Norwegian University of Life Sciences

Master's Thesis 2020 60 ECTS
Faculty of bioscience

Remodeling of H3K4me3 histone tail modifications across life stages in Atlantic salmon

Xxx. Naqing
Master of Animal breeding and genetic

Acknowledgements

In the process of writing this thesis, I have been supported by many people. I would like to sincerely thank all of you.

First, I would like to express my deepest gratitude to my supervisor, Associate professor Simen Sandve, for your constant support, guidance and motivation. It would never have been possible for me to take this work to completion without your incredible support. Second, I would like to extend my sincere gratitude to my co-supervisor Line Røsæg, provided me with help and encouragement and patience throughout the duration of this project.

Special thanks to Thomas Harvey for providing research results and patiently explained it to me. I'd also like to acknowledge the help of Gareth Gillard on bioinformatic, and Jin Yang for helping me with data visualization and helpful advice for thesis writing.

I appreciate the company of everyone in Sandve-lab group. I also had great pleasure of working with MO, Mathilde and Mariann and Tan in CIGENE lab, who never wavered in their support during my experiment. Thanks for Orion cluster supported my data analysis!

Last but not least, I would thank my parents for supporting me from afar, and my partner and my friends, for always supporting me mentally and brought me happiness.

Abstract

Atlantic salmon undergo a series of preparatory adaptation for a life in the sea, referred to as smoltification. Recent studies have focused on the genetic base of this process and revealed genome-wide transcriptional remodeling related to smoltification and seawater-transfer in different tissues, especially in the liver (Harvey, 2019; Gillard et al., 2018). However, we know very little about the mechanisms that drive the dynamics of genome regulation during smoltification.

A factor that controls the transcriptional level gene expression is Epigenetics, Epigenetics is a common term for various chemical alterations of the DNA molecule or chemical modifications to the histone tail, which is called histone tail modification. one of the well-studied histone tail modification H3K4me3 is reported to associate with active promoter. In this project, we aim to interpret this histone tail modification remodeling of gene expression during different life-stages of salmon.

We generate ChIP-seq data for H3K4me3, profiled the life-stage associated landscape of H3K4me3 in Atlantic salmon liver by using H3K4me specific antibody. 14 fish (4 parr, 6 smolt, and 4 seawater) across three life stages (parr, smolt, and seawater) were sampled.

We found H3K4me3 signals in all life stages are enriched in the promoter and 5UTR of gene region. The life stage-specific genes with the H3K4me3 signal reflecting the physiological characteristics of that stage and associated with smolt liver function and development. By linking the H3K4me3-marked genes with life-stage associated dynamic co-expressed gene cluster from previous study (Harvey, 2019), we find the histone code is weakly correlated with differential gene expression.

In conclusion, the analyses of H3K4me3-signals and gene expression fits with a model where some genes are regulated through histone tails remodeling during parr-smolt transformation, but that this level of genome regulation does not play the major role in the developmental transition from a parr to a smolt.

Unfortunately, we failed to produce high-quality ChIP-seq data from biological replicates. Hence, data analysis based on genome-wide patterns of H3K4me3-signals from pooled replicates, which only allowed us to assess the H3K4me3 signals in a qualitative way.

Abstrakt

Atlanterhavslaks gjennomgår en serie forberedende tilpasninger for et liv i havet, kalt smoltifisering. Nyere studier har fokusert på den genetiske basisen i denne prosessen og avdekket omfattende transkripsjonelle endringer relatert til smoltifisering og overgang fra ferskvann til saltvann i forskjellige vev, spesielt i leveren (Harvey, 2019; Gillard et al., 2018). Vi vet imidlertid veldig lite om mekanismene som driver dynamikken i genreguleringen under smoltifisering.

En faktor som kontrollerer genekspresjonen på transkripsjonelt nivå er epigenetikk. Epigenetikk er en vanlig betegnelse for forskjellige kjemiske endringer av DNA-molekylet eller kjemiske modifikasjoner av histonhalene. En av de godt studerte histonhalemodifiseringene, H3K4me3, assosierer med aktive promoterer. I dette prosjektet tar vi sikte på å tolke denne histonhalemodifiseringen som remodellering av genuttrykk i løpet av laksens forskjellige livsfaser.

Vi har generert ChIP-seq data for H3K4me3, beskrevet det livsfase-assosierte H3K4me3-landskapet i lever ved å bruke H3K4me-spesifikt antistoff. Det ble tatt prøver av 14 fisk (4 parr, 6 smolt og 4 saltvann) over tre livsfaser (parr, smolt og saltvann).

Resultatene våre viser at H3K4me3-signaler er beriket i promotoren og 5UTR i genregionen i alle de tre livsfasene til Atlanterhavslaks. Livsfasespesifikke gener med H3K4me3-signalet gjenspeiler de fysiologiske egenskapene i en bestemt livsfase og er assosiert med leverfunksjon hos smolt og utvikling. Ved å koble H3K4me3-merkede gener med livsfase-assosiert genklynger med samme uttrykksmønster fra en tidligere studie (Harvey, 2019), finner vi at histonkoden er svakt korrelert med differensielt genuttrykk.

Avslutningsvis passer analysene av H3K4me3-signaler og genuttrykk til en modell der noen gener reguleres gjennom ombygging av histonhaler under parr-smolt utviklingen. Dette regulatoriske nivået spiller allikevel ikke den viktigste rollen i utviklingen fra parr til smolt.

Dessverre klarte vi ikke å produsere ChIP-seq data av høy kvalitet fra biologiske replikater. Dermed ble dataanalysen basert på mønstre av H3K4me3-signaler på tvers av hele genomet fra sammenslåtte replikater, noe som bare tillot oss å vurdere H3K4me3-signalene kvalitativt.

Table of Contents

Introduction

Atlantic salmon life cycle

Atlantic salmon (*Salmo salar)* is an anadromous fish. As such, they spend the early life stage at fresh water, travel to sea water after 1-3 years, come back as adult fish to their hatched place to spawn (Figure1). After hatching Atlantic salmon feed on yolk sac attached to their body which can provide nutrition for several weeks or month. When the yolk is completely absorbed, the fish start preying, referred to as first feeding. At this stage the fish are called fry. Fry continue to grow into parr in fresh water. Parr-smolt transformation occurs in late spring, when parr go through a series of preparatory adaptations for a life in the sea. This process is called smoltification. Smolt then migrate to open sea, where they switch to a marine diet and undergo rapid growth. (Marine institute, salmon life cycle)

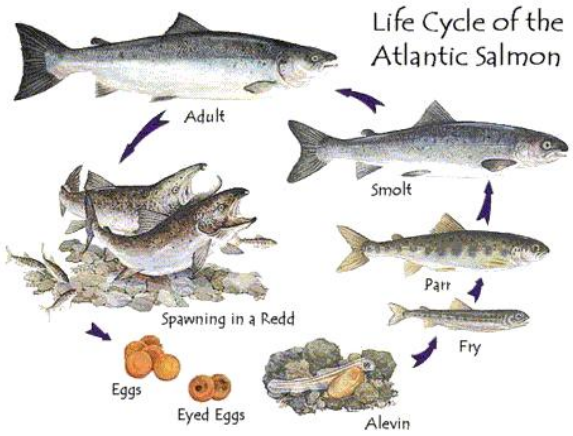


Figure1. Life cycle of Atlantic salmon. Adapted from Atlantic salmon federation. New hatched salmon call Alevin, has egg yolk attached on their body, provide all the needed nutrition. Salmon at Fry and Parr stage still spend in fresh water, migrate to sea after turn into Smolt.

Smolt development

The Parr-smolt transformation, also called smoltification, includes a series of simultaneous but independent transformation processes involving changes in morphology, behavior and physiology.

Morphology

The most intuitive difference between smolt and parr is their morphological characteristics (Figure 2) (reviewed by McCormick, 2012). Except obvious size difference, parr has vertical bands and dark spots at the sides of the body. The bands and spots fade during smoltification and are replaced by silvery skin scales, and dark fin margin. In addition to the color, the body shape changes too. Smolts have slimmer bodies than parr due to linear growth being faster than the mass growth. McCormick & Saunders (1987) suggest such a reduction of weight-to-length ratio may be related to a decrease in body lipid content. This decrease may stem from increased activity and the energetic demands during smolt development. Also, this streamlined

body shape may allow greater swimming capacity, facilitate smolt avoid predator along the way travel to sea (McCormick, 2012).

Figure 2. Morphological differences between Atlantic salmon Parr (top) and smolt reared in the wild. Parr has vertical bands and spots on the sides. Smolt has intense silvering and darkened caudal, pectoral, and dorsal fin margins. by S.D. McCormick.

Behavior

Smolt also develop migration-related behavioral changes. Parr are territorial and live near the stream bottom, whereas smolts decrease aggressive behavior, tend to form aggregates and swim at mid-depths. This swimming habit ensures smolt to follow the water flow downstream, escape the river, and start a life in the sea (reviewed by Wedemeyer, Saunders, & Clarke, 1980).

Physiology

The transition to smolt is associated with several large physiological changes. Metabolism elevates about 30% in smolt and the oxidation consumption increases because of increased catabolism of fat, carbohydrate and protein (Sheridan, 1989). Glycogen and lipid reserves are reduced, not only due to increase of catabolism but also because of a decrease in the synthesis of these molecules (Gillard et al. 2018). Last and most importantly, development of hypo-osmoregulatory ability, the gill physiology unergo a series of pre-seawater adaptations to prepare for the high salinity environment (Seera et al, 2010). All these physiological changes are making juvenile fish best adapt to marine life, and the period when the fish is best prepared for sea water migration is called “smolt window”. Fish that remain in fresh water after this time period will lose several smolt features, refer to as de-smoltification. (McCormick, 2012)

Regulation of smoltification

The timing and propensity of smoltification are control by both endogenous and environmental factors (McCormick, Shrimpton, Moriyama, & Björnsson, 2007). The most important environmental factors that trigger initiation of smolt development is photoperiod (Hoar, 1988) and temperature (ZYDLEWSKI, 1997). Several studies have reviewed the interplay between

endogenous and exogenous (i.e. temperature and photoperiod) control of smoltification is mediated by the neuroendocrine system. Growth hormones not only control growth and metabolism, but also control salinity tolerance by interacting with cortisol. Thyroid hormones also have a role in gill physiology development, morphological changes and metabolism, and possibly effecting behavior and prolactin is generally thought to be inhibitory to most aspects of smolt development (Lorgen et al. 2015; McCormick, 2012). However, the molecular events that underlie these regulatory processes is poorly understood.

Many studies have revealed gene regulatory mechanism underlying the parr-smolt transformation. Clarke et al. (1994) crossed Chinook salmon (Oncorhynchus tshawytscha) between different timing of smoltification and found that the early smolting is dominant to late smolting. While Foote et al. (1992) cross breed the anadromous and non-anadromous sockeye salmon (Oncorhynchus nerka) suggests that in these species, the propensity of smoltification is under additive gene control, rather than dominant genetic control. More recently, Nichols et al. (2008) used quantitative trait loci analyses, identified a particular region in genome is associated with multiple smoltification related traits including growth, morphology as well as osmoregulatory enzymes in O. mykiss.

Smoltification associated transcriptional regulation

To date, many genome-wide analyses were conducted to the transcriptional level regulation of smoltification. Seear et al. (2010) investigated how gene expression alters during smoltification using cDNA microarray, reveals that genes up and down regulated in gill, brain, kidney during smoltification is reflecting the physiological and biochemical observations of salmonids during smoltification, including transcriptional regulation related to growth, metabolism, oxygen transport, and osmoregulation. Robertson & McCormick(2012) used the GRASP 16K microarray investigated the transcriptional changes during smolt development in five tissues (gill, liver, olfactory rosettes, pituitary, and hypothalamus) reveal that smolts had higher mRNA abundances for genes involved in the regulation of transcription, protein biosynthesis than parr, and lower mRNA levels for genes involved in proteolysis. A more recent study has found several genes with smoltification-associated gill regulation, which is dependent on the fish having experienced a winter photoperiod (Iversen et al. 2020).

Smoltification associated transcriptional regulation in liver

As a major organ in vertebrate, Liver plays an important role in organism life, especially related to metabolism, energy production, as well as nutrition storage, especially lipid metabolism. Lipid is predominant resource of energy for fish metabolism, development, and growth (Carmona-Antoñanzas et al., 2014). As such liver play a key role in smoltification. Decades' study has revealed, during salmon travel from fresh water to seawater, the liver undergoes great alteration to better adapt to environment change. For example, change in availability of essential long-chain polyunsaturated n-3 fatty acids (LC-n-3 PUFA), which is low in fresh water and higher ins sweater. liver is known to capable to regulate endogenous lipid synthesis based on dietary availability (Leaver et al., 2008). Carmona-Antoñanzas et al., (2014). studied the role of key transcription factors (TF) in the transcriptional regulation of lipid metabolism in salmon by transfection and over expression of TFs. They found that genes of biosynthesis LC-PUFA (elovl and fads2) and cholesterol metabolism (abca1) are regulated by Lxr and Srebp TFs in salmon, and the srebp1 and srebp2 mRNA respond to low LC-PUFA diet. Thus, conclude that Atlantic salmon adjust lipid metabolism in response to dietary lipid composition through the transcriptional regulation of gene expression. Feeding experiment that study the transcriptional regulation of lipid metabolism in Atlantic salmon liver and gut and in fresh and saltwater shows that liver become less respond to diet change after transfer to saltwater due to gene expression relate to lipogenesis and lipid transfer decreasing, but the lipid intake in gut is increased (Gillard et al. 2018). It was also report earlier by Sheridan (1989) lipid storage decrease due to not only increase of break down but also decrease in synthesis. Harvey (2019) confirmed this lipid reduction is a genome wide trend, furthermore, provide evidence that this trend occurs after smoltification while the smolt are still in fresh water, in his study also find that Epigenetic remodeling genes alter in expression during both smoltification and seawater transfer, indicating that epigenetic remodeling may take place in both life stages.

Transcription regulation and histone tail modifications

Transcription regulation

Transcription regulation rely on transcription factor (TF). Transcription factor is a protein that bind to specific sequence of a gene, it determines whether a gene is transcribed or not by controlling when, where, in what level the RNA polymerase act (Figure 3). RNA polymerase is the enzyme which actually perform the transcription (LaKna, 2017). TF either works alone or

cooperate with other transcription related protein control the RNA polymorase. The sequence on DNA for TF to bind be divided into promoter and enhancer, promoter is considered as the site where TF bind and transcription initiate, enhancer can act on the gene (either enhance or repress) from distance (Andersson and Sandelin, 2019). Transcription regulating mechanism have different pattern, in this thesis we will focused on epigenetic gene expression regulation, which control gene expression by adjusting the accessibility of transcription factors.

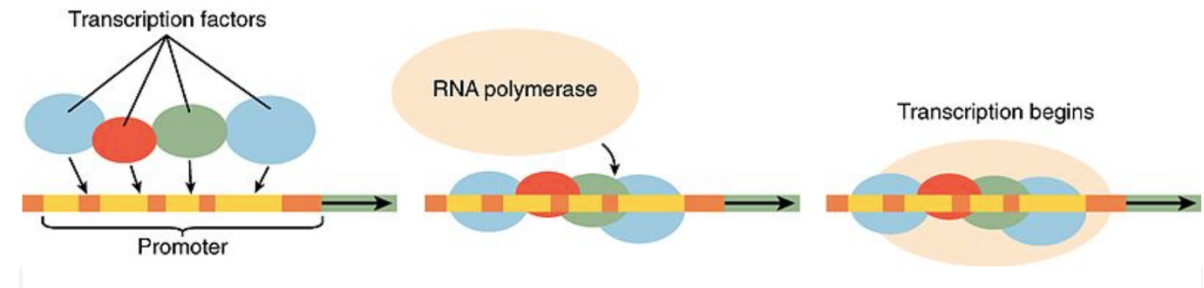


Figure 3. The role of transcription factor. Transcription factor control transcription by control RNA polymerase.

Epigenetic regulation of gene transcription

The original meaning of "epigenetics" used to denote the poorly understood processes by which a fertilized zygote developed into a mature, complex organism, now this term used for functionally relevant changes to the genome that result from environmental cue or related to general organism development that do not involve a change in the DNA sequence (reviewed by Felsenfeld, 2014). There are two major mechanism of epigenetic regulation of transcription. One is chemical modification of deoxycytidine residues of DNA where one methyl group is added to the DNA, known as DNA-methylation. Increasing DNA-methylation makes the DNA strand less accessible to transcription factor and RNA polymerase II. The second mechanism of epigenetic transcriptional regulation is chemical modification to the histone proteins which makes up the nucleosomes that DNA is wrapped around in the cell nucleus. (reviewed by Jaenisch and Bird, 2003)

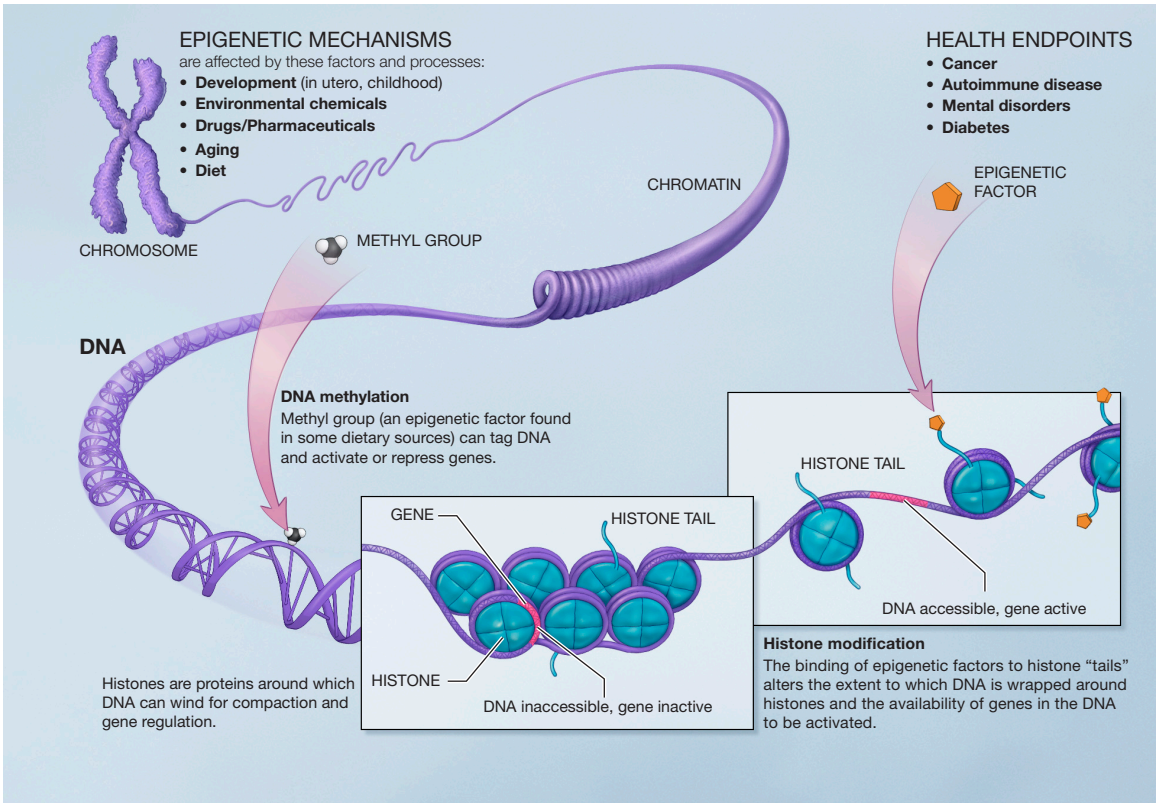


Figure 4. Schematic diagram of Epigenetics regulating gene expression. Tightly packed chromatin is heterochromatin, genes packed in heterochromatin is inactive. Euchromatin is less tightly packed, on which placed active genes. Epigenetic gene expression regulation is affected by environmental factor and organism development. Epigenetic can directly modify DNA strand by adding methyl group to DNA. Or modify the histone on the tail. Histone tail is protruding from the surface of the histone, modified by Epigenetic factors. Adapted from National Institutes of Health.

The role of histone tail modifications in Epigenetic gene regulation

Eukaryotic DNA is packaged into a structure called chromatin (Figure 4)., The backbone of this chromatin is called nucleosomes, which consists of four core histones (H2A, H2B, H3, and H4). 146 base pairs of DNA is wrapping around each nucleosome. The core histones have seemingly unstructured tail protrude from the surface of the chromatin, refer to as histone tail. The nucleosome structure and its tail modifications control gene expression by functioning as a dynamical switch between transcriptional active euchromatin and inactive heterochromatin (Huisinga, Brower-Toland, & Elgin, 2006). Studies have shown that the tails are involved in a range of intra- and inter-nucleosome contacts as well as contact with each other (du Preez & Patterton, 2013). This study also showed that the core histones tails, particularly that of H3 and H4 are influenced by post-translational modifications (du Preez & Patterton, 2013). There are a large number of posttranslational modifications that remodel the histone tail. Also, they can orchestrate the ordered recruitment of enzyme complexes to the DNA. In this way, histone

modifications have the potential to influence many fundamental biological processes (Kouzarides, 2007). To date, there are at least eight different classes that have been characterized and they have different sites functions to the chromatin. (Table 1)

Table 1. characterized histone modifications (Kouzarides, 2007)

Chromatin Modifications	Residues Modified	Functions Regulated
Acetylation	K-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	K-me1 K-me2 K-me3	Transcription, Repair
Methylation (arginines)	R-me1 R-me2a R-me2s	Transcription
Phosphorylation	S-ph T-ph	Transcription, Repair, Condensation
Ubiquitylation	K-ub	Transcription, Repair
Sumoylation	K-su	Transcription
ADP ribosylation	E-ar	Transcription
Deimination	R > Cit	Transcription
Proline Isomerization	P-cis > P-trans	Transcription

H3K4me3

Among the histone tail modifications, the best studied are methylation and acetylation of lysine (K) residues. In this project we focus on the methylation modification. Lysine residues can be mono- (me1), di- (me2), or tri-methylated (me3). Different marks on lysine residues are distribute on particular areas on genome and associated with distinct states of gene expression, and these patterns are generally conserved from yeast to humans (Gates, Foulds, & O'Malley, 2017). ENCODE project, (2007) have studied genome-wide histone tail modifications in variety of eukaryotes and demonstrate that tri-methylation at the 4th lysine residue of the histone H3 protein (H3K4me3) is positively associated with gene transcript levels and enriched near the transcriptional start sites (TSSs). In this thesis we aim to examine the role of H3K4me3 in the differential gene expression among different life stage of Atlantic salmon.

Chromatin immunoprecipitation followed by high throughput sequencing

Quantification of histone tail modifications

Chromatin immunoprecipitation is an assay to study the genome wide DNA and protein interaction. First revolutionized the genome wide understanding of DNA- protein interaction was chromatin immunoprecipitation followed by microarrays (ChIP-chip). However, following rapid development of next generation sequencing technology, chromatin immunoprecipitation combined with high throughput sequencing (ChIP-seq) replaced the Chip-chip to become the most commonly used tool to study histone tail modifications , as it

provides higher resolution, less background noise and greater coverage (Berger, 2007; Park, 2009).

Histone tail ChIP-seq: General workflow

Chromatin immunoprecipitation (ChIP) experiments enable us to enrich for DNA fragments associated with a particular protein or protein modification of interest. There are two main methods for ChIP. The most used method is called Cross-ChIP, where a cross linking reaction binds DNA to proteins prior to DNA sharing and isolation. The other method, used in this thesis, is called Native-ChIP. In this approach we take advantage of the naturally occurring binding of protein and DNA in the nucleus and the cross-linking step is not needed. The Cross-ChIP usually have higher ratio of signal-to-noise compared to Native-ChIP; however, it is also more reliant on thorough optimization of the protocol (i.e. cross-linking time and procedures). In the sections below I will briefly go through the 4 main steps of a ChIP-experiment (Figure 5)

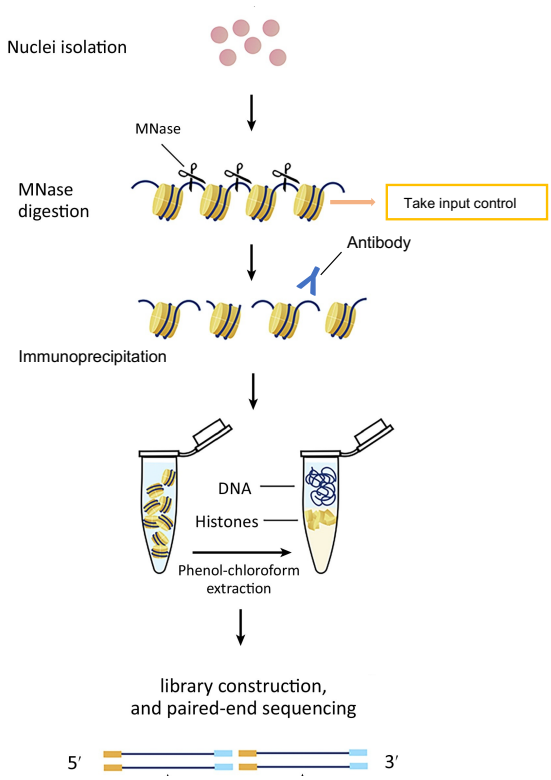


Figure5. Flow chart of ChIP-seq. First, nuclei were isolated from cell, and chromatin digested by micrococcal nuclease to nucleosomes, target nucleosome captured by Histone modification specific antibody. Input DNA are taken before conduct immunoprecipitation. The DNA from IP experiment are sent for sequencing. Finally, DNA eluted from histone, sent for pair-end sequencing. (Voong et al., 2017)

Isolation of nuclei

The first step is to isolate the nuclei. Micrococcal Nuclease (MNase) is an endo-exo nuclease, it preferably digests single strand DNA or RNA. by "endo" it means, micrococcal nuclease cut the DNA between nucleosomes, while DNA that bounding on histone remain intact. This make it a crucial tool for Native-ChIP. While it is also "exo" nuclease, because after finish cutting linker DNA, it starts to digest DNA strand wrapping on histone. This make it important to control digesting time when using Micrococcal produce mono-chromosome base, since over digestion will lead to shorten of target DNA.

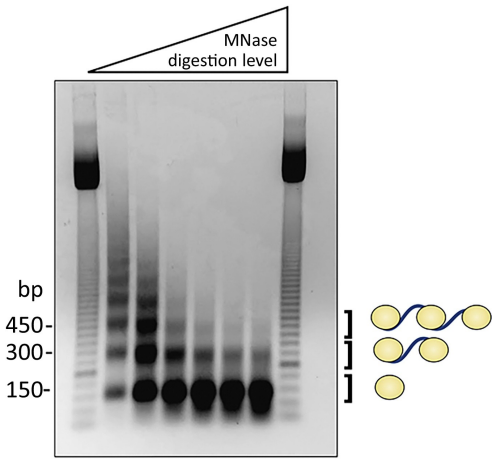


Figure 6. Distribution of chromatin fragments in the Electropherogram. Approximate locations of mono-, di-, and tri- nucleosomes are marked alongside.(Voong, Xi, Wang, & Wang, 2017)

Digestion of chromatin

The second step is to digest the chromatin into fragments with Micrococcal Nuclease. Mono-nucleosome is one histone and the DNA bound on it. It frees from chromatin when MNase digest the chromatin on linker DNA. A strand at approximately 147bp on Electrophoresis diagram is the standard for mono-nucleosome resolution of the chromatin digestion (Figure6). This is essential for downstream experiment, as our aim is histone tail modification.

Immunoprecipitation

Third, we need to use specific antibody against the protein of interest to retrieve the DNA-protein from solution, referred to as immunoprecipitation. Native ChIP experiment depend on the Antibody specifically capture the histone of interest and the bound DNA. This specificity directly decides to the success of the experiment. Dynabeads are superparamagnetic, monosized polymer particles, A wide range of bioreactive molecules can be adsorbed or coupled to the bead-surface and used in the separation of biological materials (cells, proteins, nucleic acids etc). ChIP experiment depend on dynabeads to separate the target histone-DNA complex that captured by antibody from background. While dynabeads sometimes bind to chromatin with absent of antibody, this binding is of course not specific, thus, lead to incorrect

result of ChIP experiment, so unspecific binding must be removed before conduct immunoprecipitation.

DNA elution and sequencing

Finally, target DNA is eluted from chromatin solution using Phenol chloroform: isoamyl alcohol. The chloroform denatures the proteins, separate the DNA and denatured protein suspend in an aqueous and organic phase in the solution respectively. isoamyl alcohol reduces foaming during the extraction process. In this way DNA is extracted and subjected to high throughput sequencing.

There several ways in ChIP experiment can fail. For example, ununiform fragments yield from Micrococcal nuclease digestion may result in uneven distribution of read along the genome. In addition, repetitive sequence might enrich at one site. Therefore, to determine a Statistically significant peak, a signal to noise calculation, comparing ChIP peak to the same location in a corresponding control sample is needed. (Figure 7)

In commonly, there are three way to provide control sample: input DNA, a portion of the DNA taken from sample before conducting immunoprecipitation; mock IP DNA, DNA obtained from IP without antibodies; and DNA from nonspecific immunoprecipitation (Park, 2009). Among which input DNA is most commonly used, and the control type used in this thesis.

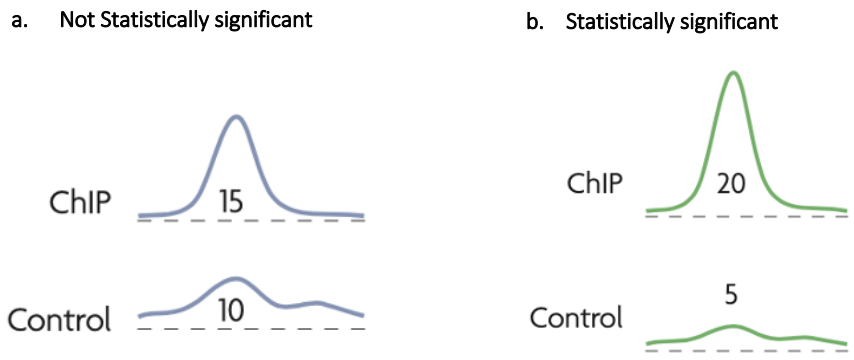


Figure.7. Example of Statistically significant and not significant peaks. a. A peak that is not statistically significant— the enrichment ratio between the ChIP and control sample is low (1.5) b. A peak can be statistically significant although the number of read is low, the enrichment ratio between the ChIP and control experiments is high (4). Figure adapted from "ChIP–seq: advantages and challenges of a maturing technology" (Park, 2009)

Background and aim

Atlantic salmon undergo a tremendous transformation in many aspects of morphology, behavior and physiology before migrating to sea water. These transformations are triggered by environmental signal, supported by endocrine system regulation. Earlier studies have demonstrated gene expression changes across life-stages is linked to physiological changes during parr-smolt transformation in Atlantic salmon (Harvey , 2019; Seear et al., 2010). However, we know very little about the mechanisms that drive the dynamics of genome regulation during smoltification. Epigenetic regulation on smoltification have studied in aspect of DNA methylation by Morán et. al, they find that the genome-wide DNA methylation patterns differ a lot between freshwater hatchery brown trout and migrant morphotype in brown trout (Morán, Marco-Rius, Megías, Covelo-Soto, & Pérez-Figueroa, 2013). Another study tested if the hatchery and wild conditions influence DNA methylation patterns in liver of steelhead, found in addition to DNA methylation difference in respond to rearing treatment, there are profound differences in DNA methylation due to age, that could indicate smoltification associated changes in liver physiology(Gavery et al., 2019). But study on histone tail modification remodeling of gene expression across parr-smolt-seawater life-stages are completely lacking. In this study we attempt to interpret this aspect of epigenetic remodeling of gene expression during different life-stages of salmon. The aim of this study is to explore the links between H3K4me3 histone modifications and gene regulation in liver of Atlantic salmon across life stages.

Materials and Method

Fish materials

Fish were provided by Norwegian University of Life Sciences (NMBU) fish lab. To compare epigenetic remodelling in Atlantic salmon in different life stage, liver samples were collected at 3 timepoints, the Parr stage, after smoltification, and after seawater transferred. Timepoint 1 (referred to as week1) was 21st weeks after first feeding, timepoint 2, was after the fish had undergone smoltification but before sea water transfer (referred to as week19), timepoint 3 was 6 weeks after transfer into seawater (referred to as week25) (Figure8). 4 fish were sampled for week1 and week25. 6 fish were sampled for week19, 2 fish were added to achieve experimental requirement. Both sexes were randomly sampled. Fish were given an artificial

winter (photo period change from L:D 24:0 to L: D8:16) to induce smoltification and sacrificed with a sharp blow to the head and liver tissue were dissected immediately after slaughter and stored at -80 degrees.

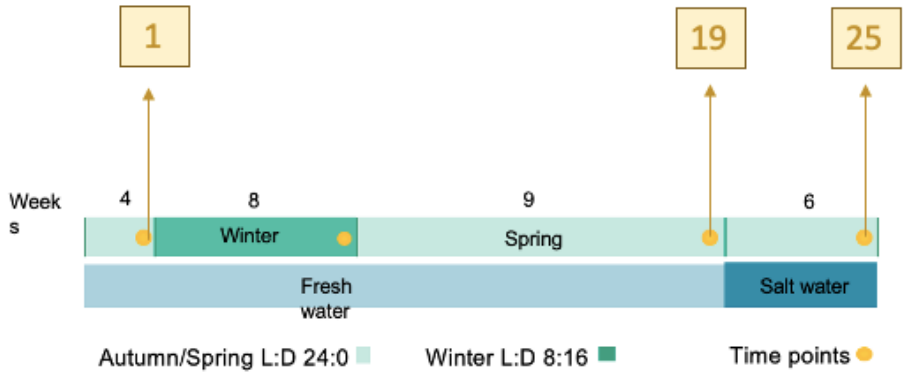


Figure8. Illustration of the sampling times across the Atlantic salmon different life stages used in this thesis. Sample start taken after 21 weeks since first feeding, therefore first timepoint named week1. Second time point is the last moment before sea water transfer after smoltification has completed, third timepoint is 6 weeks after transferred in sea. By: Line Lieblein Røsæg

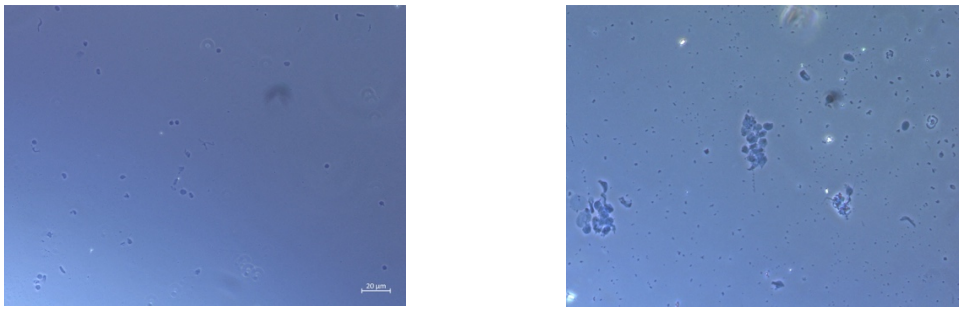
Chromatin immunoprecipitation (ChIP)

The native-ChIP protocol used in this experiment is based on Brind' Amour et el' publication (Brind'Amour et al., 2015). Buffers used in protocol attached in appendix. The original protocol is available online at https://protocolexchange.researchsquare.com/article/nprot-3501/v. In the following section I will go through the main steps of the protocol:

Step 1: Tissue disruption and nuclei isolation

The first step in the protocol aim to isolate cell nuclei without disrupting the histones in the chromatin. we collect approximately 2 million cells. Starting material of 100mg of frozen liver tissue was homogenized in cold phosphate-buffered saline (PBS) buffer with protease inhibitor cocktail (PIC), Supernatant was then discarded after 3 minutes centrifugation at 2500g. Remaining cell pellet were washed with PBS buffer 2 times and resuspended in 5ml PBS buffer. Cells were count by Bio-Rad's TC20™ Automated Cell counter. A required content of 2 million cells per millilitre can be optimized by adjusting the volume of cell solution. To isolate the nuclei from the cells, 1ml of cell solution was centrifuged for 10 minutes at 19000g. The resulting cell pellet was resuspended in 500 µl sigma nuclei isolation buffer with PIC and then kept on ice for 30 minutes while vortexed every 10 minutes. Cell membranes were then lysed by Sigma

nuclei isolation buffer, but histone will not be affected because the protection of PIC. Lastly a small sample of isolated nuclei was dyed by trypan blue and checked under microscope for quality control. Trypan blue cannot pass through living cell membrane, so the successfully isolated nuclei can be observed as blue rounds. The nucleus that separate from each other are the best for downstream experiment, as aggregation may prevent Micrococcal nuclease entering some of the nucleus. Figures below show the state of successful (Figure9a) and unsatisfactory (Figure9b) nuclei isolation states.



a b

Figure 9. Example of nuclei isolation states. a. Nuclei are abundant and independent, easy for enzyme to enter and digest chromatin. b. Limited number of nuclei are here, most of them are Sticking together, there are blue dots everywhere, probably the debris of the rupture of the nucleus.

Step 2: Chromatin digestion

Second step of protocol is to generate nucleosome-based chromatin fractions. Chromatin is digested by Micrococcal Nuclease (MNase). MNase dilution and MNase Master mix (see appendix) were prepared on ice, and 10µl was added to each sample tubes. Samples were then left for nuclease digestion shaking on an Eppendorf Thermomixer at 850g speed. Because MNase is sensitive to temperature, it is important to keep this enzyme on ice and minimize the time this enzyme is left out of the freezer. Optimal digestion time differ among different tissues. After testing aliquots at different digestion times, we decided to use a digestion time of 9 minutes at 37°C. Digestion was then stopped by adding 15 µl EDTA/ triton/deoxycholate mix (see appendix). Next, 190 µl of complete IP buffer (see appendix) was added to each sample and 50 µl sheared chromatin were taken from each sample tube as input control (Figure 5). The remaining chromatin were rotate at 4°C for 1 hour before prepared for immunoprecipitation (see Step 6).

Step 3: Check digestion state

For this step, we are testing if the digestion of chromatin at step 2 meet the requirement of continuing the protocol. Since we are aimed to digest chromatin into nucleosome-based resolution of fragments. One nucleosome is approximately 147bp, a peak around 147 pb are what we expect for. Digestion state were checked on Bioanalyzer using DNA 1000 ChIP. Below are examples of electrophoresis result and electropherogram of qualified and failed samples. For the qualified samples, a band near 150bps at in the electrophoresis image is visible (Figure 10a). In an electropherogram of successfully digested sample, a significant peak is around 150bp can be detected (Figure 10b). When the digestion is not successful, the electrophoresis would present as number 3,4,7,8 shown in (Figure 10c). For the electropherogram, no peak will be detected (Figure10d). Immunoprecipitation using unsuccessfully digested sample is meaningless, that is to say, if the electrophoresis or electropherogram suggest that the sample failed at digestion step, we need to start over again from step1.

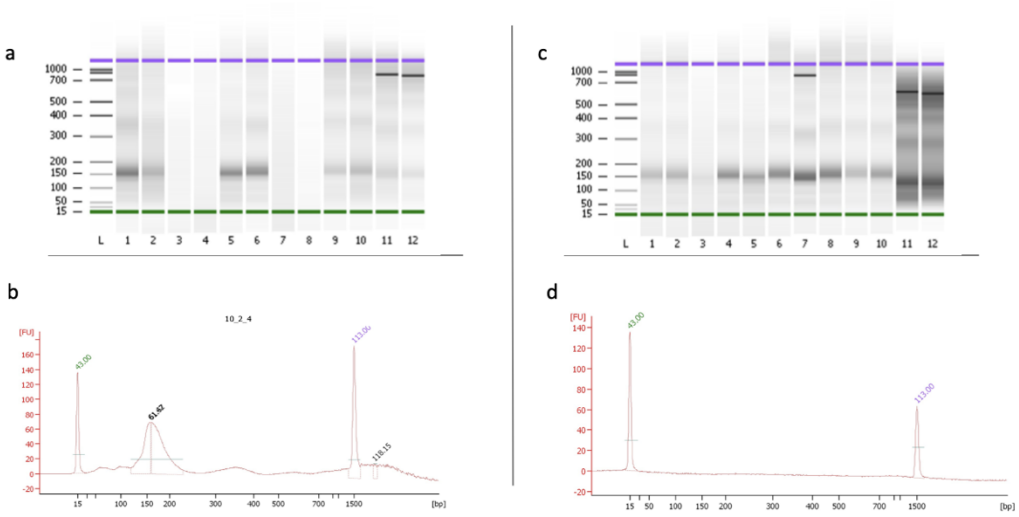


Figure 10. Bioanalyzer result of successfully and unsuccessfully digested samples. a. Electrophoresis of a group of samples all has a band near 150bp, although band at sample 3 are barely visible. b. Electropherogram of a well digested sample. some small peaks are appeared, may because of the DNA 1000 assay is not optimal, High Sensitivity kit would have had better resolution of the fragments. c. Electrophoresis on which some samples have no band appears. b. Electropherogram of a failed sample, no peak is found.

Step 4: Bead-Antibody complex preparation

To obtain the DNA fragments associated with H3K4me3 histone modification use antibodies specific for the H3K4me3 modification. The antibodies themselves are bound to proteins

complexed with magnetic beads. For each sample tube, 20 µl of 1:1 mix of dynabeads protein A and dynabeads protein G were added, diluted to 1:10 with complete IP buffer. IP buffer was subsequently taken out by placing sample tubes on magnetic rack, for washing away possible contamination from bead solution, this step repeated for two times. Protein A/G mix were resuspended in 200 PBS split in two aliquots.

Step 5: Chromatin pre-clean

To clear away any unspecific binding between chromatin and the protein A/G beads, sheared chromatin was added to one aliquots of protein A/G mix. Meanwhile, 3 µl of antibody were added to another aliquots of A/G protein mix. All sample tubes were then rotate at 4°C for 5 hours and placed on magnetic rack. The supernatant of tubes with protein beads and antibody complex was discarded. Pre-cleaned chromatin in the supernatant of the other aliquots were transferred to sample tubes with beads-antibody complex.

Step 6: Immunoprecipitation

This step is the main step of whole protocol, to pool down the chromatin fragments with histone mark H3K4me3. Sheard chromatin together with beads-antibody complex in the sample tube, were rotated at 4°C overnight. Target chromatin then be isolated from chromatin without the histone modification by placing the sample containing antibodies complexed with beads on a magnetic rack. Antibody bound chromatin is now bound to beads which again is pulled to the magnet, while chromatin lacking the histone modification is free in the supernatant, we discard the supernatant and keep the beads as well as the bound chromatin .

Step 7: Extract DNA

To obtain the DNA molecule from chromatin fragments, unbind the target DNA from the Antibody-beads complex. ChIP elution buffer were used to elute the DNA. Frist, to remove any other contamination, beads were washed with 200 µl of high salt wash buffer and low salt wash buffer (see appendix), then were resuspended in 30 µl ChIP elution buffer (see appendix) and Vortexed by max speed (2000g) on thermomixer at 65°C. Eluted DNA in supernatant was transferred to new tubes as well as another 70 µl ChIP elution buffer which used for washing the beads once again. Samples and the input controls from step2 were transferred to pre-spin phase lock gel tube, mixed with 100 µl of phenol: chloroform: isoamyl alcohol, and spin at

13000g for 5mins. The chloroform denatures the proteins, facilitates the separation of the aqueous and organic phases, while the isoamyl alcohol reduces foaming during the extraction process. Target DNA will suspend in liquid phase. The liquid phase then was mixed well with 10 µl of 3M sodium acetate, 1µl of glycogen and 275 µl pure alcohol, transferred to a new tube, precipitated in -80 C. After 1hour DNA pellet obtained by spin down sample tubes at 13000g for 30 mins. Discard supernatant add 70% ethanol, spin down again and discard supernatant, air dry to get rid of ethanol. Final DNA pellet were resuspended in 30 µl elution buffer, stored at -80.

DNA Sequencing and Raw data processing

Sequencing

Total DNA was measured by Qubit dsDNA High Sensitivity Assay Kit. Library preparation follows Illumina Pair-end protocol. Minimum of 5ng DNA for each sample was required for DNA sequencing library construction. To meet the required amount, DNA of four fish from timepoint 1 were pooled into 1 sample, named 1_pool. Three of the six fish from timepoint19 were pooled into one sample named 19_pooled. Libraries were sequenced with Illumina's pair-end sequencing assay at high throughput sequencing platform in Novogene, HONGKONG, China.

Raw read quality control

Raw reads were subjected to quality control using FastQC. (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapters in raw reads from the Illumina sequencing were removed by TrimGalore (https://github.com/FelixKrueger/TrimGalore) using the default adapter option ('AGATCGGAAGAGC'). FastQC report overall quality assessments for the sequencing both before and after adapter remove, such as per base pair quality of the reads, GC content distribution, sequence duplication statistics. Output result as html file. Illustrate the quality with colors: green as good, orange slightly abnormal, warn very unusual samples with red. An overrepresented sequence will be reported if one fraction of sequence makes up more than 0.1% of the one library.

Mapping and quality control

To align the short reads to the reference genome, the QC filtered reads were mapped using Bowtie2 (Langmead & Salzberg, 2012). The section below briefly describes the relevant steps in the read mapping procedure and explains the most important parameters affecting the mapping results. To identify the positions of the genome where reads are derived from alignment scores are calculated. Specifically, the scores are calculated by penalizing different types of character mismatches, e.g. a one-character mismatch is penalized by subtracting 6 'points', a two-character mismatch gap gets a penalty of 11, and so on. The higher the score, the more similar the read is to a genomic position. Since the reads are relatively short, they can theoretically match several positions in the genome. To handle this, a P-value is calculated that is reflecting the possibility of read to match several positions in the genome. Based on this P-value the software calculates a Q value which is a non-negative integer (Q = -10 log10 p) reflecting 'how uniquely' each reads map. A *uniquely mapped* alignment is the one with has one best score than any other possible alignment. It is possible to have more than one alignment same high Mapping quality, refer to as *multi-mapped*. we can define how many alignments can be accepted by setting the report mode. In our experiment, -k mode was used. If -k set as N, Bowtie will search up to N valid alignment report them all in descend order of the alignment score. In this project we used -k 10.

ChIP-seq data analysis

All the data analysis in this thesis were done in Rstudio version 3.6.2. server on the NMBU Orion computing cluster.

Peak calling

To find genomic regions where H3K4me3 marks are enriched, we used the Genrich software (https://github.com/jsh58/Genrich#method) to call significant 'peaks'. These peak-regions are found by identifying regions in the genome where we find more read alignments then expected by chance. Briefly, Genrich first removes reads that represent potential PCR-duplicates with identical sequences in both ends of the reads. However, multi-mapped reads were taken into account by taking a fractional count to each location. Next Genrich build up a 'background genome enrichment' signal using the input control. Based on the local genomic signatures of background enrichment, a p-value for each potential peak region in the H3K4me3 ChIP-seq

data can be calculated. The output results in a tab delimited bed file containing chromosome name, start of peak, end of peak, and the enrichment score.

Profile of H3K4me3-signals near the transcription start region

Since H3K4me3 considered as an active promoter mark, an important quality control is to check for enrichment of peaks near the transcription start site (TSS). We therefore calculated enrichment of overlaps between H3K4me3-peaks and TSS using the function getTagMatrix () in the Chipseeker R-package (Yu, Wang, & He, 2015).

Peak annotation

We performed peak annotation by determining which genomic features each peak overlaps using the R package Chipseeker. This tool reads in peak files obtained from the peak calling step and use the reference genome (ICSASG_v2) to conduct the annotation. The tool reports the genomic features covered by peaks and nearest genes according to the positional information. Since one peak can overlap several different genomic features Chipseeker prioritize genomic features as follows: Promoter, 5' UTR, 3' UTR, Exon, Intron, Downstream, Intergenic. The annotation result is output as a table containing peak annotation information, as well as location of nearest genes to each peak.

Gene ontology test and KEGG test

Enrichment tests for gene ontologies (GO) - and Kyoto Encyclopedia of genes and genome (KEGG) pathways were used to associate changes in histone tail modifications with biological functions of genes. This allows us to deduce if changes in ChIP-seq signals through development is linked to specific changes in animal physiology and/or tissue function. Since promoters are located at the 5' ends of genes surrounding the TSS (Brind'Amour et al., 2015), all the functional enrichment tests were done on genes that were classified as having a significant H3K4m3 signal in the core promoter region define as (-110, +110). We analyzed GO/KEGG enrichments across two types of gene sets: (1) gene with H3K4me3 marks across all timepoints, (2) genes with developmental stage specific H3K4me3 signals, and KEGG-pathway enrichment we carried out using the kegga in R. GO enrichment analysis were done using R function topGO, in biological process domain, use a significance threshold elim<0.001. GO terms of uniquely present genes in each timepoints were compared.

Integrating H3K4me3 signals and gene expression change during smoltification

We can envisage two distinctly different processes driving transcriptional regulation during smoltification. Either gene transcription is driven by changes in the activity of transcription factors independent of histone code, or dynamic changes in gene regulation is tightly associated with remodeling of histone tail marks. To explore the link between changes in gene expression levels and changes in H3K4me3 histone tail modifications, we used the results from a previous study on gene expression regulation during smoltification (Harvey, 2019) This study identified 5 gene co-expression clusters (including a total of 6054 genes) that have different expression changes across different life stages of Atlantic salmon. (Figure 11). To associate gene expression with H3K4me3 signals we counted how many genes in each cluster have H3K4me3 at different time points. If the H3K4me3-histone code is tightly associated with smoltification gene regulation, we expect high overlap between genes with H3K4me3 signals in week 1 and genes belonging to clusters with higher expression at the parr stage (Figure 11, clusters 1 and 2), as well as high overlap between H3K4me3 signals in week 19 and genes in co-expression clusters with increasing expression towards later stages in smoltification process (Figure 11, clusters 4 and 5).

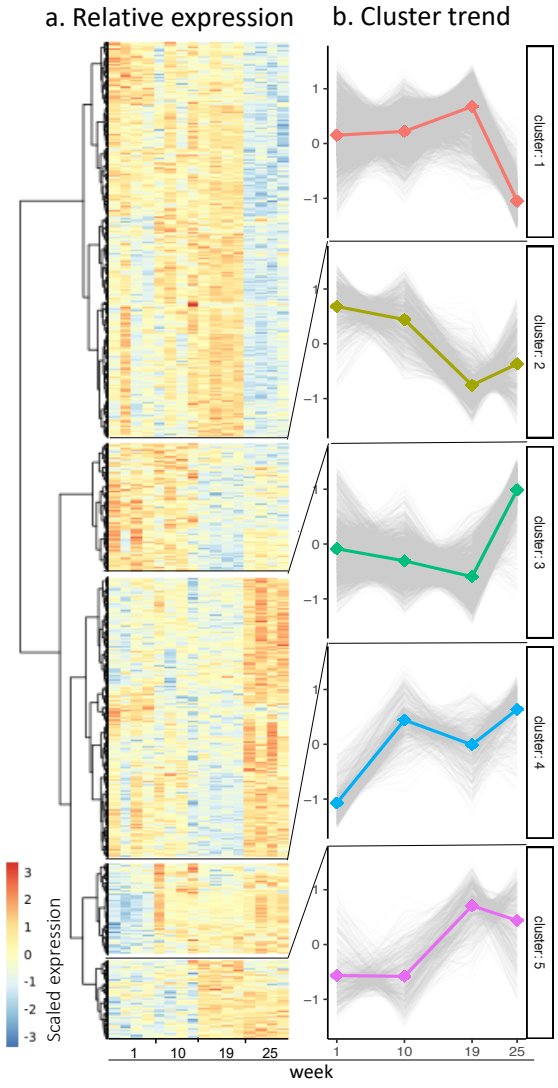


Figure 11. a. Global co-expression clusters for genes with significant changes in expression during smoltification. Each row is a gene that were found to be significantly changing (ANOVA, FDR corrected p <0.05) during smoltification. Rows are normalized to the median row-value. **b**.Co-expression clusters was identified using

Results

Sequencing and raw data processing

General ChIP-seq statistic

A total of 14 ChIP samples from week 1, week19 and week25 were generated (Table 2). The amount of precipitated DNA varied more than 18-fold, with samples from week-1 having generally lower DNA amounts compared to week-19 samples. To obtain enough amount of DNA for constructing sequencing libraries, all 4 samples from timepoint 1 were pooled into 1 sample. For week19, two more samples were added, finally sample 19_1_2, 19_1_3, 19_1_4 were pooled together in one sample. Thus, in total 9 sample were sequenced, DNA amount vary among sample from 4.7 ng to 11.175ng. Between 44-63 million pair end sequence were generated for each sample, with a high proportion aligned reads (>90%) (Table 2),

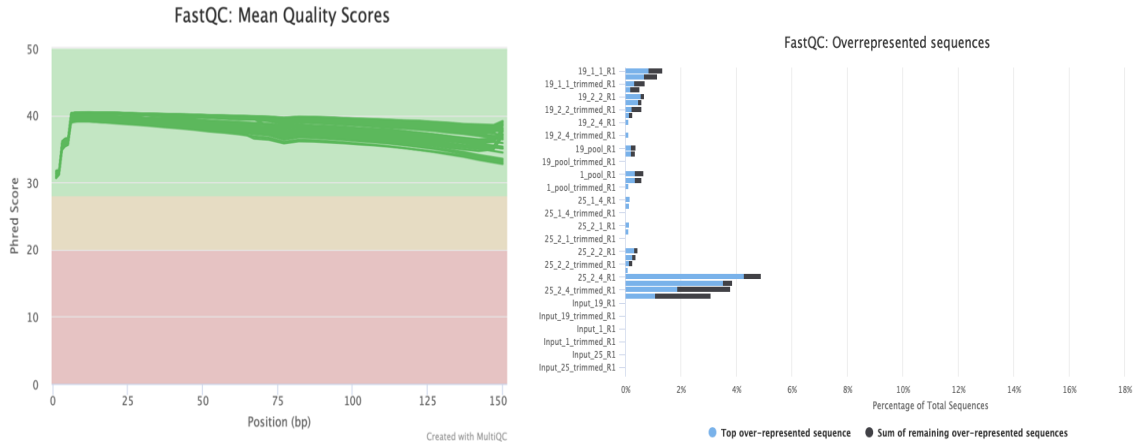
Table 2. overview of ChIP-seq data stats

Timepoint	Fish Info	DNA Concentration (ng/ µl)	Seq. Sample name	Extracted DNA (ng)	M Total seqs	% Aligned	Called peak	Merged peak
Pre-winter (week1)	1_2_4	2,8	1_pool	4,7	50.2	97.2%	20885	20888
	1_2_2	0,7						
	1_1_4	0,6						
	1_1_2	0,6						
Smolt (week19)	19_2_4	11,175	19_2_4	11,175	52.7	96.9%	186	27667
	19_2_2	6,9	19_2_2	6,9	54.6	96.2%	236	
	19_1_1	4,945	19_1_1	4,945	52.3	95.5%	110	
	19_1_4	3,835						
	19_1_2	0,925	19_pool	8,6	63.2	97.2%	22472	
	19_1_3	3,84						
Saltwater-smolt (week25)	25_2_4	6,765	25_2_4	6,765	50.8	91.8%	4149	20387
	25_2_2	8,025	25_2_2	8,025	46.9	96.4%	157	
	25_1_4	13,95	25_1_4	13,95	44.8	96.9%	84	
	25_2_1	11,64	25_2_1	11,64	44.5	97.2%	72	

Raw read quality control

In total 9 IP samples and 3 input control were assessed with FastQC indicating high quality sequencing data for all the samples. The average per base Score is >30 quality indicating successful sequencing reactions (Figure12a). All the samples had some overrepresented sequences. However, most of the sequences were significantly reduced after trimming with trimmomatic which indicates that the overrepresented sequences mainly come from adapter

sequences. Whereas, sample 25_2_4 has more overrepresented sequence that still remain considerable percent of overrepresent sequence after trimmed (Figure 12b).



a b

Figure12. Sequencing quality summary. a) Mean quality score of all sample, green region indicate reads is normal, yellow means slightly skeptical, red means low quality. This figure shows our sample are all in good quality, no sample needs to be removed. **b)** overrepresented sequences when a fraction of sequence makes up more than 0.1% of the one library, it will be report as an overrepresented sequence. Light blue is percentage of top overrepresenting reads, dark color stand for the sum up percentage of remaining overrepresented reads.

The proportion of duplicated sequences ranged from 19.7 to 33.3 (Table 3). The GC content of samples matches the Atlantic salmon genome GC content of 44.4% (Table 3) (Davidson et al., 2010), indicating that the sequencing libraries has no large-scale contamination. The length of the sequences after adapters are removed range from 135 to 141, which is normal considering our target DNA should be around 147bp. (For whole sequencing data quality including before trimmed quality, see appendix).

Table 3. Sequencing data quality. The table contains basic quality control statistics from the fastqc software.

Sample Name	% Dups	% GC	Length	M Seqs
1_pool_trimmed_R1	28.0%	43%	137 bp	25.1
1_pool_trimmed_R2	26.0%	43%	135 bp	25.1
Input_1_trimmed_R1	22.4%	44%	141 bp	26.6
Input_1_trimmed_R2	19.7%	44%	140 bp	26.6
19_1_1_trimmed_R1	24.2%	43%	139 bp	26.1
19_1_1_trimmed_R2	21.9%	43%	137 bp	26.1
19_2_2_trimmed_R1	23.8%	43%	140 bp	27.3
19_2_2_trimmed_R2	21.0%	43%	137 bp	27.3
19_2_4_trimmed_R1	22.9%	44%	141 bp	26.3
19_2_4_trimmed_R2	20.4%	44%	138 bp	26.3
19_pool_trimmed_R1	33.0%	42%	140 bp	31.6
19_pool_trimmed_R2	30.2%	42%	139 bp	31.6
Input_19_trimmed_R1	22.6%	44%	139 bp	27.1

Input_19_trimmed_R2	20.0%	44%	138 bp	27.1
25_1_4_trimmed_R1	20.6%	43%	139 bp	23.1
25_1_4_trimmed_R2	18.4%	43%	137 bp	23.1
25_2_1_trimmed_R1	21.8%	43%	138 bp	22.9
25_2_1_trimmed_R2	20.2%	43%	136 bp	22.9
25_2_2_trimmed_R1	22.9%	44%	140 bp	24.3
25_2_2_trimmed_R2	20.3%	44%	138 bp	24.3
25_2_4_trimmed_R1	27.0%	44%	138 bp	27.7
25_2_4_trimmed_R2	24.4%	44%	136 bp	27.7
Input_25_trimmed_R1	22.0%	43%	139 bp	22.4
Input_25_trimmed_R2	19.5%	44%	138 bp	22.4

Mapping quality control

Alignment rate of all samples and input controls are above 95% (Table 4), except sample 25_2_4 which had lower percent mapped reads. Input control and pooled samples have lowest error rate. Whereas pooled samples have lowest non-primary reads, input controls have most.

Table4.Mapping quality

Sample Name	Error rate	M Non-Primary	M Reads Mapped	% Proper Pairs	M Total seqs	% Aligned
1_pool	0.98%	153.6	48.8	94.0%	50.2	97.2%
Input_1	0.95%	235.5	52.3	96.7%	53.2	98.4%
19_1_1	1.37%	227.0	49.9	89.7%	52.3	95.5%
19_2_2	1.50%	225.3	52.5	88.0%	54.6	96.2%
19_2_4	1.44%	235.9	51.1	90.9%	52.7	96.9%
19_pool	0.98%	163.3	61.4	91.7%	63.2	97.2%
Input_19	0.95%	239.1	53.3	96.7%	54.2	98.4%
25_1_4	1.42%	195.7	44.8	90.2%	46.2	96.9%
25_2_1	1.22%	203.8	44.5	93.8%	45.8	97.2%
25_2_2	1.53%	214.7	46.9	90.3%	48.6	96.4%
25_2_4	1.30%	203.8	50.8	87.2%	55.4	91.8%
Input_25	0.98%	199.6	44.0	96.6%	44.8	98.3%

Generally, the alignment scores are similar across all samples (figure 13), with 40-50% of uniquely mapped pairs of reads. The biggest fraction of mapped reads is multimapping pairs, which likely reflect the history of whole genome duplication in the Atlantic salmon ancestor. It worth mentioning that the two pooled samples have more uniquely mapped pairs compared

to single biological replicates. Only small proportion of pair ends either separately mapped to different place or only one of the pair find a place.

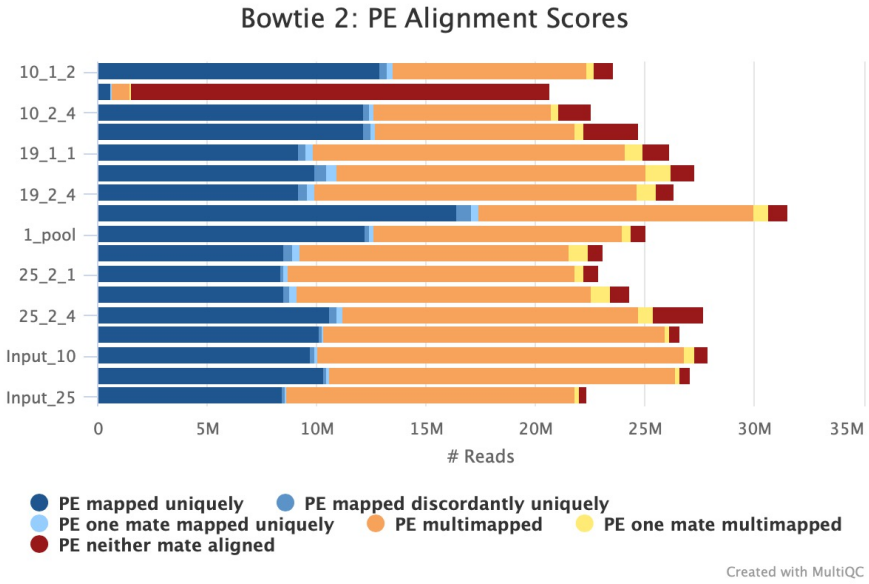


Figure 13 Bowtie2 Alignment score

Peak calling

In all the single biological replicate samples, called peak's number was extremely low compared to the pooled sample (Table 5). The pooled samples (timepoints 1 and 19) show a more expected number of peaks (compared to other liver H3K4me3 data we have generated in the lab, data not published). In all the following analyses we therefore pooled all data from all samples to call a merged peak set that was used for all downstream analyses. Timepoint 19 had most peaks (27667) followed by timepoints 1 and 25 with 20888 and 20387 peaks, respectively.

Table5. Peaks number.

Sample Name	peak.number	Merged peak
1_pool	20885	20888
19_1_1	110	27667
19_2_2	236	
19_2_4	186	
19_pool	22472	
25_1_4	84	20387
25_2_1	72	
25_2_2	157	
25_2_4	4149	

Peak enrichment around transcription start site

H3k4me3 signal enrichment in promoters across the three time points show (merged peak set from each timepoint) very similar patterns across the three peak sets, with a clear signal enrichment in the promoter/5UTR (Figure 14). The similar pattern in peak sets from all time points suggest that the strategy of pooling all data prior to peak calling has captured biological meaningful signals.

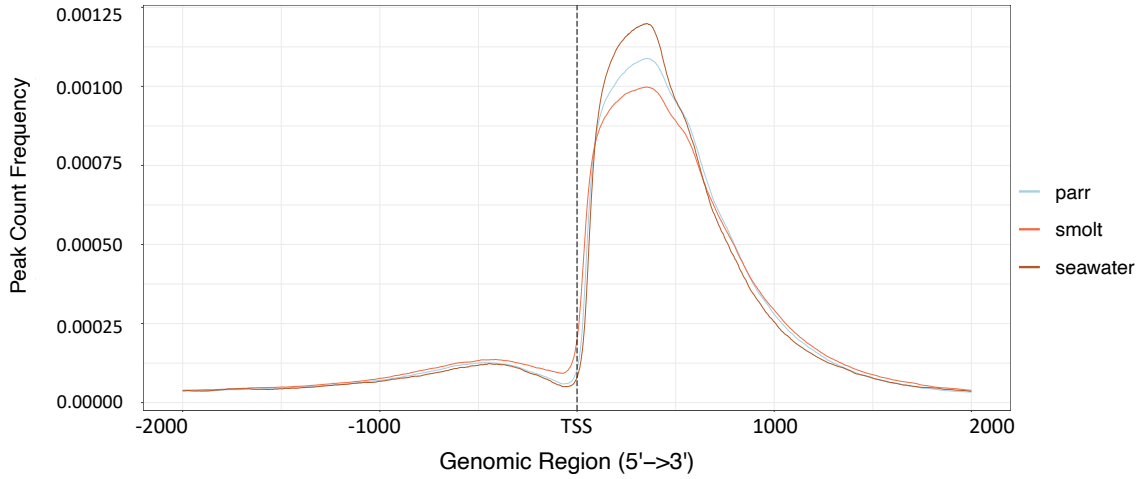


Figure 14. Peak enrichment around transcription start site. Figure shows the enrichment of peaks between -2000 to +2000 bp to the transcription starting site, x axis is distance to transcription start site, y axis is frequency of peak count. Blue line illustrates peaks in week1, red line shows week19, brown line stand for enrichment of week25.

Peak annotation

Gene feature annotation using Chipseeker shows that annotated H3K4me3 peaks across the three time points have a quite similar distribution (Figure 15), but we do find that Week-1 and week-19 have slightly a greater number of peaks overlapping promoter region compared to week-25 (light blue fraction in Figure 15). All samples have similar and relatively low proportion of peaks outside the gene space.

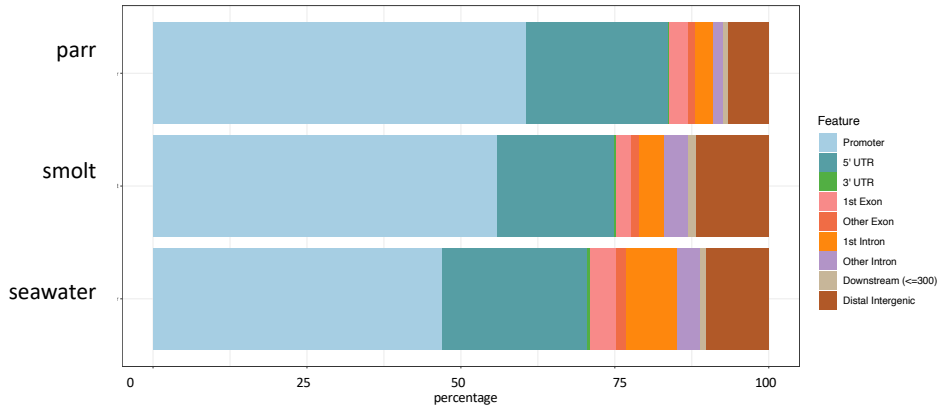


Figure 15. Gene feature overlapped by peaks in three timepoints. Y axis show the time points, x axis is percentage of peaks in total peaks in each timepoints covering different gene features. Promoter set as (-110, +110), indicating most of the peaks are very close to the TSS.

Functional enrichment analysis

Figure 16 shows the numbers of shared and unique genes with H3K4me3 signal on promoter in three sampling times. 11860 genes in week1, 13435 genes in week19, and 8492 genes in week25 have assigned a H3K4me3 peak in the promoter, most of which are shared across all timepoints (7899 genes). We found 545 and 1897 genes with unique H3K4me3-signals in week-1 and week-19 respectively. The seawater stage week-25 fish however only had 88 unique genes with H3K4me3-promoter signals.

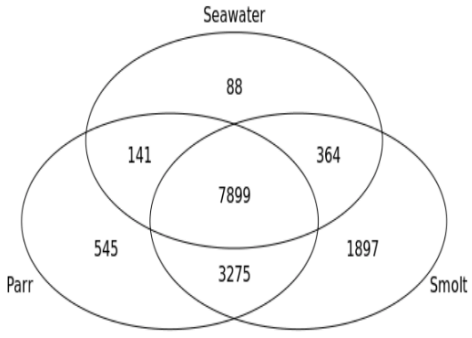


Figure16. Venn plot of genes with H3K4me3 signal on promoter in three time points. Parr is Atlantic salmon in week1, Smolt is when fish is in week19, Seawater stand for week25.

KEGG-pathway enrichment of shared H3K4me3 signals across development

We first looked at KEGG-pathway enrichment of genes with shared H3K4me3 promoter marks across all timepoints (7899 genes). Our prediction is that these genes should be enriched in general cellular housekeeping functions, and not be related to developmental stage specific processes. Indeed, except for 'Herpes simplex virus 1 infection', all significantly enriched pathways are related to basic cellular processes, supporting our prediction about these shared H3K4me3-signals relating to actively transcribed housekeeping genes (Table 6).

Table 6. KEGG pathway in all shared genes. N: number of annotated genes in this pathway. DE: detected genes, P.DE : significance of detected pathway.

Pathway	N	DE	P.DE
Spliceosome	300	153	3.0324617285243736e-26
Ribosome	288	133	4.008368532222761e-18
Oxidative phosphorylation	255	120	2.9446043201657953e-17
Protein processing in endoplasmic reticulum	395	165	6.664115523240038e-17
Autophagy – animal	368	147	2.677632380078601e-13
Autophagy – other	62	41	7.00999733731834e-13
RNA transport	319	126	3.41076626915918e-11
Ubiquitin mediated proteolysis	319	120	3.379877382329199e-9
Proteasome	115	55	5.692414009853913e-9
mRNA surveillance pathway	177	72	1.4137556981578247e-7
Mitophagy – animal	182	72	5.111354745409648e-7
Protein export	45	26	6.164924640365227e-7
Endocytosis	647	197	8.099524506542335e-6
RNA polymerase	49	25	2.0244413162318402e-5
mTOR signaling pathway	372	118	8.339472878939763e-5
Citrate cycle (TCA cycle)	72	31	1.4521994607022453e-4
Ribosome biogenesis in eukaryotes	141	52	1.641924282173981e-4
RNA degradation	156	56	2.1429811138368534e-4
Lysine degradation	123	46	2.6405631544158185e-4
Herpes simplex virus 1 infection	909	251	7.455983620409566e-4
Carbon metabolism	262	83	9.269677744053027e-4
Basal transcription factors	66	27	9.973936647316905e-4

Gene Ontology analysis of gene promoters with H3K4me3 signals unique to single time points

Next we performed functional enrichment for genes with H3K4me3 promoter marks only found in one developmental stage. These genes are candidates for developmental time specific functions. The 545 genes with unique H3K4me3 marks in week-1. Thirteen significant (adusted p-value < 0.01) gene ontology terms were found for unique genes in parr, with the many enriched terms being related to developmental processes such as "positive regulation of cell migration", "endothelial cell migration", "nephric duct morphogenesis".

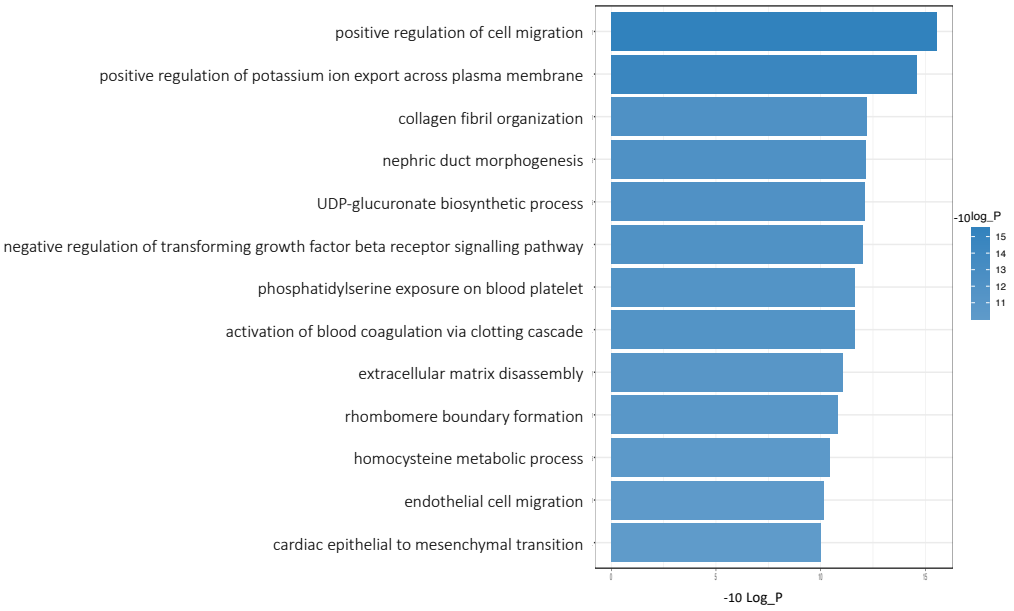


Figure17. visualization of GO term in week1. Y axis is term, x axis shows log2 of the p_value

In week 19 many functions related to cell division (e.g. DNA replication, mitotic recombination), protein biosynthesis. There are also many terms related to metabolism and energy (e.g. glycoside catabolic process, NADP regeneration)

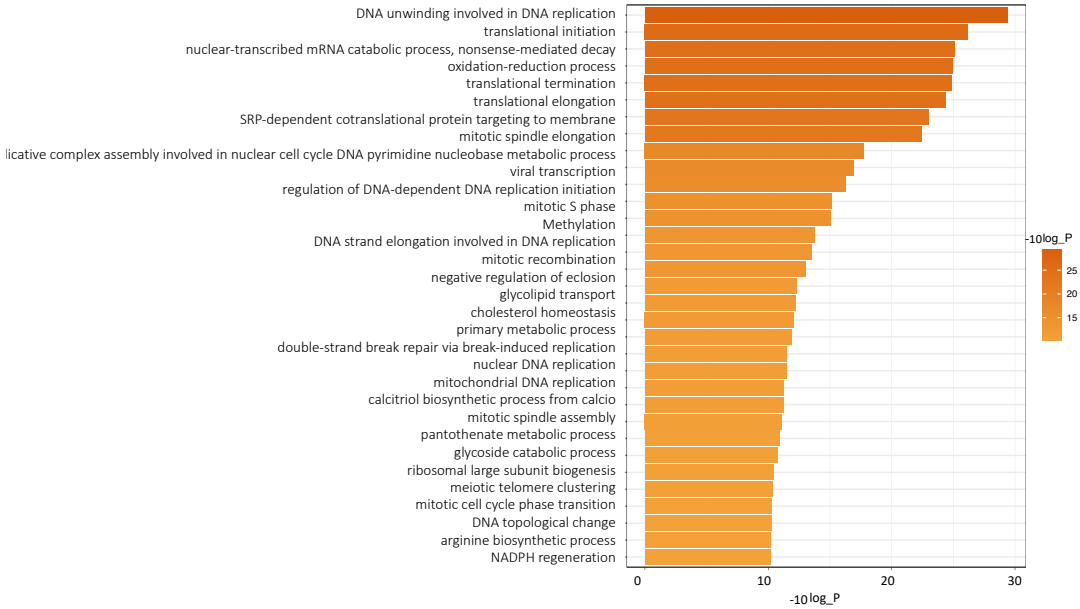


Figure18 Visualization of GO term in week19.

The H3K4me3 promoter peaks in seawater were few, and gene ontology test with a p-value cutoff of <0.01 did not yield any significant terms. We therefore increased the p-value cutoff threshold to < 0.05 for this gene set. The most significant function was positive regulation of neuron migration. In addition, we found terms related to regulation of neuron system, immune system, and chloride ion homeostasis.

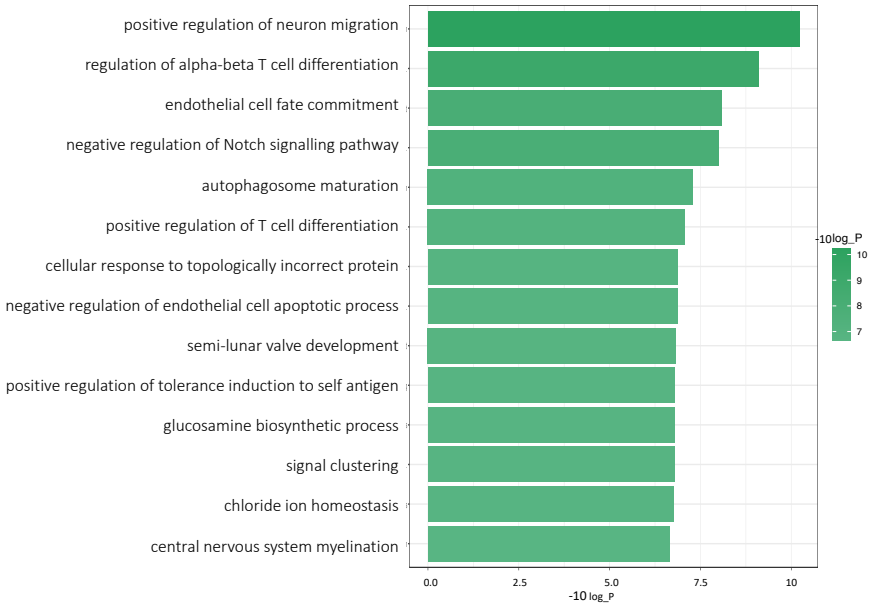


Figure19. Visualization of GO term in week25.

Comparison of H3K4me3 enrichment with gene expression level

Genes that overlapping

Most of the genes with a H3K4me3-peak in the promoter did not overlap with any differential expression clusters (NA in Table 7). By comparing the percentage of overlap between the genes in the five clusters (Figure 11b) and the genes with H3K4me3 peaks in promoters we observe a clear association between histone code changes and gene expression changes (Figure 20).

Table 7. Genes overlapping with GSF2 cluster genes in 3 timepoints.

GSF2.cluster	shared	Week1	Week19	Week25
1	801	41	250	2
2	211	26	38	3
3	436	40	78	16
4	171	3	46	5
5	62	4	61	1
NA	6218	431	1424	61

For example, genes in cluster2 which are highly expressed in Parrs and then transiently decrease in expression throughout smoltification, has the highest percentage of promoters annotated with H3K4me3-peaks (4.77%, Figure 20). Conversely, genes that tend to increase in expression from parr to smolt (clusters 4 and 5) have a large gain in H3K4me3 promoter signals in smolts (Figure 20). Genes that change dramatically in expression upon seawater transfer are genes in cluster 1 (decreasing) and cluster 3 (increasing). In line with this we find a substantial decrease in H3K4me3-promoter signals (3.2-6 fold) for cluster 1 genes, and an increase in H3K4me3-marks on promoters for cluster 3 genes.

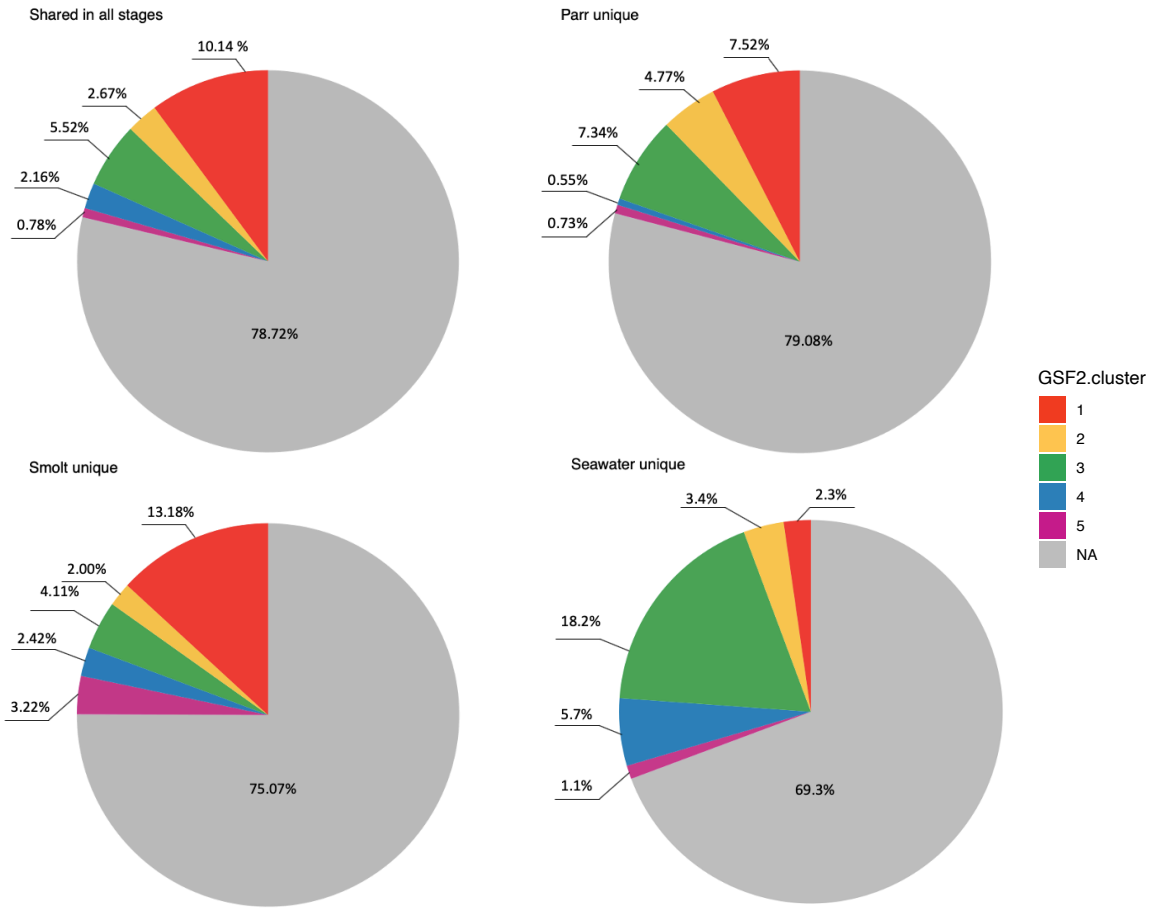


Figure 20. pie chart illustrates the percentage of GSF2 cluster genes in different time points. Only around thirty percentage of differential expressed gene are overlapping with h3K4me3-marked gene. However, these genes change over timepoints as the trend of gene expression change.

Genes not overlapping.

There are still a considerable number of genes that have H3K4me3 mark in promoter but do not overlap with differential expressed gene (69.3%-79.08%). Indicating that these genes are specifically marked with H3K4me3 at the certain timepoint, yet not showing differential

expression. Moreover, 35-65% of the genes that are regulated actively through smoltification (Figure 11a) does not have a H3K4me3-promoter signal (Table 8).

Table 8. Smoltification regulated genes with no H3K4me3 signal.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Total genes	2461	794	1745	565	489
Genes with No mark	862	418	982	265	320
No mark gene percentage	35 %	53 %	56 %	47 %	65 %

GO test of differentially expressed genes during smoltification that does not have any H3K4me3-promoter signal in any timepoint, show that many of these gene are related to biological processes that we know is part of the important physiological remodeling of liver function during parr-smolt transformation or seawater adaptation (Gillard et al., 2018;).

Table 9. GO test of smoltification regulated genes with no H3K4me3 signal

	GO.ID	Term	Annotated	Significant	Expected	elim
4	GO:0048251	elastic fiber assembly	50	11	2.87	1E-04
12	GO:0006868	glutamine transport	19	6	1.09	5E-04
5	GO:0006094	gluconeogenesis	384	41	22.03	0.00011
6	GO:0032201	telomere maintenance via semi-conservative replication	31	8	1.78	0.00028
7	GO:0050829	defense response to Gram-negative bacterium	169	22	9.69	0.00028
8	GO:0006982	response to lipid hydroperoxide	7	4	0.4	0.00033
9	GO:0006083	acetate metabolic process	7	4	0.4	0.00033
10	GO:0006114	glycerol biosynthetic process	7	4	0.4	0.00033
11	GO:0021680	cerebellar Purkinje cell layer development	161	21	9.24	0.00037
13	GO:0044245	polysaccharide digestion	8	4	0.46	0.00063
14	GO:0006633	fatty acid biosynthetic process	356	36	20.42	0.00074
15	GO:0000038	very long-chain fatty acid metabolic process	92	14	5.28	0.00074
16	GO:0019752	carboxylic acid metabolic process	2650	231	152.02	0.00078
17	GO:0042415	norepinephrine metabolic process	63	11	3.61	0.00084
18	GO:0015808	L-alanine transport	21	6	1.2	0.00091
1	GO:0061402	positive regulation of transcription from RNA polymerase II promoter in response to acidic pH	12	6	0.69	2.4e-05
2	GO:0006107	oxaloacetate metabolic process	24	8	1.38	3.7e-05
3	GO:0006271	DNA strand elongation involved in DNA replication	54	12	3.1	4.4e-05

Discussion

This discussion will be divided into two parts. First, I will discuss the biological findings related to the regulation of gene expression changes during smoltification. Then I will discuss aspects relating to the native-ChIP method and data quality.

H3K4me3 regulation is associated with smolt liver function and development

To explore the connection between H3K4me3 and gene regulatory changes across the parr-smolt transition, we used ChIP-seq to identify genes with H3K4me3-marks in promotes in three time points (Figure 5). Unfortunately, we failed to produce high quality ChIP-seq data from biological replicates. Hence, the discussion that follows will be based on genome wide patterns of H3K4me3-signals from pooled replicates, which only allowed us to assess the H3K4me3 signals in a qualitative way (i.e. presence or absence of H3K4me3 peaks in promoters).

Most of the genes (7899) with H3K4me3-signals in the promoter regions were shared across the smoltification development (Figure 16). KEGG test (table5) for these gene reveal that most significantly enriched pathways are associated to typical basic cellular functions such as Spliceosome, Ribosome, RNA transport, which are processes expected to be stable across all the life stages.

However, we also find promoters with dynamic regulation of the H3K4me3 mark during smoltification. Promoters with unique H3K4me3 signals at the parr stage are enriched in genes involved in developmental processes (Figure 17), most relevant term is “cell migration”, other terms are such as “extracellular matrix disassembly”, “rhombomere boundary formation” are also related to development and morphogenesis (Clause, 2013;Dahmann, 2011), This result could be explained by the fact that parr are still in the stage of a physiological window of development from a juvenile fish into a more ‘adult physiology (Hoar, 1988). In addition to this there are also terms associated to key metabolism processes that take place in liver, like: “UDP-glucuronate biosynthetic process”, “homocysteine metabolic process” which could be related to energy homeostasis in the fast growing parr.

The smolts timepoint have most genes with unique H3K4me3 promoter marks, with a strong bias towards cell division and metabolism related functions (Figure 18). It was report by Higgins (1985) that the growth and metabolism increase during smoltification. McCormick (2012) also report that body shape changes in smolts is related to an increased growth and basal metabolic rates, combined with reduced lipid content (i.e. high lipid turnover). A large-

scale microarray study on salmon gene expression using gill, kidney, brain found that in smolt, a gene crucial for protein synthesis (translation elongation factor 2, EF-2), was upregulated in all three tissues (Seear, 2010). Hence, this fits well with our results that the smolt stage H3K4me3-marked genes is highly enriched in genes associated with growth, translation, cell division, and metabolism. In previous studies many lipid-metabolism related genes are transcribed less towards the end of smolt development (Sheridan, 1989; Thomas N. Harvey)., however in contrast to this we find that "glycolipid transport", "cholesterol homeostasis" are enriched in genes with smolt-specific H3K4me3-signals. There are of course many things that could explain this discrepancy, one being that these genes are in chromatin which becomes enriched in the repressive H3K27me3 mark. This type of bivalent combination is considered to poise expression of developmental genes, related to timely activation caused by environmental cue (Voigt, 2013). Another explanation is that the transcription of key transcription factors regulating these lipid metabolism genes are actively repressed.

Finally, even though seawater stage has the fewest unique H3K4me3-marks, we do find biological meaningful GO enrichments such as regulation of ion homeostasis, which is known to occur following the increased chloride concentration in seawater. Seawater unique GO enrichments are also enriched in immune system regulated genes. Immune system genes in general (Johansson., et al) and in the NOTCH-pathway specifically (Shang, 2016), are previously known to be down-regulated in smolts following seawater transfer. In line with this we also find enrichment of H3K4me3-marks on genes involved in negative regulation immune system and the NOTCH-pathway.

H3K4me3 histone code is weakly correlated with differential gene expression

Study have revealed tri-methylation on K4 is present exclusively at active genes (Santos-Rosa et al., 2002). and that this histone mark interacts with the transcription initiation factor III (TAF3), which direct global TFIID recruitment to active genes and regulate the assembly of the preinitiation complex (Lauberth, 2013). In this thesis we associated the presence of H3K4me3-marks and gene expression changes (co-expression clusters) and found a weak but consistent association between gene regulation patterns during smoltification and gain/loss of H3K4me3 modification in gene promoters (Figure 20). These associations includes coordinated drop in expression and H3K4me3 marks (cluster 1: genes downregulated in seawater, cluster 2: genes downregulated in smolts relative to parr) as well as genes that are induced (cluster 3: genes

induced in seawater, cluster 5: genes induced in smolt relative to parr) (Figure20). In conclusion, our results support a link between epigenetic regulation, in particular histone code regulation, and gene regulatory programs during smoltification.

However, we also find many genes with highly regulated transcription that has H3K4me3-marks present across all timepoints (co-expressed clusters overlapped in Shared group). Since we only classified the H3K4me3 signal in promoters qualitatively, any quantitative (level of signal) difference cannot be deduced. From studies in other organisms we know that the level of the H3K4me3 (peak height) around transcription start site is correlated with level of transcription (Pokholok, 2015) (Howe.et al). Hence, if we would have succeeded in generating high quality biological replicates of H3K4me3 ChIP-seq our conclusions could have been different. It is likely that correlating quantitative RNA-seq signals with quantitative signal of H3K4me3 in promoters would have revealed a more significant associations between changes in histone tail modifications and gene regulation during smoltification.

There also large number of co- expression cluster genes that are not presenting H3K4me3 signal at any timepoints. GO test on these genes shows terms highly related to liver metabolism, such as “fatty acid biosynthetic process”, “very long-chain fatty acid metabolic process” (Harvey, 2019). These results suggest that important aspects of the smoltification-assotiated transcriptional remodeling of liver is independent of H3K4me3 regulation.

Although, our results demonstrate some associations between gene expression levels and H3K4me3-marks, there are far more genes that do not show this link. In fact, 70% of H3k4me3 signals is not associated with any differential expression at the transcription level. One explanation is that these genes are in a poised state to become active. Such poised promoters have both H3K4me3 as well as a repressive histone modification such as H3K27me3 (Voigt, 2013). Since we did not measure H3K27me3, we cannot test this hypothesis. Another reason for observing clear H3K4me3 regulation but no/little changes in gene expression could be due to the p-value cut-off for the differential expression tests. The original RNA-seq study could contain many genes with low expression levels and modest changes in regulation that would not be identified as significantly regulated during smoltification.

In conclusion, the analyses of H3K4me3-signals and gene expression fits with a model where some genes are regulated through histone tails remodeling during parr-smolt transformation, but this level of genome regulation does not play the major role in the developmental transition from a parr to a smolt.

The native-ChIP method and data quality

In this project we failed to generate high quality biological replicates using native-ChIP of histone tail H3K4me3 (Table 5). Many of the biological replicates only produced a few hundred H3K4me3 peaks, reflecting extremely low signal to noise ration. Strangely, the poor performance (i.e. low number of called peaks) for many of our biological replicates that we excluded from our analyses did not correlate strongly with the native-ChIP DNA-yields, nor the number of duplicated sequencing reads (Table 3). Possible reasons for low signal to noise ratio in ChIP-seq experimentas are listed in Table 10.

Table 10. possible causes to High background signal and low chip signal.

Effect	Possible causes
High background signal	Antibody concentration is too high
	Non-specific signal is insufficiently washed
	Nature of the histone mark
Low ChIP signal	Antibody concentration is too low
	Poor antibody quality
	Antigen is very rare
	Post-Immunoprecipitation wash are too harsh

As PCR duplicates are removed before peak calling, it is not likely causing the peak calling failure. However, PCR duplicates arise when DNA undergo many rounds of amplification (due to low starting material), which will introduce PCR-bias and reduced sample complexity when shorter fragments are amplified more efficiently (Minikel, 2012). In addition, due to the low DNA concentration from each immunoprecipitation experiments, several rounds of immunoprecipitation needed to be done for per sample. To adjust the input volume for library preparation, some DNA may have lost at a clean-up with magnetic beads process, although testing for DNA loss did not suggest that this was a problem in our experiments. Another factor

that could have caused poor library quality is that MNase digestion could lead to bias in digestion of the open chromatin regions, refered to as “regional bias” (Park, 2009).

The high background signal, making the low ChIP signal in non-pooled sample almost impossible to detect. In the pooled samples, although the signal-to-noise ratio may not change, when the reading in the peak is high enough, the signal can be distinguished from the background signal(Figure 21). This could be the reason why the pooled sample performed far better (i.e. high number of called peaks). we leaveraged the sum of several ‘weaker’ signals to and were able to increase the power to call significant peaks (Park, 2009)

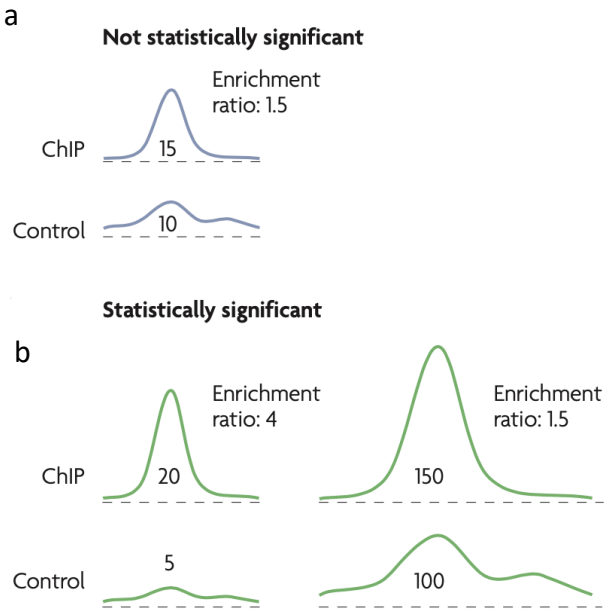


Figure 21. Illustration of statistically significant and not significant peaks. a. peak that is not statistically significant — the ratio between the ChIP and control sample is low (1.5) and the number of read counts (shown under the peaks) is also low. b. Two ways in which a peak can be statistically significant. On the left, although the number of read is low, the ratio between the ChIP and control sample is high (4). On the right, the peaks have the same enrichment ratio as those in (a) but have a larger number of reads enriched. Figure adapt from (Park, 2009, ChIP–seq: advantages and challenges of a maturing technology).

Another possible reason lead to such a different quality between the pooled and non-pooled samples might be DNA concentration measurement error caused by ultra-low DNA amount leading to bad adjustments of volume, ultimately lead to too low concentration of input DNA for the library preparation procedure. Pooling samples might therefore increase the DNA concentration to a level which keep the integrity of the fragment composition in the sample during the library preparation steps.

In conclusion, we find it unlikely that mistakes in the native-ChIP protocol itself or failed sequencing reactions can explain our failure to produce high quality biological replicates of H3K4me3 marks. The fact that pooled samples performed well indicate that the very low DNA-input concentration in most of the biological replicates combined with an overall high background noise is what caused failed ChIP-seq samples.

Reference:

ANDERSSON, R. & SANDELIN, A. 2019. Determinants of enhancer and promoter activities of regulatory elements. *Nature Reviews Genetics*, 1-17.

BAEZA-ROJANO, E., HACHERO-CRUZADO, I. & GUERRA-GARCÍA, J. M. 2014. Nutritional analysis of freshwater and marine amphipods from the Strait of Gibraltar and potential aquaculture applications. *Journal of sea research,* 85, 29-36.

BERGER, S. L. 2007. The complex language of chromatin regulation during transcription. *Nature,* 447, 407-412.

BRIND'AMOUR, J., LIU, S., HUDSON, M., CHEN, C., KARIMI, M. M. & LORINCZ, M. C. 2015. An ultra-low-input native ChIP-seq protocol for genome-wide profiling of rare cell populations. *Nature communications,* 6, 1-8.

CARMONA-ANTOÑANZAS, G., TOCHER, D. R., MARTINEZ-RUBIO, L. & LEAVER, M. J. 2014. Conservation of lipid metabolic gene transcriptional regulatory networks in fish and mammals. *Gene,* 534, 1-9.

CLARKE, W. C., WITHLER, R. E. & SHELBOURN, J. E. 1994. Inheritance of smolting phenotypes in backcrosses of hybrid stream-type× ocean-type chinook salmon (Oncorhynchus tshawytscha). *Estuaries,* 17, 13-25.

CLAUSE, K. C. & BARKER, T. H. 2013. Extracellular matrix signaling in morphogenesis and repair. *Current opinion in biotechnology,* 24, 830-833.

DAHMANN, C., OATES, A. C. & BRAND, M. 2011. Boundary formation and maintenance in tissue development. *Nature Reviews Genetics,* 12, 43-55.

DAVIDSON, W. S., KOOP, B. F., JONES, S. J., ITURRA, P., VIDAL, R., MAASS, A., JONASSEN, I., LIEN, S. & OMHOLT, S. W. 2010. Sequencing the genome of the Atlantic salmon (Salmo salar). *Genome biology,* 11, 403.

DU PREEZ, L. L. & PATTERTON, H.-G. 2013. Secondary Structures of the Core Histone N-terminal Tails: Their Role in Regulating Chromatin Structure. *Epigenetics: Development and Disease.* Dordrecht: Springer Netherlands.

FOOTE, C. J., WOOD, C. C., CLARKE, W. C. & BLACKBURN, J. 1992. Circannual cycle of seawater adaptability in Oncorhynchus nerka: genetic differences between sympatric sockeye salmon and kokanee. *Canadian Journal of Fisheries and Aquatic Sciences,* 49, 99-109.

GATES, L. A., FOULDS, C. E. & O'MALLEY, B. W. 2017. Histone marks in the 'driver's seat': functional roles in steering the transcription cycle. *Trends in biochemical sciences,* 42, 977-989.

GAVERY, M. R., NICHOLS, K. M., BEREJIKIAN, B. A., TATARA, C. P., GOETZ, G. W., DICKEY, J. T., VAN DOORNIK, D. M. & SWANSON, P. 2019. Temporal dynamics of DNA methylation patterns in response to rearing juvenile steelhead (Oncorhynchus mykiss) in a hatchery versus simulated stream environment. *Genes,* 10, 356.

GILLARD, G., HARVEY, T. N., GJUVSLAND, A., JIN, Y., THOMASSEN, M., LIEN, S., LEAVER, M., TORGERSEN, J. S., HVIDSTEN, T. R. & VIK, J. O. 2018. Life-stage-associated remodelling of lipid metabolism regulation in Atlantic salmon. *Molecular ecology,* 27, 1200-1213.

HARVEY, T. N. 2019. *Towards a genome wide understanding of salmon lipid metabolism gene regulation across tissues and life stages = Regulering av gener involvert i lipidmetabolisme i ulike vev og livsstadier i laks.* 2019:40, Norwegian University of Life Sciences, Faculty of Biosciences Cigene.

HIGGINS, P. 1985. Metabolic differences between Atlantic salmon (Salmo salar) parr and smolts. *Aquaculture,* 45**,** 33-53.
HO, J. W., BISHOP, E., KARCHENKO, P. V., NÈGRE, N., WHITE, K. P. & PARK, P. J. 2011. ChIP-chip versus ChIP-seq: lessons for experimental design and data analysis. *BMC genomics,* 12**,** 134.
HOAR, W. 1988. 4 The Physiology of Smolting Salmonids. *Fish physiology.* Elsevier.
HOWE, F. S., FISCHL, H., MURRAY, S. C. & MELLOR, J. 2017. Is H3K4me3 instructive for transcription activation? *Bioessays,* 39**,** 1-12.
HUISINGA, K. L., BROWER-TOLAND, B. & ELGIN, S. C. 2006. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma,* 115**,** 110-122.
IVERSEN, M., MULUGETA, T., GELLEIN BLIKENG, B., WEST, A. C., JØRGENSEN, E. H., RØD SANDVEN, S. & HAZLERIGG, D. 2020. RNA profiling identifies novel, photoperiod-history dependent markers associated with enhanced saltwater performance in juvenile Atlantic salmon. *PloS one,* 15**,** e0227496.
JAENISCH, R. & BIRD, A. 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature genetics,* 33**,** 245-254.
JOHANSSON, L.-H., TIMMERHAUS, G., AFANASYEV, S., JØRGENSEN, S. M. & KRASNOV, A. 2016. Smoltification and seawater transfer of Atlantic salmon (Salmo salar L.) is associated with systemic repression of the immune transcriptome. *Fish & shellfish immunology,* 58**,** 33-41.
KOUZARIDES, T. 2007. Chromatin modifications and their function. *Cell,* 128**,** 693-705.
LAKNA. 2017. *How do Transcription Factors Bind to DNA* [Online]. PEDIAA. Available: https://pediaa.com/how-do-transcription-factors-bind-to-dna/ [Accessed].
LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods,* 9**,** 357.
LAUBERTH, S. M., NAKAYAMA, T., WU, X., FERRIS, A. L., TANG, Z., HUGHES, S. H. & ROEDER, R. G. 2013. H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell,* 152**,** 1021-1036.
LEAVER, M. J., BAUTISTA, J. M., BJÖRNSSON, B. T., JÖNSSON, E., KREY, G., TOCHER, D. R. & TORSTENSEN, B. E. 2008. Towards fish lipid nutrigenomics: current state and prospects for fin-fish aquaculture. *Reviews in Fisheries Science,* 16**,** 73-94.
LORGEN, M., CASADEI, E., KRÓL, E., DOUGLAS, A., BIRNIE, M. J., EBBESSON, L. O., NILSEN, T. O., JORDAN, W. C., JØRGENSEN, E. H. & DARDENTE, H. 2015. Functional divergence of type 2 deiodinase paralogs in the Atlantic salmon. *Current Biology,* 25**,** 936-941.
MCCORMICK, S. D. 2012. Smolt physiology and endocrinology: Chapter 5.
MCCORMICK, S. D. & SAUNDERS, R. L. Preparatory physiological adaptations for marine life of salmonids: osmoregulation, growth, and metabolism. Am. Fish. Soc. Symp, 1987. 1-229.
MCCORMICK, S. D., SHRIMPTON, J. M., MORIYAMA, S. & BJÖRNSSON, B. T. 2007. Differential hormonal responses of Atlantic salmon parr and smolt to increased daylength: a possible developmental basis for smolting. *Aquaculture,* 273**,** 337-344.
MORÁN, P., MARCO-RIUS, F., MEGÍAS, M., COVELO-SOTO, L. & PÉREZ-FIGUEROA, A. 2013. Environmental induced methylation changes associated with seawater adaptation in brown trout. *Aquaculture,* 392**,** 77-83.
NICHOLS, K. M., EDO, A. F., WHEELER, P. A. & THORGAARD, G. H. 2008. The genetic basis of smoltification-related traits in Oncorhynchus mykiss. *Genetics,* 179**,** 1559-1575.

PARK, P. J. 2009. ChIP–seq: advantages and challenges of a maturing technology. *Nature reviews genetics,* 10**,** 669-680.
POKHOLOK, D. K., HARBISON, C. T., LEVINE, S., COLE, M., HANNETT, N. M., LEE, T. I., BELL, G. W., WALKER, K., ROLFE, P. A. & HERBOLSHEIMER, E. 2005. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell,* 122**,** 517-527.
PROJECT, E. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *nature,* 447**,** 799.
ROBERTSON, L. S. & MCCORMICK, S. D. 2012. Transcriptional profiling of the parr–smolt transformation in Atlantic salmon. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics,* 7**,** 351-360.
SANTOS-ROSA, H., SCHNEIDER, R., BANNISTER, A. J., SHERRIFF, J., BERNSTEIN, B. E., EMRE, N. T., SCHREIBER, S. L., MELLOR, J. & KOUZARIDES, T. 2002. Active genes are tri-methylated at K4 of histone H3. *Nature,* 419**,** 407-411.
SEEAR, P. J., CARMICHAEL, S. N., TALBOT, R., TAGGART, J. B., BRON, J. E. & SWEENEY, G. E. 2010. Differential gene expression during smoltification of Atlantic salmon (Salmo salar L.): a first large-scale microarray study. *Marine Biotechnology,* 12**,** 126-140.
SHANG, Y., SMITH, S. & HU, X. 2016. Role of Notch signaling in regulating innate immunity and inflammation in health and disease. *Protein & cell,* 7**,** 159-174.
SHERIDAN, M. A. 1989. Alterations in lipid metabolism accompanying smoltification and seawater adaptation of salmonid fish. *Aquaculture,* 82**,** 191-203.
VOIGT, P., TEE, W.-W. & REINBERG, D. 2013. A double take on bivalent promoters. *Genes & development,* 27**,** 1318-1338.
VOONG, L. N., XI, L., WANG, J.-P. & WANG, X. 2017. Genome-wide mapping of the nucleosome landscape by micrococcal nuclease and chemical mapping. *Trends in Genetics,* 33**,** 495-507.
WEDEMEYER, G. A., SAUNDERS, R. L. & CLARKE, W. C. 1980. *Environmental factors affecting smoltification and early marine survival of anadromous salmonids*, Department of Fisheries and Oceans, Biological Station.
YU, G., WANG, L.-G. & HE, Q.-Y. 2015. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics,* 31**,** 2382-2383.
ZYDLEWSKI, J. 1997. Temperature effects on osmoregulatory physiology of juvenile anadromous fish. *Global warming: Implications for freshwater and marine fish***,** 279.

Appendix

1 Buffer tables

diluted MNase 1:10	stock	mix	final conc
10 x Mnase buffer	10x	18 uL	1x
Mnase enzyme		2 uL	

MNase master mix *in uL*	stock	mix	final conc
10 x Mnase buffer	10x	6*sample	1x
200mM DTT		0.5*sample	
MNase dilution 1:10	1x	3*sample	

EDTA/Triton-Deoxycholate solution (in 10ml dH2O)	**stock**	**mix**	**final conc.**
EDTA	**0.5M**	**100ul**	**50mM**
Triton X-100	**100%**	**5ul**	**0.05ul**
sodium deoxycholate	**poder**	**0.005 g**	**0.05ul**
dH2O		**895ul**	

Complete immunoprecipitation buffer (100ml)	**stock**	**mix**	**final concentration**
Tris HCl, pH 8.0	1M	2 ml	20mM
EDTA	0.5 M	400 ul	2mM
Triton X-100	100%	100 ul	0,10%
NaCl	5M	3 ml	150mM
dH2O		94.5 ml	

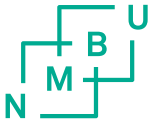
Low salt wash solution (100ml)	**stock**	**mix**	**final concentration**
Tris-HCl pH 8.0	1M	2 ml	20mM
EDTA	0.5M	400 ul	2mM
NaCL	5M	3 ml	150mM
Triton X-100	100%	1 ml	1%
SDS	20%	500 ul	0,10%
dH2O		93.1 ml	

High salt wash solution (100ml)	**stock**	**mix**	**final concentration**
Tris-HCl pH 8.0	1M	2 ml	20mM
EDTA	0.5M	400 ul	2mM
NaCL	5M	10 ml	500mM
Triton X-100	100%	1 ml	1%
SDS	20%	500 ul	0,10%
dH2O		86.1 ml	

ChIP elution buffer (100ml)	**stock**	**mix**	**final concentration**
NaHCO3 (natrium bicarbonate)		840.1mg	100mM
SDS	20%	5 ml	1%
dH2O		fill up to 100ml	

2. complete sequence quality.

Sample Name	% Dups	% GC	Length	M Seqs
19_1_1_R1	25.0%	44%	150 bp	26.3
19_1_1_R2	22.8%	44%	150 bp	26.3
19_1_1_trimmed_R1	24.2%	43%	139 bp	26.1
19_1_1_trimmed_R2	21.9%	43%	137 bp	26.1
19_2_2_R1	24.3%	44%	150 bp	27.5
19_2_2_R2	21.7%	44%	150 bp	27.5
19_2_2_trimmed_R1	23.8%	43%	140 bp	27.3
19_2_2_trimmed_R2	21.0%	43%	137 bp	27.3
19_2_4_R1	23.2%	44%	150 bp	26.4
19_2_4_R2	21.0%	44%	150 bp	26.4
19_2_4_trimmed_R1	22.9%	44%	141 bp	26.3
19_2_4_trimmed_R2	20.4%	44%	138 bp	26.3
19_pool_R1	33.3%	43%	150 bp	31.7
19_pool_R2	30.6%	43%	150 bp	31.7
19_pool_trimmed_R1	33.0%	42%	140 bp	31.6
19_pool_trimmed_R2	30.2%	42%	139 bp	31.6
1_pool_R1	28.6%	43%	150 bp	25.2
1_pool_R2	26.7%	44%	150 bp	25.2
1_pool_trimmed_R1	28.0%	43%	137 bp	25.1
1_pool_trimmed_R2	26.0%	43%	135 bp	25.1
25_1_4_R1	21.0%	44%	150 bp	23.2
25_1_4_R2	19.1%	44%	150 bp	23.2
25_1_4_trimmed_R1	20.6%	43%	139 bp	23.1
25_1_4_trimmed_R2	18.4%	43%	137 bp	23.1
25_2_1_R1	22.2%	44%	150 bp	23.0
25_2_1_R2	20.7%	44%	150 bp	23.0
25_2_1_trimmed_R1	21.8%	43%	138 bp	22.9
25_2_1_trimmed_R2	20.2%	43%	136 bp	22.9
25_2_2_R1	23.2%	44%	150 bp	24.4
25_2_2_R2	20.9%	45%	150 bp	24.4
25_2_2_trimmed_R1	22.9%	44%	140 bp	24.3
25_2_2_trimmed_R2	20.3%	44%	138 bp	24.3
25_2_4_R1	27.6%	45%	150 bp	27.9
25_2_4_R2	25.2%	45%	150 bp	27.9
25_2_4_trimmed_R1	27.0%	44%	138 bp	27.7
25_2_4_trimmed_R2	24.4%	44%	136 bp	27.7
Input_19_R1	22.9%	44%	150 bp	27.1
Input_19_R2	20.3%	44%	150 bp	27.1
Input_19_trimmed_R1	22.6%	44%	139 bp	27.1
Input_19_trimmed_R2	20.0%	44%	138 bp	27.1
Input_1_R1	22.6%	44%	150 bp	26.6
Input_1_R2	20.0%	45%	150 bp	26.6
Input_1_trimmed_R1	22.4%	44%	141 bp	26.6
Input_1_trimmed_R2	19.7%	44%	140 bp	26.6
Input_25_R1	22.3%	44%	150 bp	22.4
Input_25_R2	19.8%	44%	150 bp	22.4
Input_25_trimmed_R1	22.0%	43%	139 bp	22.4
Input_25_trimmed_R2	19.5%	44%	138 bp	22.4



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway