strategy which gives rise to multinucleated cells (> 1000 C [37]) and increases body size by increasing the cell volume rather than cell division (discussed in detail in Section 1.1.2 and 1.3.4). *Oikopleura* secretes an extracellular, cellulose-based structure through its epithelium ('tunics' or 'houses') which uses to filter particles in suspension to feed on and sustain growth. This growth strategy continues throughout most of the remaining juvenile life of *Oikopleura* until the third day of development, where gonad development starts.

Most of *Oikopleura*'s growth until the end of its life at day six is dedicated to gonad maturation, which at its peak reaches a total volume bigger than the somatic part of the whole organism. This is achieved again through the employment of endocycling in the gonad tissues, creating a giant cell with over 10000 nuclei. In the female gonad, a process of nuclei selection occurs, where some nuclei are selected to become oocytes and other nurse nuclei, providing support for the selected nuclei and then degenerating towards the end of maturation. Male gametogenesis is known to start later than the female and is a fast process occurring just before full maturation.

An overview of the different phases of development of *Oikopleura* can be seen in Figure 1.9.

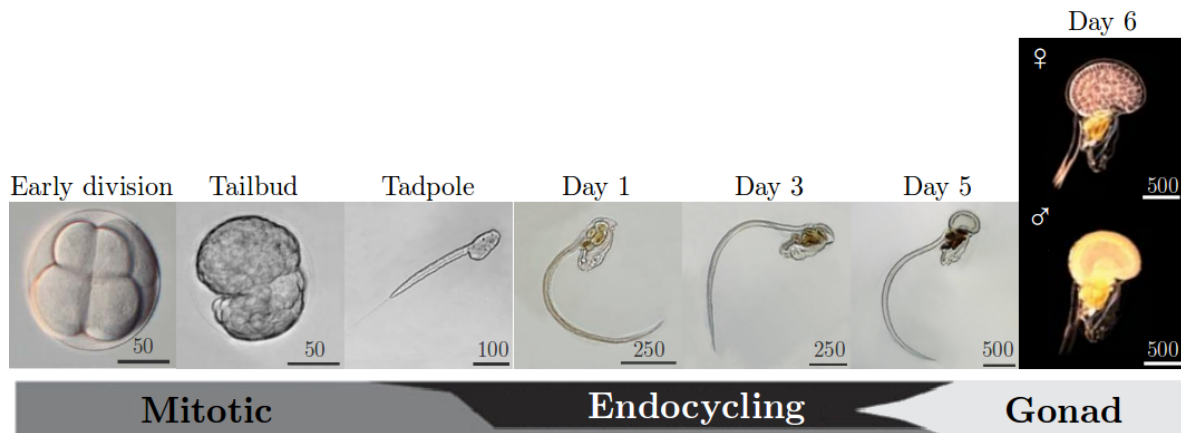


Figure 1.9: Stages of development of *Oikopleura dioica* and respective predominantly employed modes of cell cycle. Scale in micrometers. Adapted from [36] and [35].

1.3.2 The genome of *Oikopleura dioica*

A reference sequence of *Oikopleura dioica* genome was made available in 2010 [38], revealing a surprisingly plastic architecture and the species fast evolution. In this planctonic animal, the chordate genome architecture seems to have been redesigned to obtain a minimal-sized genome: approximately 70 megabases. This is greatly due to the reduction of intron and intergenic space, although some gene loss was also detected.

1.3. *Oikopleura dioica*, a marine chordate extensively employing endocycles

The dramatic reduction of intergenic space brings obvious constraints in terms of gene regulation, which is to some degree easily seen by the normal occurrence of genes in operons (28% of the 18000 total), and makes the *Oikopleura* genome one of the most compact of all known chordate genomes. Operons are significantly enriched for genes involved in house-keeping functions, while genes involved in developmental processes are significantly under-represented in Operons [38].

Although the idea of generalised simplicity in regulatory mechanisms could be a hypothesis due to the high density of the genome, insights from two distinct populations of *Oikopleura* seem to contradict this. Highly conserved elements in non-coding regions have been found through genomic alignments of Atlantic and Pacific *Oikopleura dioica* [38]. Besides those, several other conserved genomic features can be found in the *Oikopleura* genome. Long introns are more likely to be descendants of ancestrally conserved introns than the shorter ones, and genes important for development show twice as much more ancestral introns when compared to all annotated genes. Nevertheless, major changes in the number and synteny of cis regulatory modules (promoters and enhancers) compared with other Tunicates are likely to have occurred due to major loss of gene synteny and the fast evolutionary rate of *Oikopleura*.

1.3.3 Histone variants and modifications of *Oikopleura dioica*

In *Oikopleura dioica*, the diversity of histone variants and their use through development has been profiled with a mass spectrometry approach [39]. *Oikopleura* possesses a histone complement of 31 histone variants which can be grouped in distinct sets of developmental expression profiles throughout its life cycle. Most common histone variants are present, although remarkable absences can be seen in the macroH2A and H2AX histone variants. A significant expansion of the histone complement exists in *Oikopleura*, which is even further diversified by the existence of alternative splicing on some variants. Additionally, a remarkable set of 15 male-specific histone variants was discovered [39].

Histone modifications also occur extensively in *Oikopleura* [39]. Mass spectrometry identified 40 different post-translational modifications (PTMs) in structurally distinct locations. Histone modifications involved in transcription and marking of cis-regulatory elements such as mono-, di- and tri-methylation of lysine 4 in histone 3 (H3K4me1/2/3), acetylated lysine 27 on histone 3 (H3K27ac), are present. Repressive marks with different modes of action were also found such as the constitutive heterochromatin marker trimethylation of lysine 9 of histone H3 (H3K9me3) and the mark laid by the Polycomb-Repressor Group (PRG) of proteins trimethylation of lysine 27 on histone H3 (H3K27me3). Histone marks related with the control of replication such as all methylation states of lysine 20 in histone H4 (H4K20me), monoubiquitination of lysine 123 in histone H2B (H2BK123ub1) and methylation of lysine 79 in histone H3 (H3K79me), the later two of particular interest in this work.

Regarding the function of histone modifications in the cell cycle, conservation of the function of phosphorylation of serine 10 and 31 in histone 3 (H3S10p and H3S31p) was detected [40] and [41]. These modifications are a marker for the mitotic phase of cell cycle, present in the diplotene/diakinesis where they have been shown to contribute to normal chromosome segregation and transmission during mitosis and meiosis. H3S10p is also a mediator of gene silencing at different stages of the cell cycle. Phosphorylation of serine 28 of H3 (H3S28p) also occurs during mitosis where it is seen as a licensing factor and indicating the state of readiness of chromosomes to undergo separation after metaphase rather than as a driving force in chromosome condensation.

1.3.4 The cell cycle modes of *Oikopleura dioica*

Oikopleura dioica employs an unique combination of different modes of cell cycle through its life cycle (see section 1.3.1). This is unique because most cells switch from rapid mitotic division to endocycles around the event of tail shift. Nevertheless, not much is known about the regulation of the alternative cell cycle modes of endocycling in the somatic and gonadal tissues and even less on how can cells change from one to the other.

In a study of the diversity of *Oikopleura*'s Cyclin and Cyclin-dependent kinases (CDK) [42], it was shown that the complement of cyclins involved in transcriptional regulation is similar to other invertebrates, whereas the complement used for cell cycle regulation has significant amplifications, especially in the cyclin D, B and CDK1 families. Remarkably, some families are extended not only by gene duplication but by the existence of alternative splice forms, such as in the Cyclin Db. Many of the Cyclin and CDK proteins are expressed exclusively and distinctively during the later stages of development that follow sex differentiation and gonad maturation, showing an important link between complex biological features and gene regulation.

Amplification of the CDK1 family to five paralogs of this mitotic CDK is somewhat unexpected due to the necessity of suppression of CDK1 activity by the activation of APC/C (Cdh1/Fzr) in *Drosophila* and mammalian endocycles. Due to the extensive use of endocycles, amplification of the families of CDK regulating the G1/S phases (CDK 2 or 6) would have been more anticipated. Also very suspicious is the expansion of the Cyclin D family, leading to the hypothesis that regulation through CyclinD-CDK1 complexes especially during the late developmental stages could provide the regulatory refinement needed to control such a complex system as the female coenocyst [42].

Nevertheless, the *Oikopleura* somatic endocycle resembled features known to be present in endocycles in other systems such as the downregulation of cyclins B and A as in the *Drosophila* and vertebrate mammalian trophoblast giant cells endocycles [42].

Regarding other key cell cycle regulators, *Oikopleura* possesses a fairly conserved set of proteins involved in regulation of cell cycle both upstream and downstream of the Cyclin-CDK effectors. Of particular notice, the E2F family of transcription factors is reduced when compared with vertebrates or other invertebrates such as *Drosophila*. In *Oikopleura* there is only a canonical activator (E2F1) and an atypical repressor (E2F7) (see more about the E2F family of transcription factors in section 1.1.2.2.2).

1.3.5 Growth arrest

Developmental processes such as reproduction can be regulated by environmental conditions. In an array of organisms, the trigger for maturation of reproductive cells can be delayed or paused in unfavourable conditions. Two particular examples can be seen in the nematode *Caenorhabditis elegans* and the fruitfly *Drosophila melanogaster*, where the loss of germline stem cells (GSCs) has the consequence of extending the lifespan of the organism by evolutionarily conserved signalling pathways [43]. The reverse situation has also been detected, when under nutritional deprivation, these models sacrifice a portion of GSCs that have already entered meiosis, maintaining only a small pool of active GSCs [44].

In *Oikopleura dioica*, nutritional stress caused by cultivation in high density has shown to induce a condition of growth arrest (GA) where somatic endocycling ceases and the lifespan of the animal is expanded at least three-fold. In contrast with the previously mentioned models,

1.3. *Oikopleura dioica*, a marine chordate extensively employing endocycles

Oikopleura only enters GA condition before the start of meiosis. If nutritional deprivation is encountered after, GA is not induced and maturation occurs at the same time, but animals have a reduced offspring [44].

The GA condition therefore represents a state of suppression of proliferation in the somatic endocycles and could be exploited for experimental purposes in understanding endocycle regulation.

1.4 Project Goals

The general aim of this Master thesis was to study the regulation of the different *Oikopleura* cell cycle modes and the transition between them. Although much is known about the canonical eukaryotic cell cycle and to some extent about endocycles, knowledge on the mechanism cells use to transition between different modes of cell cycles is still quite reduced. To broaden knowledge on this topic, this thesis studies specifically the role of the E2F transcription factors in these cycles by identifying putatively regulated genes during developmental stages with predominant use of either the canonical mitotic cycle or endocycling.

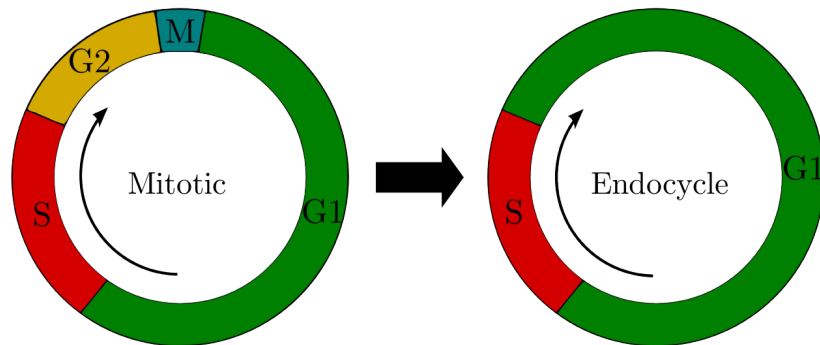


Figure 1.10: Goal: to gain understanding of the mechanisms behind the change of cell cycle modes in *Oikopleura*.

To this end, ChIP-seq on the E2F1 and E2F7 transcription factors was performed in the Tailbud (mitotic), Day 2 (somatic endocycle), Day 6 female (gonadal endocycle) and Day 6 male (mitotic) developmental stages. ChIP-chip data on the last two stages had been previously acquired and it is here explored to highlight the differences on these two modes of cell cycle.

A much less explored component of regulation of cell cycles is the influence histone modifications have. Methylation of lysine 79 on histone H3 (H3K79me) has been proposed to act as a timer for a cell's age due to the distributive mechanism of action of its histone methyltransferase - Dot1 - and the absence of a demethylase, which causes the accumulation of this histone mark along the cell's life .

Another major goal of the project was to gain insights on the regulation of the different cell cycle modes by performing functional studies on the Dot1 histone methyltransferase, and consequently in the function of H3K79me during different modes of cell cycle in *Oikopleura dioica*. This was accomplished by the use of a chemical inhibitor of Dot1 in the Tailbud stage (mitotic) and the Day 5 (endocycle) and characterization of the phenotypic effects of inhibition from a developmental and cellular point of view.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Antibodies

Table 2.1: Antibodies used for ChIP, immunohistochemistry and western blot

Antibody	Description	Supplier	Purpose
OdE2F1	Anti-E2F1 of <i>Oikopleura dioica</i>	21stCentury	ChIP
OdE2F7	Anti-E2F7 of <i>Oikopleura dioica</i>	21stCentury	ChIP
ab2886	Anti-Histone H3 K79me1 Rabbit	Abcam	Western blot
39143	Anti-Histone H3 K79me2 Rabbit	Active Motif	Western blot
ab2621	Anti-Histone H3 K79me3 Rabbit	Abcam	Western blot
05-1312	Anti-Ubiquityl-Histone H2B Mouse	Milipore	Western blot
ab32107	Anti-Histone H3 (phospho S10) [HTA28]	Abcam	Immunohistochemistry and Western blot
ab43176	Anti-ATP Synthase beta Chicken	Abcam	Western blot
A-11077	Anti-Rat IgG secondary antibody conjugated with Alexa 568 fluorochrome	Molecular Probes	Immunohistochemistry
ab6802	Donkey anti-Rabbit IgG secondary antibody conjugated with HRP	Abcam	Western blot
ab7003	Donkey anti-Mouse IgG secondary antibody conjugated with HRP	Abcam	Western blot
151-24-06	Goat-Anti-Chicken IgG secondary antibody, Phosphatase labeled	KPL	Western blot

2.1.2 Chemicals and reagents

Table 2.2: Chemicals and reagents of particular relevance

Supplier	Chemical	Purpose
Thermo Scientific	16% Formaldehyde ampoules, methanol-free	ChIP
Tocris Bioscience	SGC0946	Dot1 inhibition

2.1.3 Consumables

Table 2.3: Consumables of particular relevance

Supplier	Consumable	Purpose
Sigma-Aldrich	BD Precisionglide syringe needles - gauge 27, L 1/2 inches	ChIP
Covaris	AFA microtubes	ChIP
Sigma-Aldrich	Siliconized microtubes 1.7 mL capacity	ChIP
Invitrogen	Protein G magnetic beads	ChIP
Life Technologies	Qubit dsDNA HS assay kit	ChIP
Bio-Rad	Mini-PROTEAN TGX stain-free gel	Western blot
Bio-Rad	Trans-Blot turbo mini nitrocellulose membrane	Western blot
Bio-Rad	Precision Protein StrepTactin-HRP Conjugate	Western blot
Bio-Rad	Clarity western ECL substrate	Western blot
Molecular Probes	TO-PRO-3 Iodide stain	Immunohistochemistry
Roche	In Situ Cell Death Detection Kit, Fluorescein	Tunel assay
Life Technologies	Click-iT EdU Alexa Fluor 488 Imaging Kit	EdU detection

2.1.4 Instruments and equipment

Table 2.4: Instruments of particular relevance

Supplier	Instrument/equipment	Purpose
Covaris	S200 focused sonicator	ChIP
Thermo Scientific	Nanodrop ND-1000 spectrophotometer	ChIP
Life Technologies	Qubit 2.0 fluorometer	ChIP
BioRad	C1000 thermocycler with CFX96 module	ChIP
Leica	TCS-SP5 confocal microscope with Ar-Kr ion laser	Microscopy
Bio-Rad	Trans-Blot turbo transfer system	Western blot
Bio-Rad	ChemiDox XRS imaging system	Western blot

2.1. Materials

2.1.5 Buffers and solutions

2.1.5.1 ChIP

PBS, pH 7.4 1.37 M NaCl, 27mM KCl,100 mM Na₂HPO₄, 1.8 mM KH₂PO₄

Lysis buffer 150mM NaCl, 1% NP-40, 0.5% Na deoxycholate, 0.1% SDS, 50 mM Tris pH 8, 1mM EDTA

PBS-T 0.02% Tween-20 in PBS

Lithium chloride buffer 50 mM Tris pH 8, 250mM LiCl, 0.5% NP-40, 0.5% Na deoxycholate

TE buffer 10 mM Tris pH 8, 1 mM EDTA

Elution buffer 50 mM Tris pH 8, 1 mM EDTA, 0.1% SDS

2.1.5.2 Western blot

Laemmli extract 200 mM Tris-HCl pH 6.8, 8% SDS, 40% glycerol, 0.004% bromophenol blue (400 mM 2-mercaptoethanol - added fresh)

Polyacrilamide gel, stacking portion 6% bis-acrilamide, 125 mM Tris pH 6.8, 0.1 %SDS, 0.1% ammonium persulphate, 0.1% TEMED

Polyacrilamide gel, running portion 8-18% bis-acrilamide, 370 mM Tris pH 8.8, 0.1 %SDS, 0.1% ammonium persulphate, 0.1% TEMED

SDS-PAGE running buffer 25 mM Tris, 192 mM glycine, 0.1% SDS

TBS, pH 7.4 1.5M NaCl, 0.2M Tris

TBS-T 1.5M NaCl, 0.2M Tris, 0.1% Tween-20

PBS, pH 7.4 1.37 M NaCl, 27mM KCl,100 mM Na₂HPO₄, 1.8 mM KH₂PO₄

Transfer buffer 25 mM Tris, 192 mM glycine (10% methanol - added fresh)

Ponceau S stain 2% Ponceau S, 30% trichloroacetic acid, 30% sulfosalicylic acid

2.1.5.3 Immunohistochemistry, Tunel assay and Edu incorporation and detection

Fixative 4% paraformaldehyde, 100 mM MOPS pH 7.5, 500 mM NaCl

PBS, pH 7.4 1.37 M NaCl, 27mM KCl,100 mM Na₂HPO₄, 1.8 mM KH₂PO₄

PBS-T 0.02% Tween-20 in PBS buffer

PBS-TE 0.02% Tween-20, 1 mM EDTA in PBS buffer

PBS-TEG 0.02% Tween-20, 1 mM EDTA, 100 mM glycine in PBS buffer

Blocking solution 3 % acetylated bovine serum albumin (BSA) in PBS-TE buffer

2.1.5.4 Tunel assay

Permeabilization solution 0.1% Triton X-100 in 0.1% sodium citrate, freshly prepared

2.2 Animal culture and collection

2.2.1 Culture of *O. dioica*

The culture of *Oikopleura* was performed at the SARS Centre appendicularian facility as previously described [35]. Cultured animals are native from the coastal area outside Bergen and are frequently collected and added to the permanent culture. Animals are permanently cultured at the facility in 6L seawater beakers, with permanent stirring and daily water renewal as well as feeding twice daily with algae according to the developmental stage (see table 2.5). The use of a fixed volume for culture implies the progressive dilution of animals until the third day of life, where density remains at 150 animals per 6L beaker.

Table 2.5: Feeding regime of *Oikopleura* according to developmental stage

Developmental stage		<i>Isochrysis</i> <i>sp.</i>	<i>Chaetoceros</i> <i>calcitrans</i> (cells/mL)	<i>Rhinomonas</i> <i>reticulata</i>	<i>Synechococcus</i> <i>sp.</i> (mL)	Crushed <i>R.</i> <i>reticulata</i>
Day 1	Morning	2000	2000	0	5	5
	Evening	1000	1000	0	3	5
Day 2	Morning	2000	2000	0	5	5
	Evening	1000	1000	0	3	5
Day 3	Morning	2000	4000	0	5	5
	Evening	1000	2000	1000	3	5
Day 4	Morning	4000	4000	2000	0	5
	Evening	2000	2000	1000	0	5
Day 5	Morning	4000	4000	2000	0	5
	Evening	2000	2000	1000	0	5

2.2.2 Collection of *O. dioica*

2.2.2.1 Tailbud stage

To collect *Oikopleura* in the Tailbud developmental stage, a controlled *in vitro* fertilization was performed. Day six mature male animals were collected to a Petri dish with seawater and allowed to spawn. When all male animals had spawned, sperm quality was visually inspected on a light microscope for motility. Day six mature females were individually collected to glass salliars with seawater and allowed to spawn at 18°C. When spawned, 60µL of sperm solution was added to the salliar and after 3-4 hours tailbud animals collected to a eppendorf microtube.

2.2.2.2 Day two stage

Late day two *Oikopleura* were individually collected with the aid of a 2 mL plastic pipette to a 1 L beaker with clean seawater (no algae) and stayed there 3 to 4 hours to be allowed to empty the gut of any remaining algae as well as build a new clean house. To achieve release

2.3. ChIP-seq

of the houses, animals were individually collected again, this time using a mouth-pipette built from a 1 mL plastic pipette and poured into a glass recipient on ice with 0.125 mg/mL MS222 and allowed to sink to the bottom, where a third collection using a 200 μ L micropipette took them to a 1.5 mL microtube.

2.2.2.3 Day six, immature stage

Immature day six animals with a visible gonad but naked-eye indistinguishable features were collected from culture into a 500 mL plastic beaker with clean seawater with the aid of a 25 mL plastic pipette and from there to a 1.5 mL microtube with a 2 mL plastic pipette.

2.3 ChIP-seq

Preparation of *Oikopleura* chromatin was performed as previously described [45] with minor modifications.

2.3.1 Animal fixation

Tailbud and day two animals were briefly spun at 5000 g for 10 seconds to be collected at the bottom of the microtube, while day six immature voluntarily sank in seconds time. Seawater was exchanged for 0,5 mL PBS and 0,5 ml of either 1 or 2% Formaldehyde in PBS was added to have a final fixative concentration of 0.5 or 1% respectively, and let rotating for a variable amount of time as shown in Table 2.6, depending of the developmental stage.

Formaldehyde was quenched with glycine solution to stop fixation and again let rotating for 5 min at 18°C. Again by either spinning at 5000 g for 30 seconds or allowing animals to freely sink, Formaldehyde was removed and animals washed with cold PBS with protease inhibitors and PMSF three times on ice, after which all solution was removed and animal pellets frozen with liquid nitrogen and stored at -80°C for future use.

Table 2.6: **Fixation conditions for each tested developmental stage of *Oikopleura***

Stage	Fixative concentration	Time (min)	Temperature (°C)
Tailbud	1%	5	18
Day two	0.5%	5	18
Day six immature	1%	10	18

2.3.2 Cell lysis and chromatin sonication

Frozen animals pellets were thawed on ice and pooled with the aid of cold Lysis buffer to a total volume multiple of 130 μ L but never less than 390 μ L depending on the total amount of animal material. Mechanical lysis was conducted with a 27-gauge needle and 2 mL syringe passing the whole volume no less than three times through the needle and incubated on ice 15 min.

Sonication was performed on a S200 Covaris focused sonicator on a 4°C water bath with 130 μ L AFA glass microtubes. The instrument settings used to sonicate the chromatin to a

approximate gaussian distribution within 100-800 bp can be seen on Table 2.7.

Whole sonicated lysates were centrifuged for 10 min, at 21000 g at 4°C and the chromatin-enriched supernatant was collect to a siliconized microtube.

Table 2.7: **Settings used on the S200 Covaris sonicator for *Oikopleura* chromatin**

Setting	Value
Duty cycle	5%
Intensity	7.5
Processing time	5 minutes
Bath temperature	5-6 °C
Power mode	Frequency sweeping
Degassing mode	Continuous
Volume	130 μ L
AFA intensifier	Integrated
Water level	8

2.3.3 Chromatin quality assessment

Total protein yield of chromatin was quantified with the Bradford assay with a 30:1 ratio of Coomassie reagent to chromatin and measured on a Nanodrop instrument.

To check the distribution of the sonicated chromatin fragments, 1% SDS, 100mM NaCl and 50 μ g/mL RNase A were added to a fraction of the chromatin and incubated for 30 min at 55°C. Proteinase K was added to 200 μ g/mL and samples incubated 90 min to O/N at 65°C to reverse crosslinks. DNA was purified with the Phenol-Chloroform method and precipitation with ethanol. Briefly, an equal volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was mixed to the samples, and after centrifugation the 30 μ g Glycogen, 300 mM Sodium Acetate and cold 70% Ethanol (final concentrations) were added to the aqueous phase in a new microtube and incubated at -80°C for 1 hour. After centrifugation at 4°C, DNA pellets were washed with 70°C Ethanol and dissolved in 3 to 5 μ L TE buffer.

DNA quality and amount was quantified on a Nanodrop instrument and loaded on an Agilent DNA High Sensitivity digital electrophoresis chip, to assess the distribution of DNA fragments with an Agilent Bioanalyzer instrument.

2.3.4 Immunoprecipitation

Protein G magnetic beads were washed once with PBS-T in siliconized tubes, incubated at 4°C rotating for two hours with 10 μ g of antibody and washed again with PBS-T and Lysis buffer.

800 μ g of chromatin were added to the magnetic beads with the already bound antibody and left O/N rotating at 4°C, after which a series of 15 min, 4°C washes followed: three with Lysis buffer without any inhibitors; two with Lithium chloride buffer; one with TE buffer.

Elution of IP-enriched DNA followed with the addition of elution buffer and incubation at 65°C for 15 min horizontally stirring at 900 rpm. A second, more stringent elution was

2.4. Western blot

performed with TE buffer with 500 mM NaCl and both supernatants were pooled.

Reversal of crosslink took place during a three to five hour incubation at 65°C, after which 50 $\mu\text{g}/\text{mL}$ RNase A were added, incubated at R/T for 10 min, and incubated again this time with 200 $\mu\text{g}/\text{mL}$ Proteinase K at 55°C O/N. Phenol-Chloroform extraction followed as previously described on section 2.3.3, with the exception of final pellet dilution in 50 μL TE buffer.

1 μL of sample was used in a Qubit high-sensitivity assay to measure total dsDNA yield from the ChIP assay.

2.3.5 Illumina library construction and high-throughput sequencing

ChIP libraries were made using the Ovation Ultralow Library kit according to the manufacturer. Selected DNA fragments were within 150-350 bp and each set of biological replicates (16 samples) was sequenced in a single lane of a Illumina HiSeq 2500 instrument with 100 bp paired-end reads.

2.4 Western blot

2.4.1 Sample preparation

Animals of the required stage were collected in microtubes and washed once with PBS. Laemmli buffer with fresh 2-mercaptoethanol was added to a final 1x concentration and samples were boiled at 99°C for 10 minutes with mild agitation. After a 5 minute centrifugation at 16000 g, samples were snap frozen in liquid nitrogen and stored at -20°C for later use.

2.4.2 SDS-PAGE

SDS-polyacrylamide gel electrophoresis was performed using either precasted or self-casted gels (see section 2.1.3 or 2.1.5.2 respectively). The total volume of samples, as well as 8 μL of protein marker were loaded and gels were run for 20 minutes at 50 V and then approximately 90 minutes at 150 V in running buffer.

2.4.3 Protein transference

Electrophoretically separated proteins were transferred to a nitrocellulose membrane using a semi-dry transfer system with a program optimized for proteins with mixed molecular weight (7 minutes transfer at constant 1.3 A).

The nitrocellulose membrane was washed once in TBS buffer for 10 minutes, incubated in 1:10 Ponceau stain in TBS for 5 minutes and briefly rinsed with distilled water to allow visualization of protein, and an assessment of the transference efficiency. Ponceau stain was removed with two TBS-T washes.

2.4.4 Blotting

To reduce unspecific antibody binding, the membrane was blocked with 5% fat-free milk or 3% BSA in TBS-T for 60 minutes at R/T with mild horizontal shaking.

The membrane was incubated with primary antibody diluted 1:500-2000 in 5% fat-free milk O/N at 4°C, and washed for 10 minutes at R/T with mild horizontally shaking, once with TBS, twice with TBS-T and again with TBS.

Secondary antibody diluted 1:2000-5000 in 5% fat-free milk was added to the membrane and incubated for one hour at R/T together with 1:5000 StrepTactin-HRP conjugate. Following washes were as after primary antibody incubation: 10 minutes at R/T with mild horizontally shaking, once with TBS, twice with TBS-T and again with TBS.

2.4.5 Detection

The membrane was washed once with PBS and incubated for 5 minutes in ECL substrate to allow detection of secondary antibody binding through chemiluminescence, being scanned with an exposure time between 3 and 30 seconds depending on signal strength.

2.5 Whole mount Immunohistochemistry

Collected animals were fixed in 4% paraformaldehyde at 4°C O/N and washed once with PBS-TE and twice with PBS-TEG to quench remaining formaldehyde. Two TBS-TE washes followed, after which animals were inculcated O/N in blocking solution.

Primary antibodies were diluted 1:100 in blocking solution and added to the animals after the previous blocking solution was removed from the microtubes. An incubation of at least six days with the primary antibodies was performed at 4°C

After primary antibody binding, six PBS-TE washes in 10 min intervals followed to remove any unbound primary antibody traces. A post-fixation similar to the primary one followed: 4% formaldehyde in PBS-TE was added, and fixation took place O/N at 4°C, after which one wash with PBS-TE and two with PBS-TEG followed.

Secondary antibodies diluted 1:500 in blocking solution were added to the animals and these were left for another minimum of six days at 4°C in darkness, to allow antibody binding and avoid weakening of fluorescent signal. Another six washes of PBS-TE followed, with the particularity that ToPro3 was added to the fourth wash, diluted 1:1000, therefore allowing DNA staining.

Animals were mounted in glass slides with Vectashield and imaged with confocal microscopy.

2.6 Dot1 inhibition

SGC0946 or DMSO were diluted to 30 μ M in sea water previously heated to 60°C and left at 60°C for 10 minutes, after which it was left to cool to R/T.

Animals of the Tailbud developmental stage were obtained by *in vitro* fertilization as described in section 2.2.2.1. After hatching, animals were pooled and transferred in equal number to a 6-well plate coated with agarose with previously prepared sea water with either SGC0946 or DMSO.

Animals of D5 stage were transferred to a 50 mL Falcon tube filled sea water with SGC0946 or DMSO ddiluted as described previously. Falcon tubes were left rotaing slowly at R/T for 3,5 hours and animals were fed with 4000 cells/mL of *Isochrysis sp.*, *Chaetoceros calcitrans* and *Rhinomonas reticulata*.

2.7. *Tunel assay*

2.7 Tunel assay

Collected animals were fixed in 4% paraformaldehyde for 1 hour at RT and washed twice with PBS-TE.

Permeabilization was performed by the addition of 500 μL of Permeabilisation solution and incubation for 10 min on ice.

Tunel reaction was prepared as indicated by the manufacturer with the exception that higher volume was used. Reaction mixture was prepared by the addition of 50 μL of Enzyme solution to 450 μL of Label solution. At least 100 μL (250 μL recommended) of the Tunel reaction was added to the samples and these were incubated for 60 min at 37°C without light.

Samples were washed five times with PBS-TE with the particularity that ToPRo3 was added to the third wash, diluted 1:1000, to stain DNA.

Animals were mounted in glass slides with Vectashield and imaged with confocal microscopy.

2.8 EdU incorporation and detection

Edu incorporation

Animals of required stage were incubated in either 10 μM EdU solution or DMSO in sea water for 1 hour.

Edu detection

Animals were collected and fixed in 4% paraformaldehyde at 4°C O/N, washed once with PBS-TE and twice with PBS-TEG to quench remaining formaldehyde. Two TBS-TE washes followed, after which animals were washed three times in blocking solution.

The Click-iT reaction cocktail was prepared according to the manufacturer for each sample (Table 2.8), and 500 μL were added to each sample, after which a 2,5 hour incubation time with mild shaking followed.

Table 2.8: **Reaction mix used in detection of EdU**

Reagent	Volume (μL)
10X Click-iT reaction buffer	43
CuSO	20
Alexa Fluor azide	1.2
Reaction buffer additive	50
MQ H ₂ O	386

Samples were washed once with blocking solution and six times with PBS-TE, with the particularity that ToPRo3 was added to the fourth wash, diluted 1:1000, therefore allowing DNA staining.

Animals were mounted in glass slides with Vectashield and imaged with confocal microscopy.

2.9 Data analysis

2.9.1 Data preprocessing

ChIP-chip data was obtained by performing ChIP on E2F1 and E2F7 transcription factors and hybridized to a high density tiling array designed for the *Oikopleura* genome previously used [46]. This data was processed using the the R/Bioconductor package Ringo [47].

Quality of raw microarray data was assessed by observing the spatial distribution of the intensities on the chips and autocorrelation.

Raw probe intensities were normalized and fold changes of reporters' intensities were derived from the enriched sample divided by their intensities in the non-enriched input sample. The logarithm of these ratios is used as normalized probe intensity values.

Smoothing over individual probe intensities was performed to reduce the bias from probe GC content, melting temperature and secondary structures.

2.9.2 Finding of ChIP-enriched regions

ChIP-enriched regions (ChIP-ERs) were derived from the smoothed intensities of reporters that exceed the 0.95 threshold in a region with at least three contiguous positions but where none of the probes is in a distance higher than 48 bp to the others.

2.9.3 Genomic characterization of ChIP-enriched regions

ChIP-ER overlap with genomic features and distances to nearest transcription start site were calculated using the Bedtools [48] *closest* function with gene annotation information retrieved from *Oikobase* [46].

2.9.4 Transcription factor motif finding in ChIP-enriched regions

DNA sequence was retrieved from ChIP-ER and search of motifs was performed using MEME-ChIP [49] with default settings and the JASPAR core vertebrate database.

2.9.5 Local epigenetic landscape of ChIP-enriched regions

Genome segmentation by chromatin states was derived from profiles of 19 histone modifications and gently provided by P. Navratilova, G. Danks and E. Thompson prior to publication. Annotation of ChIP-ERs with chromatin states was performed using the Bedtools *intersect* function [48].

2.9.6 Assignment of ChIP-enriched regions to genes

Each ChIP-ER was assigned to the closest transcription start site by use of the *closest* function of Bedtools [48] and that gene is annotated as a target. ChIP-ERs lying in genome scaffolds without genes were discarded.

2.9.7 Testing of significant Gene Ontology terms

A hypergeometric tests for overrepresentation of Gene Ontology terms was performed with a 0.001 p-value, and using the universe of all GO term annotation of *Oikopleura* as

2.9. Data analysis

background using the GOstats R/Bioconductor package [50]. The GO tree of significant terms was reduced by reducing redundant GO terms joining terms under the same parent. This was performed with REViGO [51].

2.9.8 Expression levels of genes

Expression values of genes through development of *Oikopleura* were retrieved from *Oikobase* [46]. When comparing two developmental stages, expression values were normalized by the mean expression of all genes within each stage and transformed with a logarithm of base 2. When two developmental stages were compared (ovary and testis tissue), a smoothed conditional mean was fit and quantiles of the distribution were calculated from expression values.

Chapter 3

Results

3.1 E2F factors in *Oikopleura*'s late developmental stages

Oikopleura dioica employs endocycles extensively through its accelerated life cycle. Endocycles are the predominant mode of cell cycle in somatic tissues from the Tailshift developmental stage on, and from the Day 3 of development endocycles are again employed extensively during gonad growth and maturation. During the last day of the life cycle, the final maturation of the gonad occurs. In female animals, the gonad consists of a coenocyst containing thousands of nuclei of different nature - meiotic nuclei, which arose from germ stem cells and will give rise to progeny, and nurse nuclei which sustain the meiotic. During the development of the male gonad, endocycles are employed extensively for proliferation and growth and gametogenesis starts at a later point than in females. The meiotic differentiation resembles more closely the canonical eukaryotic cell cycle than endocycling. We want to discover the regulatory mechanisms behind the use of each type of cell cycle in *Oikopleura* and in the transition between them.

E2F transcription factors (TFs) are known to play a major role as key cell cycle progression effectors particularly in the transition from the G1 and G2 phases to S and M phases respectively. Both E2F TFs are ubiquitously expressed throughout the development of *Oikopleura* regardless of the cell cycle mode predominantly employed (see Figure 3.1), pointing to a possible function of the TFs in all types of cell cycle. An exception to the ubiquitous pattern of E2F expression is in male gonads during Day 6 of development, where E2F7 is very lowly expressed. The trunk of Day 6 animals has very low expression of both TFs as well and is known to have many key regulators downregulated due to lack of proliferation of somatic tissues.

Despite the ubiquitous expression pattern of E2F TFs, differential binding and regulation of alternative genes can occur throughout development, leading to different biological outcomes. To investigate if E2F factors contribute to the use of different cell cycle modes through differential binding targets, we analyse ChIP-chip experiments of E2F1 and E2F7 TFs in dissected gonads of animals in Day 6 developmental stage of separate sexes.

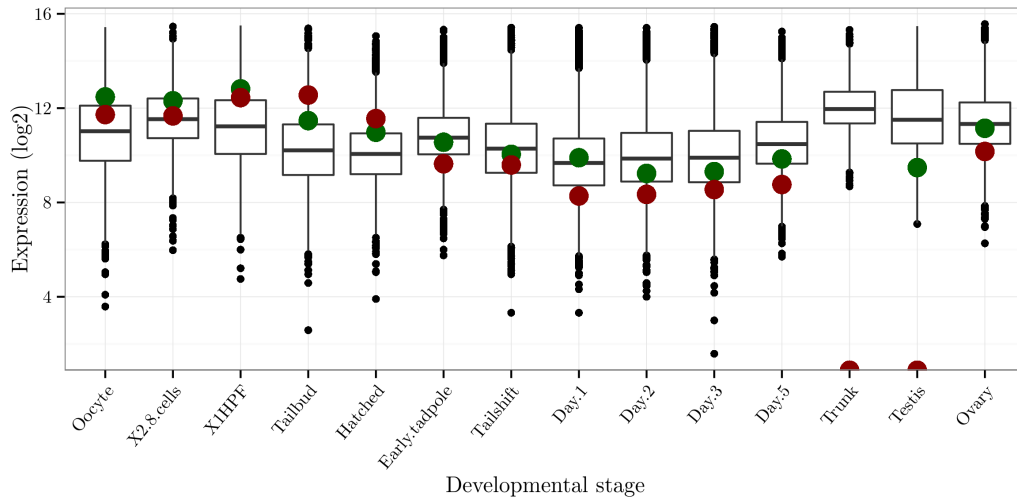


Figure 3.1: Expression levels of E2F factors during *Oikopleura dioica*'s development. Green: E2F1; Red: E2F7; Boxplots: All genes. Expression values transformed with a logarithm of base 2.

3.1.1 Detection of E2F binding through ChIP-chip

3.1.1.1 Validation of E2F antibodies

Antibodies used for immunoprecipitation of E2F TFs were validated by Western Blot (Figure 3.2) and immunohistochemistry as previously published by Subramaniam *et. al* [44] and found to detect the proteins in a specific manner.

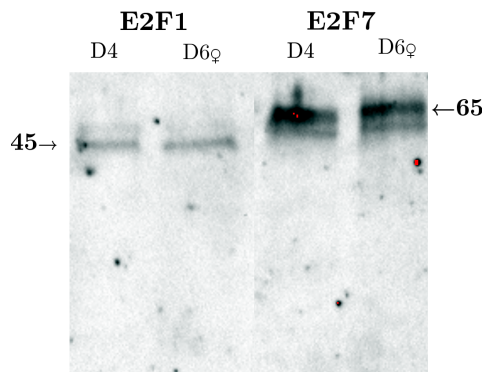


Figure 3.2: Detection of E2F proteins in two developmental stages of *Oikopleura dioica*. Molecular weight in kDa.

3.1.1.2 Genomic context of E2F binding sites

When conducting genome-wide analysis of features of interest (*e.g.* TF binding sites, cis-regulatory elements), the distribution of genetic and regulatory elements that make the background features of the genome under analysis mirror its architecture and can influence

3.1. E2F factors in *Oikopleura*'s late developmental stages

the statistical significance of the distribution of features of interest when compared with the background.

Oikopleura dioica possesses the most compact genome known yet for a chordate organism [38], with reduced intergenic and intronic space, and enlarged proportion of coding sequence when compared with other model systems, both invertebrates and vertebrates (Figure 3.3).

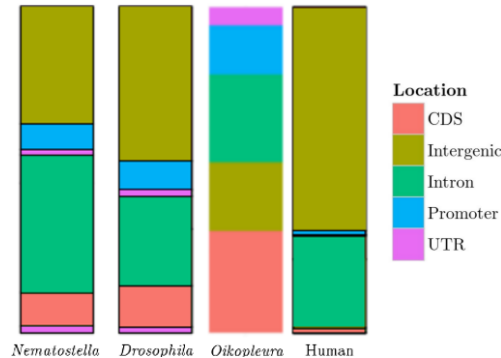


Figure 3.3: Comparison of the relative distribution of genomic features in *Oikopleura dioica* and other animal models. From left to right: *Nematostella vectensis*, *Drosophila melanogaster*, *Oikopleura dioica*, *Homo sapiens*. Data from all species except *Oikopleura* from [52].

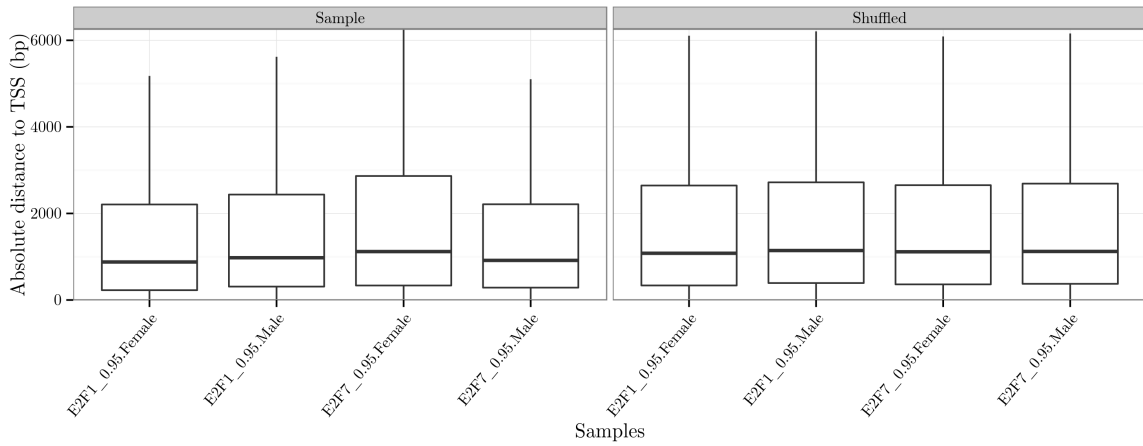
The knowledge of the background of features in the genome of *Oikopleura* will be useful when analysing the genomic context of E2F TF binding sites obtained by ChIP-chip. After preprocessing ChIP-chip data, putative binding sites of E2F TFs were obtained by selecting statistically significant ChIP-enriched regions (ChIP-ER) (see section 2.9.2 on methodology). We found tens of thousands of ChIP-ERs, where E2F1 samples have roughly similar size (16000) and the E2F7 sample in female D6 animals presented roughly the double of the male (24000 against 12000) (Table 3.1).

Table 3.1: N° of ChIP-enriched regions detected for all samples analysed

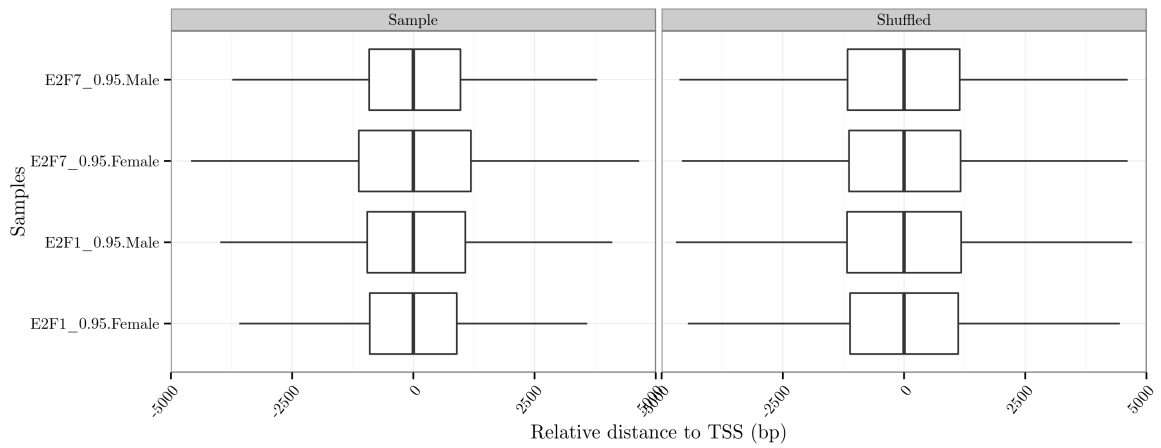
Stage	Sex	TF	N° ChIP-ERs
D6	Female	E2F1	16708
D6	Female	E2F7	24899
D6	Male	E2F1	16630
D6	Male	E2F7	12992

Transcription factors are known to regulate gene expression by binding cis-regulatory elements (CREs) (*e.g.* promoters, enhancers). Although CREs locate primarily in promoters and distal or intronic enhancer elements [53], virtually every type of genomic feature can serve as a TF binding site. In *Oikopleura*, E2F ChIP-ER were found to locate in close proximity to gene transcription start sites (TSS), regardless of the function of the TF (activator or repressor) and the developmental stage profiled (Figure 3.4a/b).

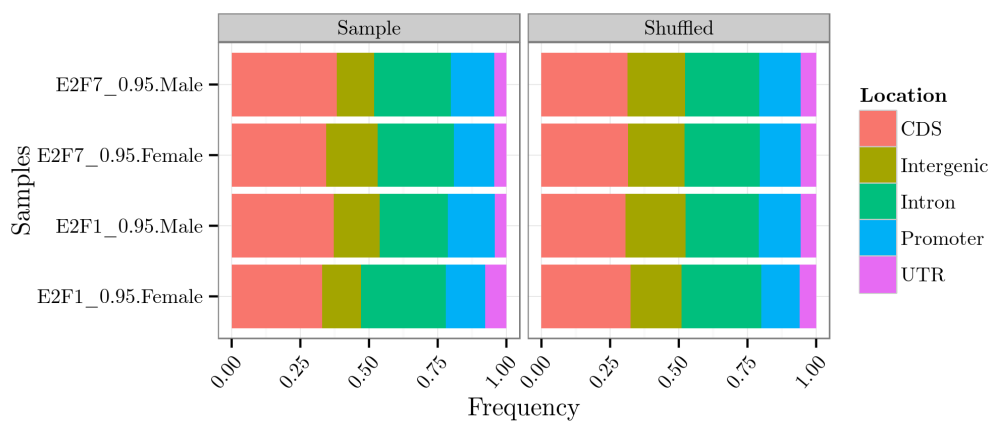
When comparing the distribution of observed distances with randomly positioned data of same size and widths, no statistically significant differences were found. Furthermore in both observed and randomized data the median distance to the nearest TSS coincided with



(a) Absolute distance of E2F ChIP-ER to the nearest gene TSS.



(b) Distance of E2F ChIP-ER relative to the nearest gene TSS.



(c) Location of E2F ChIP-ERs in the genome.

Figure 3.4: Genomic context of *Oikopleura* E2F ChIP-enriched regions. Shuffled data has same number and widths as ChIP-ERs.

3.1. E2F factors in *Oikopleura*'s late developmental stages

it, revealing the structural constraints imposed by the compact genome. Nevertheless, the width of the distributions of observed data were smaller than the randomized, showing some propensity for the observed data to cluster around gene TSSs.

Regarding the type of genomic features which E2F putatively bind, a considerable portion of ChIP-ERs overlaps regions of coding sequence(CDS) (Figure 3.4c) - this was unexpected, due to the general sequence constrains (*e.g.* in terms of codon usage) of CDS and due to the preferential location of TF binding sites in non-coding regions. The portion of ChIP-ERs overlapping CDS was in some cases higher than expected by chance. This could be explained by the combination of close proximity of ChIP-ERs to gene TSS and the wideness of ChIP-ER provided by the microarray technology, making overlaps between ChIP-ERs and CDS more likely.

Overall, the distribution of ChIP-ERs in the *Oikopleura* genome for both TFs in both stages analysed (Figure 3.4c) resembles what is expected by chance, and once again represents the constrains imposed by the genomic architecture.

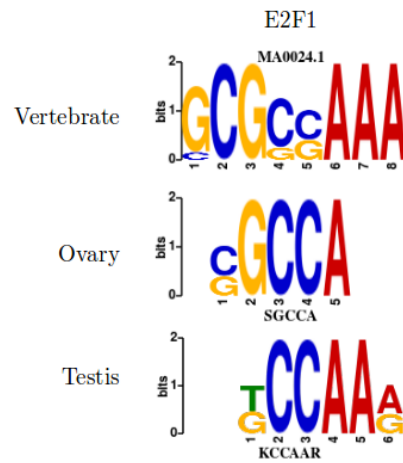
3.1.1.3 Insights into the mechanism of E2F binding

Transcription factors are by definition sequence-specific DNA-binding proteins, but TF binding has also been shown to be constrained by the local epigenetic state of chromatin. Current findings have shown that the existence of a balance between local epigenetic context and TF sequence recognition is a likely underlying the mechanism of TF binding at least for TFs important for development and differentiation.

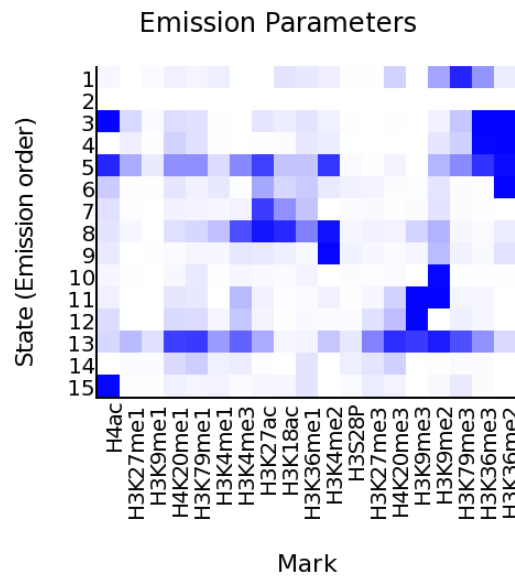
In E2F ChIP-ERs we found significantly overrepresented motifs that aligned significantly to the known E2F TF binding motif from vertebrate models (Figure 3.5a). The motifs found in *Oikopleura* are shorter in length than the vertebrate. This could be due to the evolutionary distance to vertebrates and the rapid rate of evolution of *Oikopleura*, and also by the fact that the vertebrate motif was acquired with a combination of techniques (*e.g.* SELEX, ChIP) which allows higher resolution. Nevertheless, both E2F TFs of *Oikopleura* seem to bind locations with sequence content conserved with vertebrates.

To explore the possibility that the local epigenetic environment conditions the binding of the E2F TFs, we investigated the distribution of E2F ChIP-ERs in relation to a genome-wide annotation of chromatin states based on histone modifications (P. Navratilova, G. Danks and E. Thompson, unpublished data). This annotation provides 15 discrete epigenetic states based on genome-wide profiles of 19 histone modification marks obtained by ChIP-chip (Figure 3.5b).

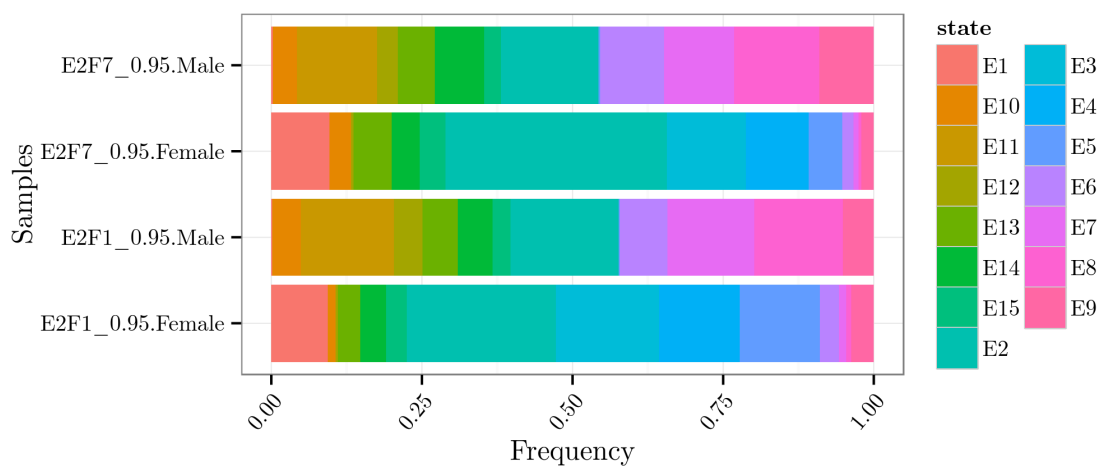
The pattern of distribution of chromatin states in ChIP-ERs is highly similar within the same developmental stage as opposed to the the expectation that TFs with the same regulatory function (activator, repressor) would be more similarly constrained and more similarly likely to transform the epigenetic environment. E2F ChIP-ERs didn't show any preference for a particular local epigenetic signature (Figure 3.5c). A possible explanation lies on the wideness of ChIP-ERs provided by the microarray technology in a very compact genome, for if a high proportion of ChIP-ERs overlap with many chromatin states, the overall resolution would drop, reflecting mostly the overall landscape of chromatin states and epigenetic regulation within a particular developmental stage. It is also possible that E2F factors are not particularly sensitive to or involved in remodeling the epigenetic lanscape by recruiting chromatin remodelers, such as known for TFs involved in core developmental functions and differentiation.



(a) Significant motifs found in E2F ChIP-ER that align with the known vertebrate E2F motif.



(b) Composition of chromatin states based on histone modifications in *Oikopleura dioica*. Unpublished data kindly provided by P. Navratilova, G. Danks and E. Thompson.

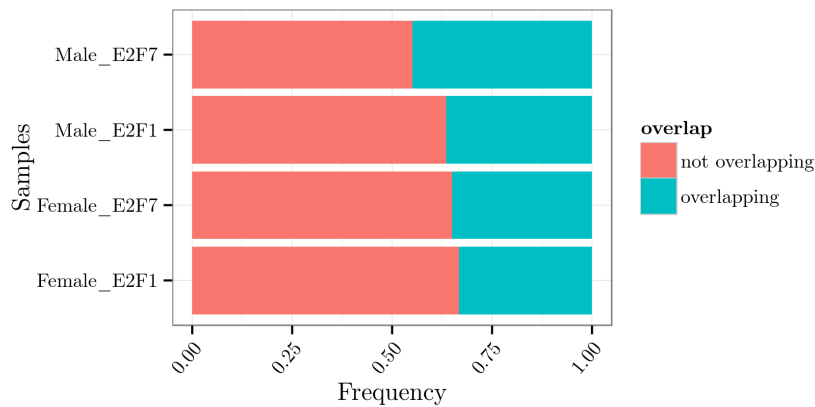


(c) Location of ChIP-ERs in the chromatin state annotation from (b).

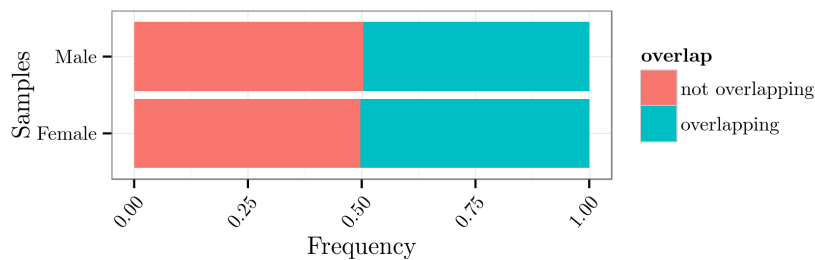
3.1. E2F factors in *Oikopleura*'s late developmental stages

3.1.1.4 Spatial and temporal co-localization of E2F factors

E2F activator and repressor transcription factors are known to antagonize the effect of each other by regulating an overlapping set of genes in opposed ways. To explore such relation in E2F TFs of *Oikopleura*, we measured the degree of overlap between the two. Also in *Oikopleura* ChIP-ERs of the activator (E2F1) and repressor (E2F7) showed high (45% as maximum) spatial and temporal co-localization as seen by the overlap of ChIP-ERs within the same developmental stage (Figure 3.6a). From the ChIP-ERs of the two TFs that overlap within developmental stages, roughly half of them overlap between developmental stages (Figure 3.6b), showing that there is indeed a high degree of overlap between the genomic regions bound and likely the set of genes regulated.



(a) Portion of ChIP-ERs that are shared (overlap) between transcription factors within the same developmental stage.



(b) Portion of shared (overlapping) ChIP-ERs between transcription factors that also overlap between developmental stages.

Figure 3.6: Portion of E2F ChIP-ERs shared (overlapping) within and between developmental stages.

However, since ChIP experiments were performed on tissues with unsynchronised nuclei, the signal from the ChIP experiment most likely gives an overall picture of the binding of E2F TFs during this developmental stage by masking the possibly more subtle changes in E2F action during a cell cycle, rather than providing an accurate description of events inside an "ideal nucleus".

3.1.2 Putative E2F regulated genes (target genes)

Transcription factor binding has the function of altering the expression of genes. TFs binding diverse genomic locations have been shown to regulate genes near its binding site or in wide distances including different chromosomes. To attribute a putative binding event to the regulation of a gene, we assigned each ChIP-ER to the nearest transcription start site of a gene and said that this gene is regulated by this E2F ChIP-ER. This approach has the advantage that it assumes less about the putative binding events, allowing genes to have multiple E2F regulatory elements in any genomic position. To study the functional properties of genes with an assigned ChIP-ER (from here on *target genes*, for simplicity), we divided the genes into groups based on whether they were the targets of one TF exclusively or both in one of the tissues studied (ovary or testis) in Day 6 animals. The total number of genes putatively regulated by ChIP-ERs can be seen in Table 3.2, along with the number of genes that are exclusively regulated by one TF in a tissue.

Table 3.2: N° of target genes from each TF

Stage	Sex	TF	N° of target genes	N° of exclusive target genes
D6	Female	E2F1	7327	5907
D6	Female	E2F7	11209	5504
D6	Male	E2F1	7342	5429
D6	Male	E2F7	7508	4814

3.1.2.1 Gene ontology of target genes

From genes that are targets of both TFs within one tissue, a significant overrepresentation of genes whose biological processes is related with DNA replication and nucleotide metabolism was found (Figure 3.7). Other prominent overrepresented groups of gene ontology (GO) terms found among these genes were related with cell cycle regulation and progression, formation of protein-DNA complexes (with TFs overrepresented among these) and processes related with development of adult traits.

Co-regulation of genes by the two TFs in the same development stage is likely due to the necessity of gaining fine control over the expression of these genes and highlights the importance of the E2F factors as key regulators of processes such as cell cycle progression and processes inherent to it (DNA replication, nucleic acid metabolism). Regulation of protein-DNA complexes and TFs is also not surprising due to the fact that regulation at the transcriptional level often exists in the form of gene regulatory networks of which TFs are the major regulatory force. Regardless, of the type of outcome from the action of a TF (activator or repressor), it is likely that other TFs will also be regulated accordingly. Since the stages of development studied here correspond to the last phase of *Oikopleura*'s life cycle, general regulation of genes involved in biological processes important for acquisition of adult traits (*e.g.* gametogenesis).

Genes exclusively targeted by the E2F1 TF in the female gonad have a set of overrepresented GO terms associated with them highly similar to genes bound by both TFs, but in addition terms related with cytoskeletal activity. This could be related to the fact that during maturation of the female gonad oocytes expand their cytoplasmatic domain drastically, occupying a volume higher than the somatic part of the organism - such accomplishment re-

3.1. E2F factors in *Oikopleura*'s late developmental stages

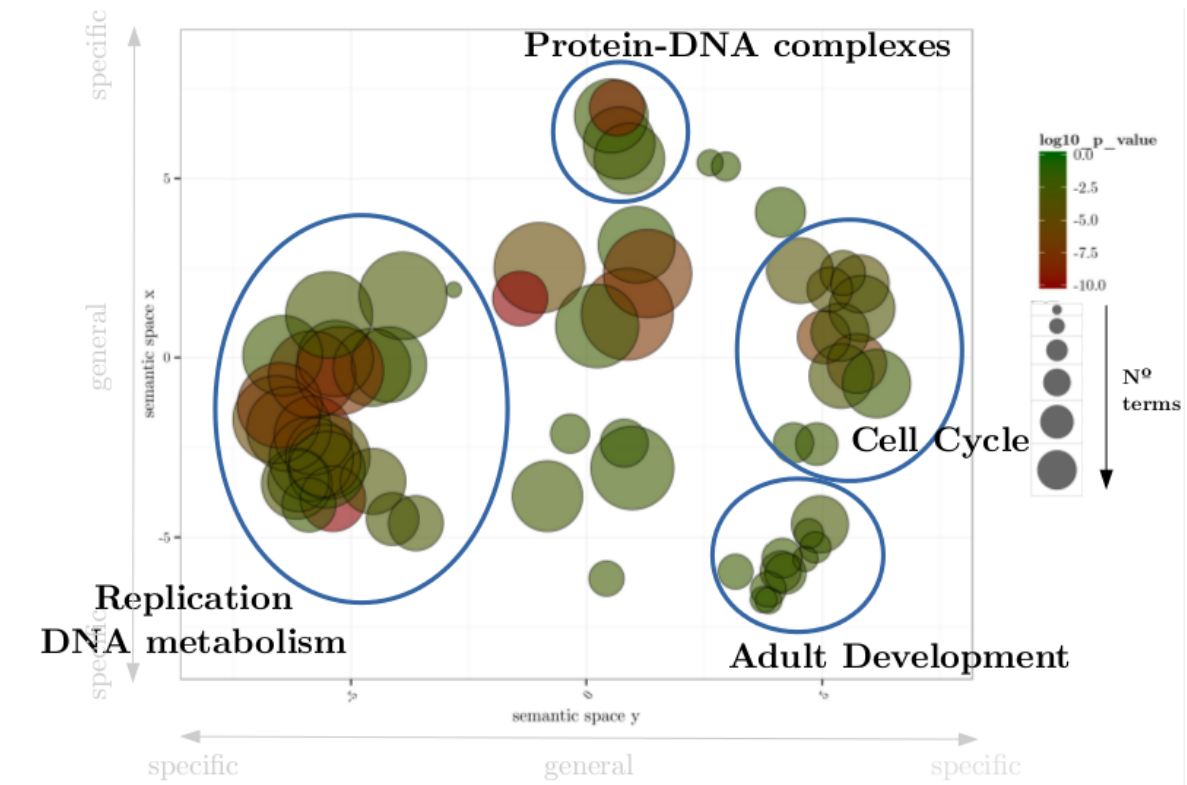


Figure 3.7: Gene Ontology of genes putatively regulated by both E2F transcription factors in testis of animals in D6 developmental stage. Data spread over two semantic spaces (axis). Gene Ontology terms were nested within the same parent and the amount of terms clustered is shown by the area of the circle. Significance values are shown in a color spectra. Ovary of animals in the D6 developmental stage have a high degree of similarity.

quires extensive production and coordination of structural proteins in the cytoskeleton. In the analogous situation (exclusively bound E2F1 genes) in the male gonad, much fewer overrepresented GO terms were found. Of particular notice is the presence of the "*metaphase/anaphase transition of mitotic cell cycle*" term, which is highly specific for M phase. The activation by E2F1 of genes involved in M phase is coincident with the resumption of cell cycles with mitotic phase which occurs in late maturation of the male gonad and during spermatogenesis (meiosis).

Genes under exclusive action of E2F7 in the female gonad have a scarce and reduced set of overrepresented GO terms associated with them but where again the "*metaphase/anaphase transition of mitotic cell cycle*" term stands out. The female gonad employs mainly endocycles throughout its maturation and it is foreseeable that genes involved in the promotion on M phase should be repressed to allow the endocycle. Genes exclusively bound by E2F7 in the male gonad have a rich set of overrepresented terms resembling the ones in Figure 3.7, but where terms such as "*meiosis*" and terms related with transcription and RNA metabolism also appear. This could be due to the minimalistic nature of sperm when compared with the oocyte - sperm is mainly a carrier of genetic information in the form of the genome, as opposed to the oocyte which possesses maternal transcripts and reserve substances to support early development. E2F7 binding of genes involved in these functions could be contributing

to their suppression.

This mirror situation of genes involved in the same biological processes being activated or repressed inversely in the male and female gonads is indicative of the differences underlying the different cell cycle modes employed in the two tissues. However, although analysis of gene function through testing of under/overrepresented GO terms is useful, it assumes that genes have kept the same functional properties (biological process, cellular localization, molecular function) as the genes from which the terms were originally derived. Full lists of significantly overrepresented GO terms can be seen on appendix section A.

3.1.2.2 Differences of expression in E2F target genes

The ultimate goal of transcription factor binding is to steer a developmental program in one direction by altering gene expression. Activator TFs positively influence gene expression whereas repressors have a negative effect by repressing or diminishing gene expression. We decided to explore the dynamics of gene expression in E2F target genes in *Oikopleura* to see to which degree do they condition gene expression in the tissues analysed.

There are major and significant changes of expression at the transcriptome level between ovary and testis tissue of animals in the D6 developmental stage (Figure 3.8a). Overall, there is a bias for genes being more highly expressed in the female gonad. In the male, a high fraction of genes are not expressed at all while being expressed at various levels in the ovary (note points stacked on x axis in Figure 3.8a). This illustrates the intrinsic Biology of the maturing gonads - in the ovary, meiotic and nurse nuclei coexist and the expression values reflect genes expressed in both (possibly even different sets of genes).

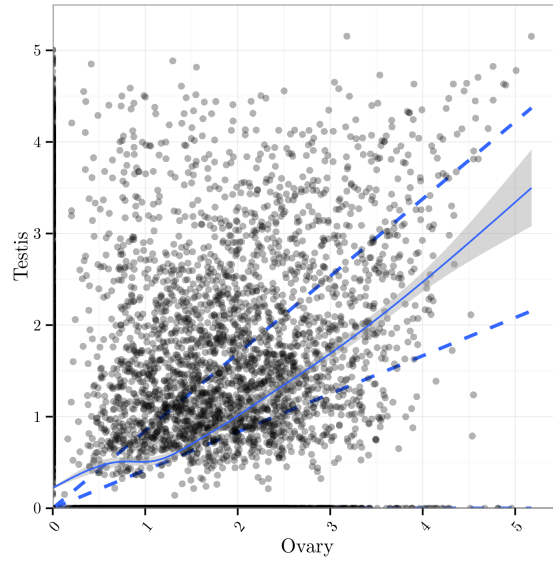
We tested the hypothesis that genes that are exclusively targeted by the E2F1 TF (activator) in a particular tissue would be more highly expressed in that tissue than in the other, and the inverse situation for genes exclusively targeted by E2F7 (repressor). There were no significant differences between the means of the distribution of these two groups in either tissue. However, a trend exists for genes that are bound by the repressor transcription factor (E2F7) in a stage to be less expressed in that stage and more in the other as shown by the inversion of position of smoothed and quantile lines in Figure 3.8(b) and (c) between tissues.

3.1.2.3 E2F regulation of key cell cycle effectors

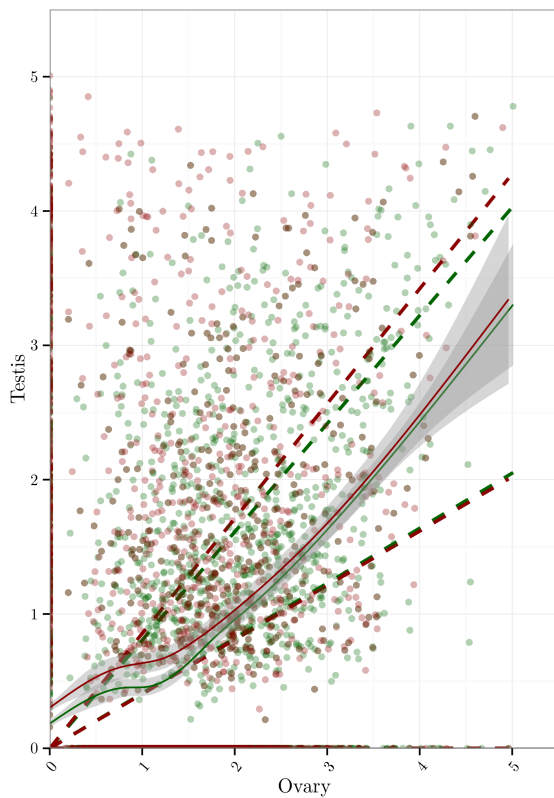
Since differential regulation of the canonical cell cycle and endocycles arise by the alternative regulation of key cell cycle effectors of progression through the cycles, it is foreseeable that some might be differentially regulated by the E2F TFs in the two gonadal tissues. The vast majority of such key cell cycle regulators components were detected to be putatively regulated by at least one E2F TF in at least one of the tissues. Such effectors are Cyclin, CDK proteins and CKI proteins as well as proteins directly involved in E2F regulation and other known regulators of cell cycle progression such as the co-transcriptional factor DP, the Retinoblastoma protein (Rb), Geminin, Cdt1, Cdh1 and PCNAs. Interestingly, all four PCNA genes existent in *Oikopleura* are bound by both TFs in both tissues.

However, due to the lack of replication and the wideness of ChIP-ERs provided by the microarray technology, analysis of discrete cases of gene regulation by E2F factors should be taken with caution. Lists of key cell cycle effectors regulated by E2F TFs can be found in the appendix section B).

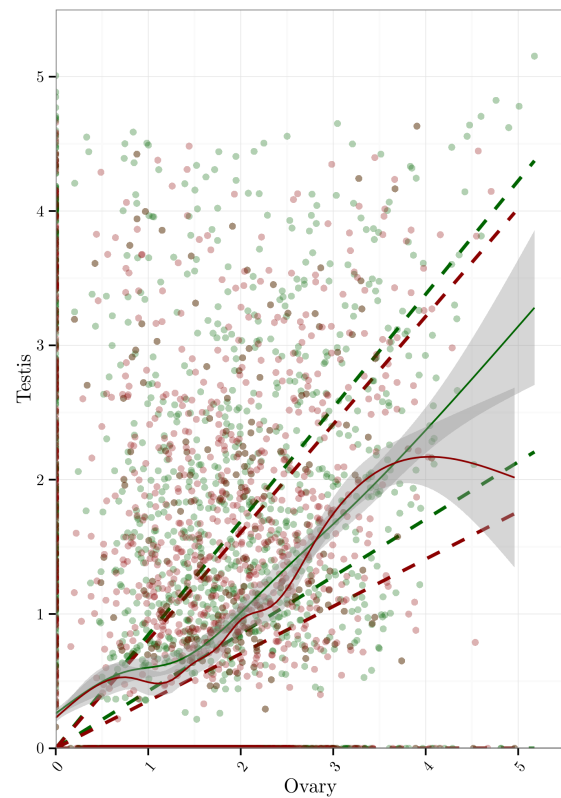
3.1. E2F factors in *Oikopleura*'s late developmental stages



(a) Expression values of all genes in ovary and testis tissue of animals in the D6 developmental stage.



(b) D6 Female developmental stage (Ovary).



(c) D6 Male developmental stage (Testis).

Figure 3.8: Expression values of genes in ovary and testis tissue of animals in the D6 developmental stage. (a) All genes; (b, c) Genes exclusively targeted by one E2F TF in (b) ovary; (c) testis; Green: E2F1-bound genes; Red: E2F7-bound genes; Expression values were normalized by the mean expression of all genes within each stage and transformed with a logarithm of base 2; Dashed lines indicate quantiles of the distribution of expression values; Smoothed line represents the smoothed conditional mean of expression values.

3.2 Dot1 and H3K79 methylation in *Oikopleura dioica*'s cell cycle modes

Methylation of lysine 79 in histone H3 (H3K79me) has been recently shown to be laid in a distributive manner by the Dot1 methyltransferase leading to its accumulation during a cell cycle [30]. This led to the proposal H3K79me levels might act as memory for the age of the cell and contribute to the regulation of cell cycle progression [31]. Studies on H3K79me and the function of Dot1 in Human cells have shown that knock-out of Dot1 produces cells with a high degree of ploidy due to re-replication in the same cell cycle [33], a situation similar but inverse to *Trypanosoma brucei*, where overexpression of Dot1 has the aforementioned effect [32].

3.2.1 Use of H3K79me along *Oikopleura*'s life cycle

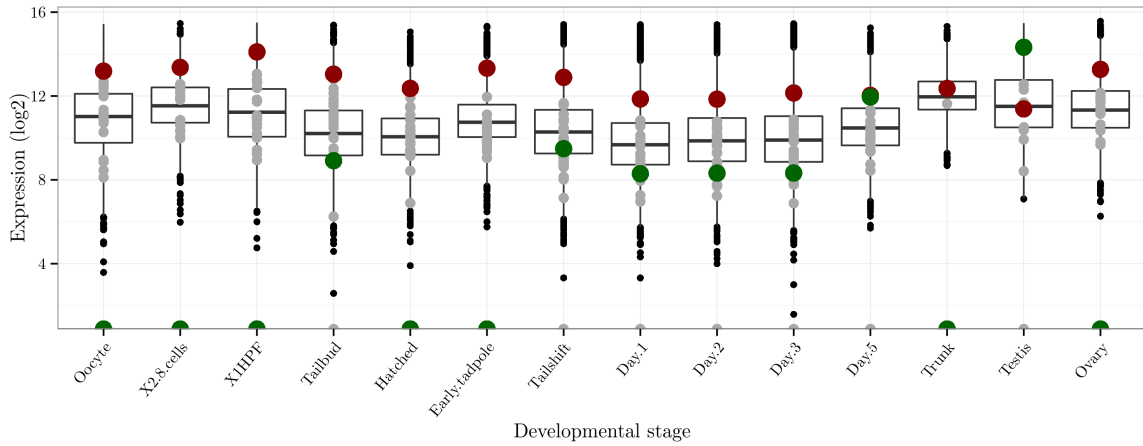
To investigate whether H3K79me would have a function in the regulation of *Oikopleura dioica*'s cell cycle modes and contribute to the differences between them, we conducted an initial functional study on Dot1 through the use of a chemical inhibitor in the Tailbud stage (mitotic) and the Day 5 (endocycle) and characterization of the phenotypic effects.

The Dot1 methyltransferase is ubiquitously expressed throughout the development of *Oikopleura* with no exception (Figure 3.9a), indicative of its vital function. Nevertheless, the Bre1 ubiquitin ligase, responsible for laying monoubiquitination of lysine 123 on histone H2B (H2BK123ub1), of which Dot1 depends to lay its mark, is mostly expressed from the Tailshift developmental stage on (Figure 3.9a), coincident with the employment of endocycles in *Oikopleura*. In the last day of the life cycle (D6), where sexual dimorphism is most apparent, Bre1 ceases to be expressed in the somatic tissues of the Trunk and in the gonads of female animals, while having the highest expression level of the whole cycle in the male testis, indicating a sex-specific differential use of Bre1 during gonad maturation.

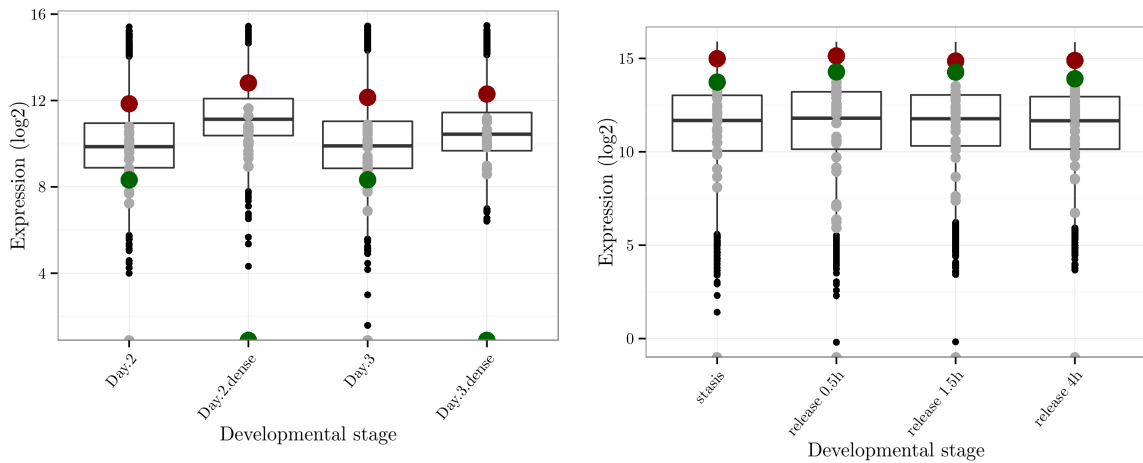
Oikopleura is known to possess a developmental condition known as growth arrest (GA) with reduced proliferation and growth when cultured densely due to reduced nutritional resources (see section 1.3.5). During GA, expression of Bre1 is significantly reduced whereas the Dot1 methyltransferase is relatively stable (Figure 3.9b). When animals are released from the growth arrest condition by restoring the normal culture density and nutritional supply, Bre1 expression is restored to levels compared with normal development during endocycling stages (Figure 3.9c). This suggests that Bre1 is more highly expressed during periods of proliferation and reinforces the idea it might have a function during endocycles, for this type of cell cycle mode is rapidly inactivated when in GA and resumed when animals are released from this condition [44]. Nevertheless, this putative increase of activity of Bre1 in endocycles does not imply any particular connection with the increase of H3K79me levels.

To gain more knowledge on the dynamics of H3K79me during *Oikopleura*'s life cycle we assessed the different states of H3K79me by western blot on whole animals (Figure 3.10a). Developmental stages with a predominant use of endocycles show overall high levels of K79 methylation, while stages with a predominant use of a mitotic cell cycle show decreased levels of K79 methylation. Additionally, in the Tailbud stage (mitotic), the ratio of H3K79me1 to H3K79me3 is in favor of the least methylated form, while on Day 5 all three forms seem equally present and in Day 6 female animals the most methylated form becomes more predominant. In Day 6 male animals, H3K79me3 was atypically detected, probably due to the existence of testis-specific variant of H3, H3.t, which lacks the lysine 79 residue. This suggests that higher

3.2. *Dot1* and H3K79 methylation in *Oikopleura dioica*'s cell cycle modes



(a) Life cycle with optimal conditions.



(b) Normal culture and growth arrest (dense) conditions.

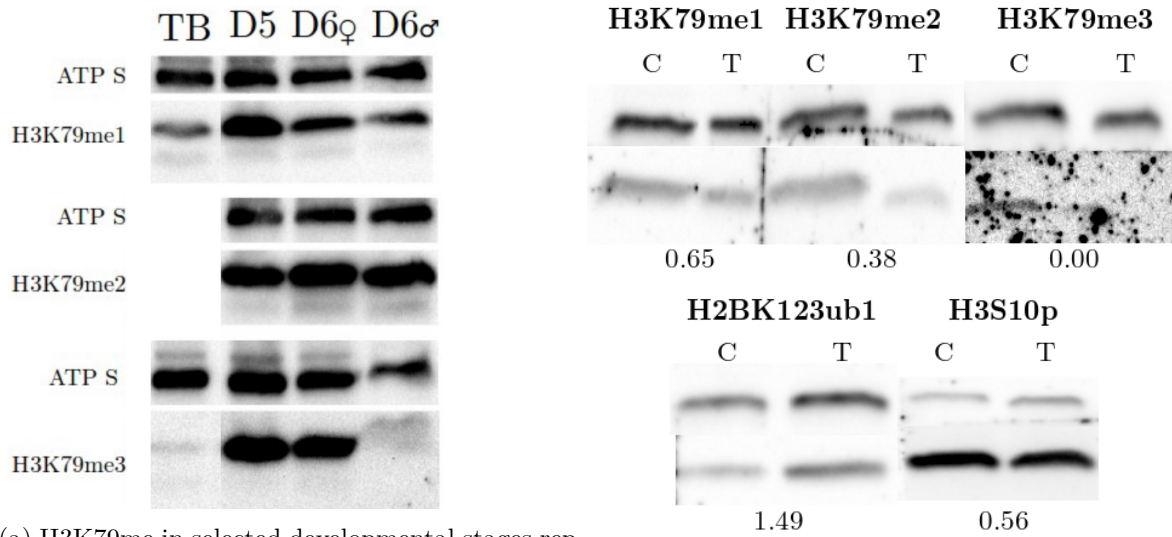
(c) After release of growth arrest condition.

Figure 3.9: Expression values of *Dot1* (red), *Bre1* (green), all SET domain-containing enzymes (grey) and all genes (boxplots) during (A) optimal *Oikopleura* development. Comparison of (B) normal development and growth arrest, and (C) after release from the growth arrest condition. Expression values were transformed with a logarithm of base 2.

methylation forms of H3K79 tend to be more present towards the end of the life cycle of *Oikopleura*, concurrent with the idea of accumulation of H3K79me with time.

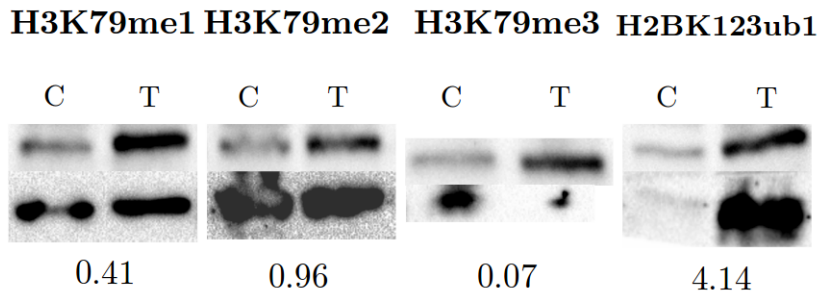
3.2.2 Functional studies on *Dot1* in two cell cycle modes of *Oikopleura*

To gain knowledge of the function of H3K79me we used a potent and specific chemical inhibitor of *Dot1*, SGC0946 (from now on *Dot1*ⁱⁿ) [54] on animals of the Tailbud stage. Treatment with 30 μ M of *Dot1*ⁱⁿ for 3.5 hours was shown to be effective in inhibiting *Dot1*, as assessed by the reduced levels of H3K79me compared with control animals (treated with the solvent DMSO) (Figure 3.10b). Depletion of H3K79me was most successful in higher



(a) H3K79me in selected developmental stages representative of the different cell cycle modes.

(b) Tailbud stage.



(c) D5 stage.

Figure 3.10: Quantification of H3K79 methylation states during selected developmental stages of *Oikopleura dioica*'s life cycle and in *Dot1*ⁱⁿ-treated animals by Western blot. In (b) and (c), values under blots show the relative amounts compared with control normalized by ATP synthase *beta* subunit (upper band). C - DMSO-treated control animals; T - *Dot1*ⁱⁿ-treated animals.

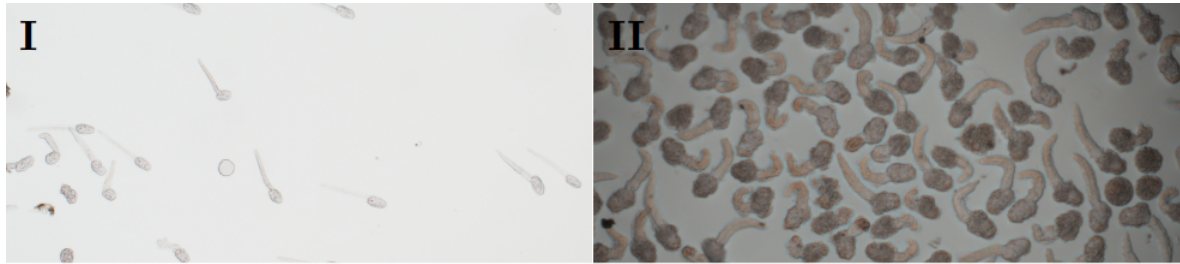
methylation states probably due to the reduced activity of Dot1 on newly deposited histones during replication, since the lower methylation states are necessary for the addition of higher.

Since the presence of H2BK123ub1 is known to increase the deposition of H3K79me by Dot1, we decided to explore the connection between H3K79me and H2BK123ub1, by assessing H2BK123ub1 levels in Tailbud animals treated with *Dot1*ⁱⁿ. Levels of H2BK123ub1 were higher in *Dot1*ⁱⁿ-treated animals when compared with control (treated with the solvent DMSO) (Figure 3.10b). Although the inverse relationship (dependency of H3K79me on H2BK123ub1) had already been described, increase of H2BK123ub1 in response to decreased levels of Dot1 and H3K79me had not and was therefore surprising.

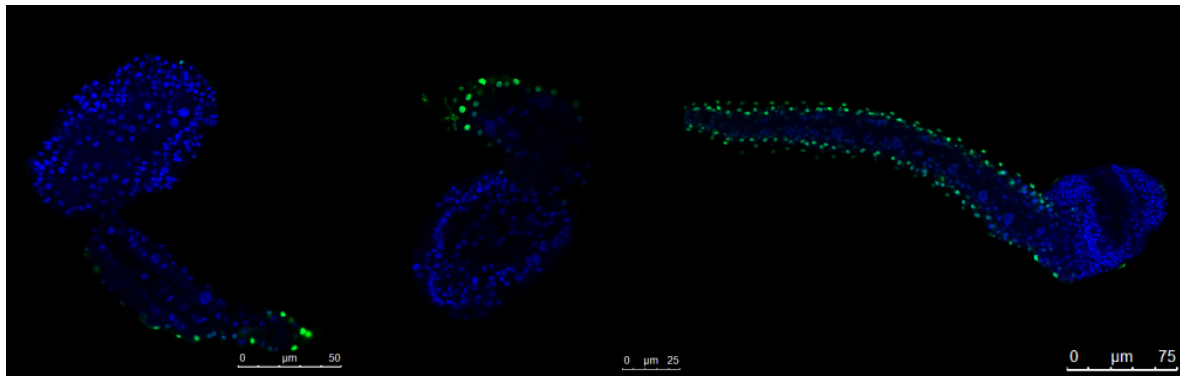
Tailbud animals treated with *Dot1*ⁱⁿ presented phenotypic effects that revealed developmental defects: the surface of the trunk was largely affected, with cells presenting a granulous appearance (Figure 3.11a), failure to maintain swimming activity and later death. To further characterize the phenotype, we assessed the state of cell proliferation by the use of the mito-

3.2. *Dot1* and *H3K79* methylation in *Oikopleura dioica*'s cell cycle modes

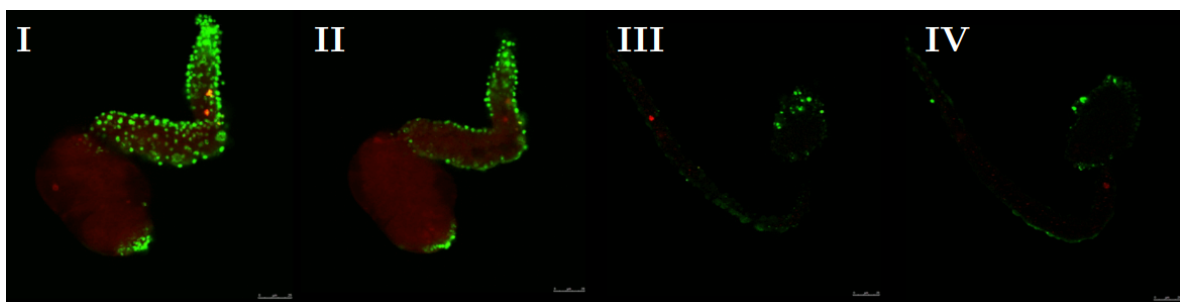
sis marker phosphorylation of serine 10 in histone H3 (H3S10p) and apoptosis by the TUNEL assay. The TUNEL assay in normal animals revealed cells marked for apoptosis along the tail of animals, starting from the posterior tip and progressing to anteriorly with developmental time (Figure 3.11b), indicating a possible use of apoptosis in regulation of morphogenic processes during this stage of development of *Oikopleura*. *Dot1*ⁱⁿ-treated animals showed reduced levels of H3S10p, indicative of reduced cell proliferation (Figure 3.10b) signs of apoptosis in the granulose cells of the trunk, especially at its surface (Figure 3.11c).



(a) Phenotypic effects of *Dot1*ⁱⁿ-treated animals compared with DMSO-treated. I: DMSO treated; II: *Dot1*ⁱⁿ-treated. Animals cultured at same densities.



(b) TUNEL assay on untreated control animals during development of the tadpole.



(c) Detection of apoptotic cells using the TUNEL assay. I-II: DMSO treated; III-IV: *Dot1*ⁱⁿ-treated.

Figure 3.11: Effects of the treatment with *Dot1*ⁱⁿ in Tadpole animals. Green: TUNEL assay; Red: H3S10p; Blue: DNA;

Inhibition of the *Dot1* methyltransferase seems to have a severe effect, causing deficiency of proliferation, developmental defects and death in the Tadpole developmental stage of *Oikopleura*, a stage where a mitotic cell cycle is employed. Since *Dot1* has been shown to be

involved in other developmental processes that may or may not be related with the addition of H3K79me marks (*e.g.* transcription), it is not safe to assume that the phenotypic effects seen are the consequence of H3K79me depletion.

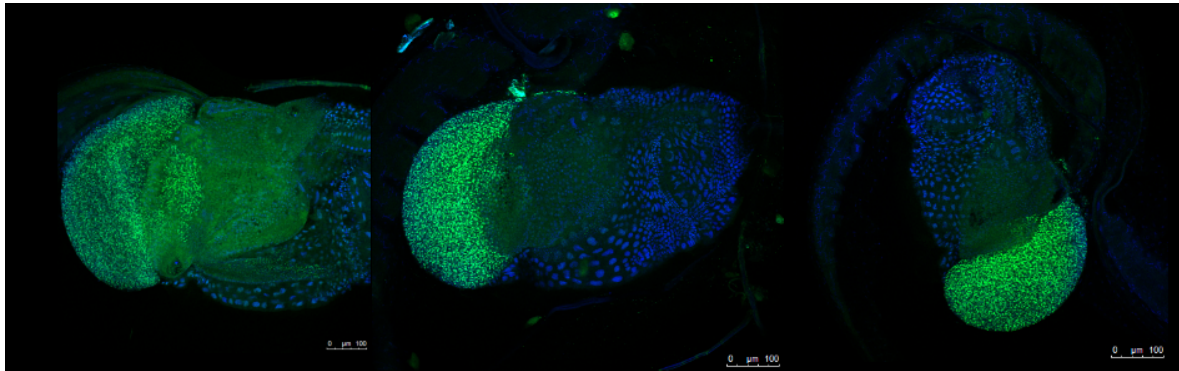
Since the overall amount of H3K79me levels is higher in the endocycling mode of proliferation and the ratio of the methylation states appears to change during the different cell cycle modes, we studied the effects of Dot1 inhibition in the D5 developmental stage, where endocycling is employed extensively in both somatic and gonadal tissues.

Overall levels of H3K79me were reduced although not as much as in the Tailbud stage (Figure 3.10c). This might be due to the fact that although H3K79me is not copied into new histones during replication, old histones will still carry the same methylation state of H3K79 and because on the D5 stage overall levels of H3K79me are high. This is furthermore amplified by the fact that in the female gonad most nuclei (either nurse or selected oocytes) in the coenocyst are already present at this time, making the most important step of H3K79me dilution - replication - more infrequent. Increased H2BK123ub1 was also detected on Dot1ⁱⁿ-treated when compared with control, reinforcing the hypothesis of an indirect negative feedback loop between H3K79me and H2BK123ub1 in *Oikopleura*.

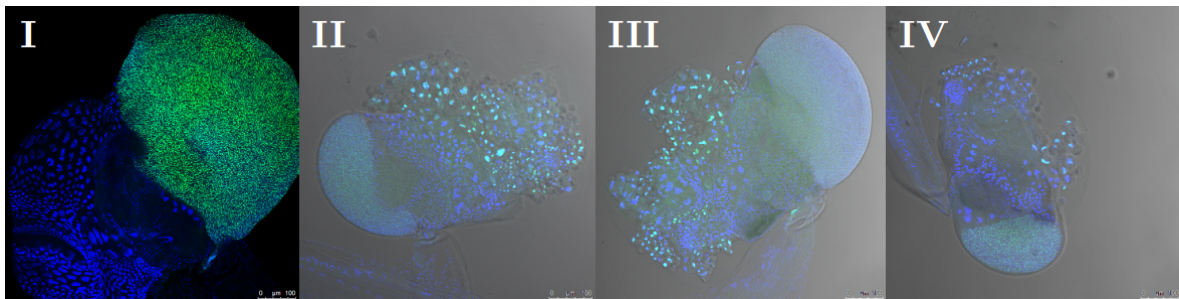
To assess the rate of proliferation of cells in this developmental stage, we measured the incorporation of a labeled nucleotide analog 5-ethynyl-2-deoxyuridine (EdU) using 1 hour pulses and quantifying fluorescence of the labeled nucleotides. In this developmental stage, under optimal conditions cells proliferate actively throughout the *Oikopleura* body, but towards the end of the life cycle (entering Day 6), proliferation is restricted to the gonadal tissue, when somatic tissue ceases to proliferate (Figure 3.12a). In Dot1ⁱⁿ-treated D5 animals, gonad proliferation is severely reduced although still detectable (Figure 3.12b) and proliferating somatic cells were detected as opposed to control animals which terminated somatic proliferation as measured by EdU incorporation. Somatic cells marked with EdU were abnormally large and had propensity to detach from the animal body (Figure 3.12c). Surprisingly, animals exhibiting this phenotype remained alive and capable of filtration functions before dying, much later than in the Tailbud stage when comparing time after treatment.

The surprising inversion of the proliferation pattern in the body of Dot1ⁱⁿ-treated D5 animals might be connected with the function of Dot1 in providing a molecular clock for the cell's age by laying H3K79me and indicate a function for Dot1 in the regulation of proliferation, perhaps by regulating the timing of replication.

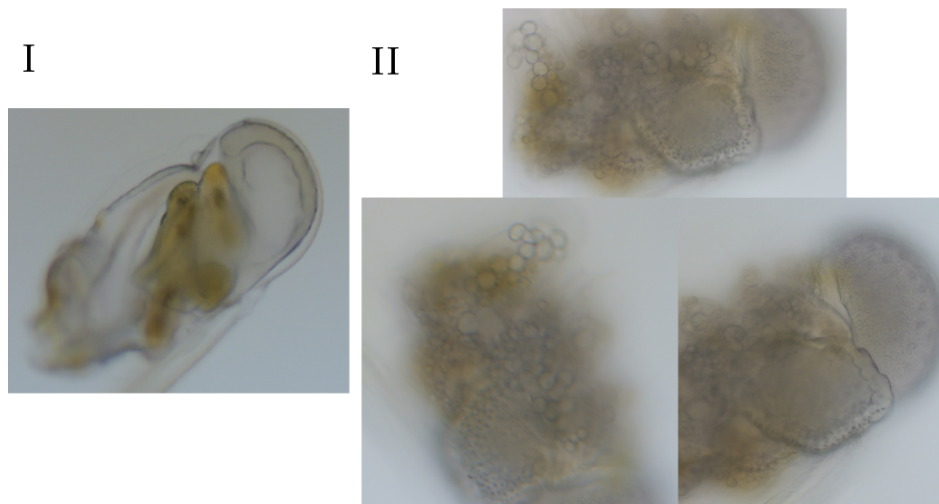
3.2. *Dot1* and *H3K79* methylation in *Oikopleura dioica*'s cell cycle modes



(a) Changes in proliferation during the D5 stage in normal untreated animals.



(b) Changes in proliferation in *Dot1lin*-treated animals compared with DMSO-treated. I: DMSO treated; II-IV: *Dot1lin*-treated - images overlaid with light microscopy for better detection of the cell's shape.



(c) Phenotypic effects of *Dot1lin*-treated animals compared with DMSO-treated. I: DMSO treated; II: *Dot1lin*-treated.

Figure 3.12: Assessment of the proliferative state of animals during a predominant endocycle mode of by detection of EdU incorporation. Green: EdU; Blue: DNA;

Chapter 4

Discussion

4.1 E2F factors in *Oikopleura*'s late developmental stages

Alternative cell cycle modes lacking a mitotic phase and creating polyploid cells have been shown to be more predominant than previously expected. Deregulation of progression through phases of cell cycle and ploidy are often at the core of disorders such as cancer. Current approaches to study such disorders often start from a state where deregulation is already well advanced and often irreversible. The marine chordate *Oikopleura dioica* presents the opportunity to study a system naturally employing different cell cycle modes through its life cycle. The study of the mechanisms governing the change between cell cycle modes could contribute in providing a fresh take on disorders which present deregulation of cell cycle progression.

In this work, we focused on the role of the two E2F transcription factors (E2F1 - activator; E2F7 - repressor) in *Oikopleura*'s cell cycle modes. They are known to play a major role as key cell cycle progression as regulators of an extensive array of genes. By analysing the binding of E2F TFs and their putatively regulated genes, one can identify differentially regulated targets in the different modes of cell cycle.

We identified thousands binding sites for both E2F TFs in ovary and testis tissue of *Oikopleura* animals of Day 6 developmental stage through ChIP-chip. The compact genome of *Oikopleura*, with its reduced intergenic and intronic space, conditions heavily the available genomic space for E2F TFs binding. E2F TFs bind in close proximity to gene's TSSs and a unexpectedly high portion of the binding sites overlap with regions of coding sequence. This might be due to the combination of TSS proximity and wideness of detected ChIP-ERs provided by the microarray technology. It is also possible that these represent *bona-fine* binding sites that have a function on transcription regulation by altering splicing patterns, or interfering with the transcription process by interacting with the components of the transcriptional machine (*e.g.* RNA Polymerase II. Regardless of the functional outcome of the presence of cis-regulatory elements (CRE) (*e.g.* enhancers, insulators) in coding sequences, the hypothesis that evolutionary pressure on the non-coding DNA sequence caused by a major event of genome compaction as occurred in *Oikopleura* [38], may have caused widespread relocation of distal CREs to coding sequences, cannot be discarded. This would imply heavy rewiring of the gene regulatory networks underlying gene expression and high sequence constraints for the codons being used as CREs as has been recently proposed by Stergachis and colleges [55] with controversy. The binding of *Oikopleura* E2F TFs seems more dependent on the sequence-

specificity than on the local context of chromatin. Motifs aligning significantly to vertebrate E2F TFs motifs show that the sequence specificity of E2F TFs is largely conserved.

The two E2F TFs (activator and repressor) were shown to have a high degree of spatial and temporal co-localization in their binding patterns throughout the genome, with many genes being regulated by both TFs in the same tissue. Our data does not possess resolution of their possible dynamic binding during cell cycles, but can still be used to infer differences of binding sites between different cell cycle modes, rather than along the progression of these cycles. In the set of genes that are regulated by both TFs, there is overrepresentation of functions related with DNA replication and metabolism, cell cycle regulation - highlighting the function of E2F TFs in the control of cell cycle progression, specifically S phase entry. These genes are likely regulated in either opposite directions by the TFs in different phases of the cell cycle, thus implying a oscillating expression during a cell cycle or the coincident binding of both might functions as a way to gain finer control over gene expression across or during phases of the cell cycles.

Both E2F factor also regulate a significant portion of genes independently. When exploring the function of genes regulated by only one of the TFs in one tissue, we found a curious inversion of a very specific function related with cell cycle regulation. The "*metaphase/anaphase transition of mitotic cell cycle*" GO term is highly specific for transitions within M phase, the cell cycle phase being skipped in endocycles. In the ovary, this term was associated with genes being solely regulated by E2F7 indicating that effector genes in M phase are being repressed, while in testis their expression happened to be induced by E2F1 binding. This inversion of E2F targets in the two tissues might be at the core of the suppression of M phase and endocycle regulation.

Many key cell cycle effectors are under E2F regulation in *Oikopleura*. Cyclin E, for instance, which modulation of expression during endoreplication in various models has been shown to be crucial [11] [13] [12], is differentially regulated by E2F factors in the two tissues studied. Nevertheless, although a certain degree of conservation between the regulation of endocycles and endoreplicative cycles between species is expected, further dissection of specific cell cycle effectors crucial for these alternative cycles must occur to attest the conservation of these mechanisms.

Although an enrichment in cell cycle control functions is evident, many genes under E2F regulation are likely to participate in regulation of other biological processes. The reduced family of E2F TFs in *Oikopleura* compared with other models could also have lead to diversification of E2F function as is suggested by the significant overrepresentation of GO terms related with adult development.

Differences of mean expression between genes targeted by one or other E2F TF in a tissue were found not to be significant. Nevertheless, there was a trend for genes regulated by E2F7 in a particular tissue being less expressed in the same tissue when compared with the other. Many genes, notably transcription factors are known to be regulated transcriptionally in regulatory networks that can enhance or buffer a given regulatory signal through feedback loops to allow high control over gene expression [56]. In light of this, we propose that most binding events of E2F TFs modulate gene expression in a smoothed way through interaction of various regulators, rather than controlling it in a boolean fashion.

Overall, the relatively low resolution of the microarray technology, coupled with the high number of ChIP-ERs detected and putatively regulated genes, makes detection of subtle differences in E2F regulation in the two cycles harder. This is further exacerbated by the possible function of E2F factors in other developmental processes. Nevertheless, profiling

4.2. *Dot1* and H3K79 methylation in *Oikopleura dioica*'s cell cycle modes

of E2F binding through ChIP-seq in more developmental time points with replicates as performed experimentally in this work (data not analysed) will allow higher resolution in binding site identification and more statistical confidence on the relation between E2F binding and gene regulation.

4.2 *Dot1* and H3K79 methylation in *Oikopleura dioica*'s cell cycle modes

In this work we started by exploring the use of different H3K79me states in the life cycle of *Oikopleura*. The overall levels of H3K79me are higher towards the end of *Oikopleura*'s life cycle, when endocycles are employed. Furthermore, the ratio between H3K79me1 and H3K79me3 methylation states decreases in endocycles, indicating the predominance of the higher methylation form. This is coincident with the higher expression level of the Bre1 enzyme (the ubiquitin ligase responsible for H2BK123ub1) in developmental stages employing endocycles as opposed to mitotic cycles. Taken together, results show that during endocycles there is a more widespread use of H3K79me agreeing with the view that it accumulates with cell cycles and might act as a timer for a cell age proposed by De Vos and colleges [31].

To gain more insights into the mechanism of H3K79me regulation we disrupted *Dot1*'s function with a chemical inhibitor (*Dot1*ⁱⁿ). It was found that in mitotic cycles, treated animals gain developmental defects, present reduced proliferation overall, have cells marked for apoptosis and die. An interesting observation came from the increase of H3K123ub1 in *Dot1*ⁱⁿ-treated animals. H3K123ub1 was known to upstream of H3K79me but the opposite had never been reported. A possible mechanistic explanation could be inspired by high levels of H3K79me3 found in the locus of the Bre1 enzyme in testis of animals in D6 developmental stage (P. Navratilova, G. Danks and E. Thompson, unpublished data), a stage where Bre1 expression is notably low (Figure 3.9a). In this situation, increased H3K79me levels in the Bre1 locus could be responsible for its very low expression, which in turn would reduce the activity of *Dot1* creating a negative feedback loop. If this mechanism would be present throughout *Oikopleura*'s life cycle, in the present situation of *Dot1* inhibition in the Tailbud stage, H3K79me levels are decreased and Bre1 expression would be enhanced, thus being responsible for the upregulation of the H2BK123ub1 mark in *Dot1*ⁱⁿ-treated animals.

Dot1 function has been manipulated in these two model systems (the unicellular parasite *Trypanosoma brucei* and in Human). In both, disruption of normal *Dot1* function causes re-replication ending in cells with increased ploidy content, but this outcome proceeds from inverse manipulation (overexpression in *Trypanosoma*, depletion in human cell culture). The high expression of *Dot1* and Bre1 together with high presence of overall and highly methylated forms of H3K79 during endocycles in *Oikopleura* suggests *Oikopleura*'s use of and H3K79me is more similar to the situation in *Trypanosoma* than in human cell culture. In this case high levels of H3K79me would allow and/or promote or be the consequence of cell cycles where replication has occurred more than once before cell division. Nevertheless, there is no indication of shared molecular mechanisms regulating H3K79me in the two species. Experiments of overexpression of *Dot1* in *Oikopleura*, particularly in stages of mitotic division could shine light over this possibility.

In a developmental stage employing mainly endocycles (D5), animals acquired severe deregulation of proliferation, causing exaggerated proliferation of somatic cells (at this stage normally quiescent) and reduced proliferation in the gonadal tissue. These highly proliferative

somatic cells exhibited large volume and easily detached from the body of the animal. This odd cellular behaviour has some parallels with cancer cells, where highly proliferative epithelial cells, which are often polyploid lose contact with its natural substrate. However speculative this hypothesis is, it cannot be disregarded that Dot1 knockout has been shown to contribute to reprogram somatic cells to a pluripotent state [57] and that this change of potency states is also often correlated with the appearance of cancer.

Future experiments including the impairment Dot1 function by an alternative mean (*e.g.* dsRNA injections) would be valuable to ascertain whether these effects are specific to Dot1 inhibition and to gain insights into the mechanisms of action (*e.g.* due to H3K79me deposition or Dot1 involvement in other functions). Assessment of the ploidy state of cells of Dot1ⁱⁿ-treated animals by flow cytometry could provide a more accurate description of the events of proliferation deregulation and the use of cell cycle modes. ChIP-seq of H3K79me in Dot1ⁱⁿ-treated animals to detect if depletion of H3K79me is generalized or locus-specific through *Oikopleura's* genome. Levels of H3K79me on origins of replication have been correlated with the inhibition of re-replication in the same cell cycle [33].

Although it is clear that Dot1 has a critical function in regulation of cell cycle and proliferation in *Oikopleura*, it is not clear if that influence is a direct consequence of the lack of H3K79me or arises by the impairment of Dot1 in other cellular functions.

Chapter 5

Conclusion and Future Perspectives

The goal of this thesis was to study the mechanisms of regulation and transition between different cell cycle modes of *Oikopleura dioica*. Our main focus was the study of the role of the family of E2F transcription factors, of which we identified their binding sites genome-wide and characterized their putatively regulated genes. We also explored the contribution of histone modifications to cell cycle regulation. This study shows that regulation at both transcriptional and epigenetic levels is required and probably both play a role in the regulation and transition between *Oikopleura*'s cell cycle modes.

We've described the mode of action of E2F TFs by showing its close-to-TSS binding, restricted by the compact *Oikopleura* genome. We've shown that E2F TFs regulate a network of thousands of genes enriched in functions related to cell cycle progression and regulation and that genes under differential regulation by the activator and repressor E2Fs during maturation of female and male gonads are regulated inversely, which likely is behind the alternative use of cell cycle modes in these tissues.

In this work, ChIP-seq of both E2F TFs in 4 developmental stages of *Oikopleura* was performed. This will allow the construction of a more resolved map of E2F binding and thus, more informative of the differential gene regulation between cell cycle modes. Validation of identified TF binding sites by ChIP-qPCR and functional characterisation will show if E2F binding is the major determinant for gene expression regulation or if other co-regulator also play a role.

On a much less studied field, we set ourselves to explore the influence of a particular histone modification (H3K79me) on the regulation of cell cycle. Our results agreed with previous showing that H3K79me accumulates with time and give strength to the hypothesis that it may act as a timer for a cell's age. Through functional studies on the H3K79me methyltransferase, Dot1, we showed that H3K79me is likely to have an important role during mitotic cell cycles and that inhibition of Dot1 during endocycles causes major phenotypic defects by increasing proliferation of somatic tissues and reducing it in the gonad. Furthermore, we discovered an indirect negative feedback loop between two histone modifications (H3K79me and H2BK123ub1), something likely never described.

Future work would imply the validation of Dot1 inhibition results through alternative methods and more thorough dissection of the mechanisms underlying the function of Dot1 and the various states of H3K79me.

This study also shows the viability and advantages of using *Oikopleura dioica* as a model system for the study of alternative cell cycle modes.

Bibliography

- [1] D. T. Fox and R. J. Duronio, “Endoreplication and polyploidy: insights into development and disease.,” *Development (Cambridge, England)*, vol. 140, pp. 3–12, Jan. 2013.
- [2] L. Comai, “The advantages and disadvantages of being polyploid.,” *Nature reviews. Genetics*, vol. 6, pp. 836–46, Nov. 2005.
- [3] B. a. Edgar, N. Zielke, and C. Gutierrez, “Endocycles: a recurrent evolutionary innovation for post-mitotic cell growth,” *Nature Reviews Molecular Cell Biology*, vol. 15, pp. 197–210, Feb. 2014.
- [4] R. A. Sclafani and T. M. Holzen, “Cell cycle regulation of DNA replication.,” *Annual review of genetics*, vol. 41, pp. 237–80, Jan. 2007.
- [5] D. A. Orlando, C. Y. Lin, A. Bernard, J. Y. Wang, J. E. S. Socolar, E. S. Iversen, A. J. Hartemink, and S. B. Haase, “Global control of cell-cycle transcription by coupled CDK and network oscillators,” *Nature*, vol. 453, pp. 944–947, May 2008.
- [6] J. M. Trimarchi and J. a. Lees, “Sibling rivalry in the E2F family.,” *Nature reviews. Molecular cell biology*, vol. 3, pp. 11–20, Jan. 2002.
- [7] S. M. Frisch and J. S. Mymryk, “Adenovirus-5 E1A: paradox and paradigm.,” *Nature reviews. Molecular cell biology*, vol. 3, pp. 441–52, June 2002.
- [8] A. Besson, S. F. Dowdy, and J. M. Roberts, “CDK inhibitors: cell cycle regulators and beyond.,” *Developmental cell*, vol. 14, pp. 159–69, Feb. 2008.
- [9] N. Zielke, K. J. Kim, V. Tran, S. T. Shibusaki, M.-J. Bravo, S. Nagarajan, M. van Straaten, B. Woods, G. von Dassow, C. Rottig, C. F. Lehner, S. S. Grewal, R. J. Duronio, and B. a. Edgar, “Control of *Drosophila* endocycles by E2F and CRL4(CDT2).,” *Nature*, vol. 480, pp. 123–7, Dec. 2011.
- [10] Z. Ullah, C. Y. Lee, M. a. Lilly, and M. L. DePamphilis, “Developmentally programmed endoreduplication in animals.,” *Cell cycle (Georgetown, Tex.)*, vol. 8, pp. 1501–9, May 2009.
- [11] M. a. Lilly and R. J. Duronio, “New insights into cell cycle control from the *Drosophila* endocycle.,” *Oncogene*, vol. 24, pp. 2765–75, Apr. 2005.
- [12] A. Eliades, N. Papadantonakis, and K. Ravid, “New roles for cyclin E in megakaryocytic polyploidization.,” *The Journal of biological chemistry*, vol. 285, pp. 18909–17, June 2010.

- [13] K. Narbonne-Reveau, S. Senger, M. Pal, A. Herr, H. E. Richardson, M. Asano, P. Deak, and M. a. Lilly, “APC/CFzr/Cdh1 promotes cell cycle progression during the *Drosophila* endocycle.,” *Development (Cambridge, England)*, vol. 135, pp. 1451–61, Apr. 2008.
- [14] L. Weng, C. Zhu, J. Xu, and W. Du, “Critical role of active repression by E2F and Rb proteins in endoreplication during *Drosophila* development.,” *The EMBO journal*, vol. 22, pp. 3865–75, Aug. 2003.
- [15] A. Weiss, A. Herzig, H. Jacobs, and C. Lehner, “Continuous Cyclin E expression inhibits progression through endoreduplication cycles in *Drosophila*.,” *Current biology*, pp. 239–243, 1998.
- [16] S. van den Heuvel and N. J. Dyson, “Conserved functions of the pRB and E2F families.,” *Nature reviews. Molecular cell biology*, vol. 9, pp. 713–24, Sept. 2008.
- [17] N.-S. Moon and N. Dyson, “E2F7 and E2F8 keep the E2F family in balance.,” *Developmental cell*, vol. 14, pp. 1–3, Jan. 2008.
- [18] J. Li, C. Ran, E. Li, F. Gordon, G. Comstock, H. Siddiqui, W. Cleghorn, H.-Z. Chen, K. Kornacker, C.-G. Liu, S. K. Pandit, M. Khanizadeh, M. Weinstein, G. Leone, and A. de Bruin, “Synergistic function of E2F7 and E2F8 is essential for cell survival and embryonic development.,” *Developmental cell*, vol. 14, pp. 62–75, Jan. 2008.
- [19] H.-Z. Chen, M. M. Ouseph, J. Li, T. Pécot, V. Chokshi, L. Kent, S. Bae, M. Byrne, C. Duran, G. Comstock, P. Trikha, M. Mair, S. Senapati, C. K. Martin, S. Gandhi, N. Wilson, B. Liu, Y.-W. Huang, J. C. Thompson, S. Raman, S. Singh, M. Leone, R. Machiraju, K. Huang, X. Mo, S. Fernandez, I. Kalaszczynska, D. J. Wolgemuth, P. Sicinski, T. Huang, V. Jin, and G. Leone, “Canonical and atypical E2Fs regulate the mammalian endocycle.,” *Nature cell biology*, vol. 14, pp. 1192–202, Nov. 2012.
- [20] B. Westendorp, M. Mokry, M. J. a. Groot Koerkamp, F. C. P. Holstege, E. Cuppen, and A. de Bruin, “E2F7 represses a network of oscillating cell cycle genes to control S-phase progression.,” *Nucleic acids research*, vol. 40, pp. 3511–23, Apr. 2012.
- [21] C. Havens and J. Walter, “Mechanism of CRL4Cdt2, a PCNA-dependent E3 ubiquitin ligase,” *Genes & development*, vol. 7, pp. 1568–1582, 2011.
- [22] S. T. Shibutani, A. F. a. de la Cruz, V. Tran, W. J. Turbyfill, T. Reis, B. a. Edgar, and R. J. Duronio, “Intrinsic negative cell cycle regulation provided by PIP box- and Cul4Cdt2-mediated destruction of E2f1 during S phase.,” *Developmental cell*, vol. 15, pp. 890–900, Dec. 2008.
- [23] H. Hochegger, D. Dejsuphong, E. Sonoda, A. Saberi, E. Rajendra, J. Kirk, T. Hunt, and S. Takeda, “An essential role for Cdk1 in S phase control is revealed via chemical genetics in vertebrate cells.,” *The Journal of cell biology*, vol. 178, pp. 257–68, July 2007.
- [24] J. H. Meserve and R. J. Duronio, “Atypical E2Fs drive atypical cell cycles.,” *Nature cell biology*, vol. 14, pp. 1124–5, Nov. 2012.
- [25] B. Strahl and C. Allis, “The language of covalent histone modifications,” *Nature*, vol. 403, no. January, pp. 41–45, 2000.

Bibliography

- [26] T. Kouzarides, “Chromatin modifications and their function.,” *Cell*, vol. 128, pp. 693–705, Feb. 2007.
- [27] A. J. Bannister and T. Kouzarides, “Regulation of chromatin by histone modifications.,” *Cell research*, vol. 21, pp. 381–95, Mar. 2011.
- [28] C. Raynaud, A. C. Mallory, D. Latrasse, T. Jégu, Q. Bruggeman, M. Delarue, C. Bergounioux, and M. Benhamed, “Chromatin meets the cell cycle.,” *Journal of experimental botany*, Feb. 2014.
- [29] K. M. Trujillo and M. A. Osley, “A role for H2B ubiquitylation in DNA replication.,” *Molecular cell*, vol. 48, pp. 734–46, Dec. 2012.
- [30] F. Frederiks, M. Tzouros, G. Oudgenoeg, T. van Welsem, M. Fornerod, J. Krijgsveld, and F. van Leeuwen, “Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states.,” *Nature structural & molecular biology*, vol. 15, pp. 550–7, June 2008.
- [31] D. De Vos, F. Frederiks, M. Terweij, T. van Welsem, K. F. Verzijlbergen, E. Iachina, E. L. de Graaf, a. F. M. Altelaar, G. Oudgenoeg, A. J. R. Heck, J. Krijgsveld, B. M. Bakker, and F. van Leeuwen, “Progressive methylation of ageing histones by Dot1 functions as a timer.,” *EMBO reports*, vol. 12, pp. 956–62, Sept. 2011.
- [32] A. Gassen, D. Brechtefeld, N. Schandry, J. M. Arteaga-Salas, L. Israel, A. Imhof, and C. J. Janzen, “DOT1A-dependent H3K76 methylation is required for replication regulation in *Trypanosoma brucei*.,” *Nucleic acids research*, vol. 40, pp. 10302–11, Nov. 2012.
- [33] H. Fu, A. K. Maunakea, M. M. Martin, L. Huang, Y. Zhang, M. Ryan, R. Kim, C. M. Lin, K. Zhao, and M. I. Aladjem, “Methylation of histone H3 on lysine 79 associates with a group of replication origins and helps limit DNA replication once per cell cycle.,” *PLoS genetics*, vol. 9, p. e1003542, June 2013.
- [34] F. Delsuc, H. Brinkmann, D. Chourrout, and H. Philippe, “Tunicates and not cephalochordates are the closest living relatives of vertebrates.,” *Nature*, vol. 439, pp. 965–8, Mar. 2006.
- [35] J.-M. Bouquet, E. Spriet, C. Troedsson, H. Otterå, D. Chourrout, and E. M. Thompson, “Culture optimization for the emergent zooplanktonic model organism *Oikopleura dioica*.,” *Journal of plankton research*, vol. 31, pp. 359–370, Apr. 2009.
- [36] S. Fujii, T. Nishio, and H. Nishida, “Cleavage pattern, gastrulation, and neurulation in the appendicularian, *Oikopleura dioica*.,” *Development genes and evolution*, vol. 218, pp. 69–79, Feb. 2008.
- [37] P. Ganot and E. M. Thompson, “Patterning through Differential Endoreduplication in Epithelial Organogenesis of the Chordate, *Oikopleura dioica*.,” *Developmental Biology*, vol. 252, pp. 59–71, Dec. 2002.
- [38] F. Denoëud, S. Henriët, S. Mungpakdee, J.-M. Aury, C. Da Silva, H. Brinkmann, J. Mikhaleva, L. C. Olsen, C. Jubin, C. Cañestro, J.-M. Bouquet, G. Danks, J. Poulain,

- C. Campsteijn, M. Adamski, I. Cross, F. Yadetie, M. Muffato, A. Louis, S. Butcher, G. Tsagkogeorga, A. Konrad, S. Singh, M. F. Jensen, E. H. Cong, H. Eikeseth-Otteraa, B. Noel, V. Anthouard, B. M. Porcel, R. Kachouri-Lafond, A. Nishino, M. Ugolini, P. Chourrout, H. Nishida, R. Aasland, S. Huzurbazar, E. Westhof, F. Delsuc, H. Lehrach, R. Reinhardt, J. Weissenbach, S. W. Roy, F. Artiguenave, J. H. Postlethwait, J. R. Manak, E. M. Thompson, O. Jaillon, L. Du Pasquier, P. Boudinot, D. a. Liberles, J.-N. Volff, H. Philippe, B. Lenhard, H. Roest Crollius, P. Wincker, and D. Chourrout, “Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate.,” *Science (New York, N.Y.)*, vol. 330, pp. 1381–5, Dec. 2010.
- [39] A. Moosmann, C. Campsteijn, P. W. Jansen, C. Nasrallah, M. Raasholm, H. G. Stunnenberg, and E. M. Thompson, “Histone variant innovation in a rapidly evolving chordate lineage.,” *BMC evolutionary biology*, vol. 11, p. 208, Jan. 2011.
- [40] A. Schulmeister, M. Schmid, and E. M. Thompson, “Phosphorylation of the histone H3.3 variant in mitosis and meiosis of the urochordate *Oikopleura dioica*.,” *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, vol. 15, pp. 189–201, Jan. 2007.
- [41] P. Ganot, A. Moosmann-Schulmeister, and E. M. Thompson, “Oocyte selection is concurrent with meiosis resumption in the coenocystic oogenesis of *Oikopleura*.,” *Developmental biology*, vol. 324, pp. 266–76, Dec. 2008.
- [42] C. Campsteijn, J. I. Ovrebo, B. r. O. Karlsen, and E. M. Thompson, “Expansion of cyclin D and CDK1 paralogs in *Oikopleura dioica*, a chordate employing diverse cell cycle variants.,” *Molecular biology and evolution*, vol. 29, pp. 487–502, Feb. 2012.
- [43] C. J. Kenyon, “The genetics of ageing.,” *Nature*, vol. 464, pp. 504–12, Mar. 2010.
- [44] G. Subramaniam, C. Campsteijn, and E. M. Thompson, “Lifespan extension in a semelparous chordate occurs via developmental growth arrest just prior to meiotic entry.,” *PloS one*, vol. 9, p. e93787, Jan. 2014.
- [45] A. Rendeiro, P. Navratilova, and E. Thompson, “Chromatin preparation for ChIP-seq in *Oikopleura dioica*.,” *figshare* <http://dx.doi.org/10.6084/m9.figshare.884562>, 2014.
- [46] G. Danks, C. Campsteijn, M. Parida, S. Butcher, H. Doddapaneni, B. Fu, R. Petrin, R. Metpally, B. Lenhard, P. Wincker, D. Chourrout, E. M. Thompson, and J. R. Manak, “OikoBase: a genomics and developmental transcriptomics resource for the urochordate *Oikopleura dioica*.,” *Nucleic acids research*, vol. 41, pp. D845–53, Jan. 2013.
- [47] J. Toedling, O. Skylar, O. Sklyar, T. Krueger, J. J. Fischer, S. Sperling, and W. Huber, “Ringo—an R/Bioconductor package for analyzing ChIP-chip readouts.,” *BMC bioinformatics*, vol. 8, p. 221, Jan. 2007.
- [48] A. R. Quinlan and I. M. Hall, “BEDTools: a flexible suite of utilities for comparing genomic features.,” *Bioinformatics (Oxford, England)*, vol. 26, pp. 841–2, Mar. 2010.
- [49] P. Machanick and T. L. Bailey, “MEME-ChIP: motif analysis of large DNA datasets.,” *Bioinformatics (Oxford, England)*, vol. 27, pp. 1696–7, June 2011.

Bibliography

- [50] S. Falcon and R. Gentleman, “Using GOstats to test gene lists for GO term association.,” *Bioinformatics (Oxford, England)*, vol. 23, pp. 257–8, Jan. 2007.
- [51] F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc, “REVIGO summarizes and visualizes long lists of gene ontology terms.,” *PLoS one*, vol. 6, p. e21800, Jan. 2011.
- [52] M. Schwaiger, A. Schönauer, A. F. Rendeiro, C. Pribitzer, A. Schauer, A. F. Gilles, J. B. Schinko, E. Renfer, D. Fredman, and U. Technau, “Evolutionary conservation of the eumetazoan gene regulatory landscape,” *Genome Research*, vol. 24, pp. 639–650, Mar. 2014.
- [53] A. S. Evgeny Z. Kvon, Tomas Kazmar, Gerald Stampfel, J. Omar Yanez-Cuna, Michaela Pagani, Katharina Schernhuber, Barry J. Dickson, “Genome-scale functional characterization of Drosophila developmental enhancers in vivo,” *Nature*, 2014.
- [54] W. Yu, E. J. Chory, A. K. Wernimont, W. Tempel, A. Scopton, A. Federation, J. J. Marineau, J. Qi, D. Barsyte-Lovejoy, J. Yi, R. Marcellus, R. E. Iacob, J. R. Engen, C. Griffin, A. Aman, E. Wienholds, F. Li, J. Pineda, G. Estiu, T. Shatseva, T. Hajian, R. Al-Awar, J. E. Dick, M. Vedadi, P. J. Brown, C. H. Arrowsmith, J. E. Bradner, and M. Schapira, “Catalytic site remodelling of the DOT1L methyltransferase by selective inhibitors.,” *Nature communications*, vol. 3, p. 1288, Jan. 2012.
- [55] A. Stergachis, E. Haugen, A. Shafer, and W. Fu, “Exonic Transcription Factor Binding Directs Codon Choice and Affects Protein Evolution,” *Science*, vol. 1367, no. 2013, 2013.
- [56] L. T. Macneil and A. J. M. Walhout, “Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression.,” *Genome research*, vol. 21, pp. 645–57, May 2011.
- [57] T. T. Onder, N. Kara, A. Cherry, A. U. Sinha, N. Zhu, K. M. Bernt, P. Cahan, B. O. Marcacci, J. Unternaehrer, P. B. Gupta, E. S. Lander, S. a. Armstrong, and G. Q. Daley, “Chromatin-modifying enzymes as modulators of reprogramming.,” *Nature*, vol. 483, pp. 598–602, Mar. 2012.

Appendix A

Gene Ontology terms associated with genes putatively regulated by E2F factors

A.1 Genes bound only by E2F1 in ovary tissue

Term ID	Description	$\log_{10}(\text{p-value})$
GO:0006278	RNA-dependent DNA replication	-2.71
GO:0009987	cellular process	-8.9625
GO:0016192	vesicle-mediated transport	-3.1982
GO:0033002	muscle cell proliferation	-4.1117
GO:0050896	response to stimulus	-2.915
GO:0065007	biological regulation	-4.2604
GO:0071840	cellular component organization or biogenesis	-2.5529
GO:0044237	cellular metabolic process	-7.5965
GO:0009058	biosynthetic process	-5.9322
GO:0006006	glucose metabolic process	-2.4011
GO:0006807	nitrogen compound metabolic process	-6.7505
GO:0043170	macromolecule metabolic process	-3.861
GO:0044238	primary metabolic process	-3.0041
GO:0060059	embryonic retina morphogenesis in camera-type eye	-3.0939
GO:0035556	intracellular signal transduction	-3.7117
GO:0022406	membrane docking	-2.3702
GO:0030029	actin filament-based process	-2.8771
GO:0044260	cellular macromolecule metabolic process	-7.0356
GO:0090317	negative regulation of intracellular protein transport	-3.0939
GO:0007010	cytoskeleton organization	-3.2969
GO:0006139	nucleobase-containing compound metabolic process	-6.8149
GO:0044249	cellular biosynthetic process	-6.2634
GO:0015812	gamma-aminobutyric acid transport	-2.2094
GO:0051338	regulation of transferase activity	-2.0189
GO:0032879	regulation of localization	-2.4154
GO:0043412	macromolecule modification	-2.4469
GO:0033036	macromolecule localization	-3.1818
GO:0010604	positive regulation of macromolecule metabolic process	-3.4616
GO:0051641	cellular localization	-2.0583
GO:0007154	cell communication	-2.2118
GO:0006259	DNA metabolic process	-3.7403

Appendix A. Gene Ontology terms associated with genes putatively regulated by E2F factors

GO:0071495	cellular response to endogenous stimulus	-2.1562
GO:0032268	regulation of cellular protein metabolic process	-2.7219
GO:0009117	nucleotide metabolic process	-2.0027
GO:0010608	posttranscriptional regulation of gene expression	-2.0002
GO:0043062	extracellular structure organization	-2.1337
GO:0051246	regulation of protein metabolic process	-2.9956
GO:0010467	gene expression	-3.7199
GO:0006464	cellular protein modification process	-2.3409
GO:0034641	cellular nitrogen compound metabolic process	-7.079
GO:0016197	endosomal transport	-2.9234
GO:0006471	protein ADP-ribosylation	-2.1064
GO:0016070	RNA metabolic process	-2.4737
GO:0015074	DNA integration	-2.6682
GO:0034645	cellular macromolecule biosynthetic process	-5.6595
GO:0090304	nucleic acid metabolic process	-5.5239
GO:0009059	macromolecule biosynthetic process	-5.3032
GO:0008360	regulation of cell shape	-3.0133
GO:0045184	establishment of protein localization	-2.2365
GO:0035278	negative regulation of translation involved in gene silencing by miRNA	-2.2094
GO:0060255	regulation of macromolecule metabolic process	-5.9469
GO:0045974	regulation of translation, ncRNA-mediated	-2.2094
GO:0046777	protein autophosphorylation	-2.5191
GO:0060419	heart growth	-2.6516
GO:0044267	cellular protein metabolic process	-2.335
GO:0019637	organophosphate metabolic process	-2.0529
GO:0007264	small GTPase mediated signal transduction	-2.4515
GO:0048523	negative regulation of cellular process	-2.6714
GO:0051649	establishment of localization in cell	-2.267
GO:0006281	DNA repair	-3.1591
GO:0006904	vesicle docking involved in exocytosis	-2.4221
GO:0051716	cellular response to stimulus	-4.0826
GO:0009913	epidermal cell differentiation	-2.4019
GO:0006260	DNA replication	-2.9098
GO:0006888	ER to Golgi vesicle-mediated transport	-2.3636
GO:0016043	cellular component organization	-2.7253
GO:0006351	transcription, DNA-dependent	-2.6185
GO:0016567	protein ubiquitination	-2.1172
GO:0070647	protein modification by small protein conjugation or removal	-2.0504
GO:0006996	organelle organization	-2.0441

A.2 Genes bound only by E2F7 in ovary tissue

Term ID	Description	log ₁₀ (p-value)
GO:0001539	ciliary or flagellar motility	-4.1085
GO:0031667	response to nutrient levels	-2.2702
GO:0043094	cellular metabolic compound salvage	-3.3526
GO:0034199	activation of protein kinase A activity	-2.5983
GO:0006040	amino sugar metabolic process	-2.1516
GO:0043101	purine-containing compound salvage	-2.3941
GO:0048739	cardiac muscle fiber development	-2.3941
GO:0006278	RNA-dependent DNA replication	-2.2707
GO:0007091	metaphase/anaphase transition of mitotic cell cycle	-3.3526
GO:0006928	cellular component movement	-2.1772
GO:0072511	divalent inorganic cation transport	-3.6156
GO:0008361	regulation of cell size	-2.4811
GO:1901071	glucosamine-containing compound metabolic process	-2.0935

A.4. Genes bound only by E2F7 in testis tissue

GO:0006816	calcium ion transport	-3.8459
GO:0030001	metal ion transport	-2.0366
GO:0007512	adult heart development	-2.1734

A.3 Genes bound only by E2F1 in testis tissue

Term ID	Description	log ₁₀ (p-value)
GO:0001539	ciliary or flagellar motility	-3.2899
GO:0006260	DNA replication	-2.2387
GO:0048739	cardiac muscle fiber development	-2.4506
GO:0051298	centrosome duplication	-2.4506
GO:0090316	positive regulation of intracellular protein transport	-2.866
GO:0006816	calcium ion transport	-2.6053
GO:0006928	cellular component movement	-2.4741
GO:0006898	receptor-mediated endocytosis	-2.9111
GO:0006139	nucleobase-containing compound metabolic process	-2.1489
GO:0034220	ion transmembrane transport	-2.2164
GO:0045727	positive regulation of translation	-2.1629
GO:0072511	divalent inorganic cation transport	-2.1809
GO:0006278	RNA-dependent DNA replication	-2.1966
GO:0030838	positive regulation of actin filament polymerization	-2.1432
GO:0007091	metaphase/anaphase transition of mitotic cell cycle	-2.2368
GO:0007512	adult heart development	-2.2368

A.4 Genes bound only by E2F7 in testis tissue

Term ID	Description	log ₁₀ (p-value)
GO:0009653	anatomical structure morphogenesis	-3.5006
GO:0009894	regulation of catabolic process	-4.3726
GO:0009987	cellular process	-4.0159
GO:0023052	signaling	-4.9272
GO:0032501	multicellular organismal process	-2.1902
GO:0032502	developmental process	-2.4275
GO:0050896	response to stimulus	-2.9353
GO:0065007	biological regulation	-5.9131
GO:0033692	cellular polysaccharide biosynthetic process	-3.1836
GO:0007020	microtubule nucleation	-2.6003
GO:0006807	nitrogen compound metabolic process	-3.3346
GO:0006112	energy reserve metabolic process	-2.7232
GO:1901071	glucosamine-containing compound metabolic process	-3.0878
GO:0044238	primary metabolic process	-3.023
GO:0043170	macromolecule metabolic process	-3.7098
GO:0050905	neuromuscular process	-2.1167
GO:0007141	male meiosis I	-2.0799
GO:0043112	receptor metabolic process	-3.1582
GO:0006913	nucleocytoplasmic transport	-2.5909
GO:0006139	nucleobase-containing compound metabolic process	-3.4415
GO:0044260	cellular macromolecule metabolic process	-2.6489
GO:0006366	transcription from RNA polymerase II promoter	-2.0226
GO:0009966	regulation of signal transduction	-3.7948
GO:0051641	cellular localization	-2.093
GO:0034329	cell junction assembly	-2.1167
GO:0015698	inorganic anion transport	-2.0856
GO:1900542	regulation of purine nucleotide metabolic process	-2.6926

Appendix A. Gene Ontology terms associated with genes putatively regulated by E2F factors

GO:0010629	negative regulation of gene expression	-2.39
GO:0007154	cell communication	-5.007
GO:0031329	regulation of cellular catabolic process	-3.8914
GO:0040029	regulation of gene expression, epigenetic	-2.0888
GO:0006022	aminoglycan metabolic process	-2.9965
GO:0006357	regulation of transcription from RNA polymerase II promoter	-2.2083
GO:0010604	positive regulation of macromolecule metabolic process	-2.1272
GO:0048583	regulation of response to stimulus	-2.5882
GO:0006040	amino sugar metabolic process	-2.6847
GO:0071310	cellular response to organic substance	-2.688
GO:0050794	regulation of cellular process	-6.1247
GO:0007167	enzyme linked receptor protein signaling pathway	-2.006
GO:0016070	RNA metabolic process	-2.7768
GO:0006621	protein retention in ER lumen	-2.0799
GO:0034641	cellular nitrogen compound metabolic process	-3.3173
GO:0007264	small GTPase mediated signal transduction	-2.9759
GO:0023051	regulation of signaling	-3.0402
GO:0090304	nucleic acid metabolic process	-2.9958
GO:0001840	neural plate development	-2.6003
GO:0005976	polysaccharide metabolic process	-2.5276
GO:0021871	forebrain regionalization	-2.0799
GO:0006351	transcription, DNA-dependent	-2.1927
GO:0034446	substrate adhesion-dependent cell spreading	-2.0799
GO:0032288	myelin assembly	-2.0799
GO:0001708	cell fate specification	-2.6684
GO:0046661	male sex differentiation	-2.0622
GO:0005979	regulation of glycogen biosynthetic process	-2.8713
GO:0035556	intracellular signal transduction	-2.5872
GO:0051716	cellular response to stimulus	-4.8572
GO:0045165	cell fate commitment	-2.3637
GO:0001763	morphogenesis of a branching structure	-2.8468
GO:0010033	response to organic substance	-2.0303
GO:0007389	pattern specification process	-2.0547
GO:0007166	cell surface receptor signaling pathway	-2.0848
GO:0035239	tube morphogenesis	-2.7748

Appendix B

List of key cell cycle effectors putatively regulated by E2F factors

Genes bound by both TFs in ovary Rb, Geminin, Cdt1, CDK1b, CDK1c, CDK1e, CDK7, p21, CyclinB1, Cyclin Dc, Cyclin E, PCNA1, PCNA2, PCNA3, PCNA4

Genes bound by both TFs in testis DP, Rb, Cdt1, CDK1b, CDK1e, CyclinB1, CyclinB3, PCNA1, PCNA2, PCNA3, PCNA4

Genes bound only by E2F1 in ovary E2F1, DP, Cdh1, CDK1d, CDK1f, CDK2/3, CDK5, CDK9b, Cyclin A1, Cyclin A2, Cyclin C

Genes bound only by E2F7 in ovary none

Genes bound only by E2F1 in testis E2F7, Geminin, CDK1c, CDK5, p21, Cyclin Dc, Cyclin E

Genes bound only by E2F7 in testis CDK1d, CDK1f, Cyclin A2