


Transcriptomic profiling of gill biopsies to define predictive markers for seawater survival in farmed Atlantic salmon

Lars Grønvold¹ | Mattis J. van Dalum² | Anja Striberny³ | Domniki Manousi¹ |
Trine Ytrestøl⁴ | Turid Mørkøre¹ | Solomon Boison⁵ | Bjarne Gjerde⁶ |
Even Jørgensen² | Simen R. Sandve¹  | David G. Hazlerigg²

¹Department of Animal and Aquaculture Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway

²Department of Arctic and Marine Biology, UiT – the Arctic University of Norway, Tromsø, Norway

³Department of Production Biology, Nofima, Tromsø, Norway

⁴Department of Nutrition and Feed Technology, Nofima, Tromsø, Norway

⁵MOWI Genetics AS, Bergen, Norway

⁶Department of Breeding and Genetics, Nofima, Ås, Norway

Correspondence

Simen R. Sandve, Department of Animal and Aquaculture Sciences, Faculty of Bioscience, Norwegian University of Life Sciences, Ås, Norway.

Email: simen.sandve@nmbu.no

David G. Hazlerigg, Department of Arctic and Marine Biology, UiT – the Arctic University of Norway, Tromsø, Norway.

Email: david.hazlerigg@uit.no

Funding information

Fiskeri - og havbruksnæringens forskningsfond, Grant/Award Number: 901589

Abstract

Wild Atlantic salmon migrate to sea following completion of a developmental process known as parr-smolt transformation (PST), which establishes a seawater (SW) tolerant phenotype. Effective imitation of this aspect of anadromous life history is a crucial aspect of commercial salmon production, with current industry practice being marred by significant losses during transition from the freshwater (FW) to SW phase of production. The natural photoperiodic control of PST can be mimicked by exposing farmed juvenile fish to a reduced duration photoperiod for at least 6 weeks before increasing the photoperiod in the last 1–2 months before SW transfer. While it is known that variations in this general protocol affect subsequent SW performance, there is no uniformly accepted industry standard; moreover, reliable prediction of SW performance from fish attributes in the FW phase remains a major challenge. Here we describe an experiment in which we took gill biopsies 1 week prior to SW transfer from 3000 individually tagged fish raised on three different photoperiod regimes during the FW phase. Biopsies were subjected to RNA profiling by Illumina sequencing, while individual fish growth and survival was monitored over 300 days in a SW cage environment, run as a common garden experiment. Using a random forest machine learning algorithm, we developed gene expression-based predictive models for initial survival and stunted growth in SW. Stunted growth phenotypes could not be predicted based on gill transcriptomes, but survival the first 40 days in SW could be predicted with moderate accuracy. While several previously identified marker genes contribute to this model, a surprisingly low weighting is ascribed to sodium potassium ATPase subunit genes, contradicting advocacy for their use as SW readiness markers. However, genes with photoperiod-history sensitive regulation were highly enriched among the genes with highest importance in the prediction model. This work opens new avenues for understanding and exploiting developmental changes in gill physiology during smolt development.

KEYWORDS

Atlantic salmon, gene expression, photoperiod, random forest model, seawater survival, smolt

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Author(s). *Journal of Fish Biology* published by John Wiley & Sons Ltd on behalf of Fisheries Society of the British Isles.

1 | INTRODUCTION

In anadromous salmonids, the transformation of freshwater (FW) resident juvenile fish (parr) into a migratory form (smolt) which goes to sea is known as smoltification, smolting or parr-smolt transformation (PST). PST entails a complex combination of physiological and behavioral changes, among which the acquisition of the ability to efficiently maintain water and ionic balance is critical (McCormick, 2013).

PST typically occurs after several years of slow growth in FW streams where spawning took place and requires parr to have exceeded a certain minimum threshold size at the end of the growing season preceding PST (Handeland et al., 2013; Sigholt et al., 1995; Strand et al., 2018; Thorpe, 1977). Exposure to a declining autumn photoperiod followed by an increasing photoperiod the following spring triggers the process of PST, apparently through a temporally coordinated sequence of endocrine change (McCormick, 2013). The photoperiod-dependent seasonal gating of PST ensures a synchronous springtime migration to sea, which is thought to reduce predation losses (Furey et al., 2016; Handeland et al., 1996).

Aquacultural production of Atlantic salmon (*Salmo salar*) depends on the ability to mimic natural PST to produce large numbers of sea-water (SW) tolerant juvenile fish for transfer to sea cages, where they can grow rapidly. There is no single universally accepted protocol for this commercial process. A widely used strategy for artificial smolt production is to rear hatchlings on a fast growth regime (i.e., continuous light [LL] and a fast growth diet for 6–12 months) to rapidly achieve threshold size, and then to expose them to a short photoperiod (SP) for several weeks before finally returning them to LL for the last 1–2 months of the FW phase. Based on observations of SW performance, it has been shown that the duration of exposure to SP should be at least 6 weeks long for LL to induce smolting (Duncan & Bromage, 1998). The mechanisms through which PST runs as a photoperiodic history-dependent process remain unknown, and untangling the role of SP exposure in smolt development is of considerable practical interest, as SP exposure reduces growth rates and slows aquaculture production (Imslund et al., 2014; Sigholt et al., 1998).

To overcome this delaying effect of SP exposure on production schedules, alternative strategies have emerged in which fish are maintained on constant light throughout the FW phase and SW tolerance is achieved through dietary manipulation with salt/tryptophan enriched diets, water salinity manipulation, or by increasing the size of fish (Striberny et al., 2021) in the run up to SW transfer. At present there is no unifying consensus on the relative merits of these different approaches, and difficulties in determining SW-readiness under different protocols probably contribute to industry losses during the SW-transfer phase of salmon production.

Because of its pivotal role in the regulation of water and ionic fluxes, undergoing a dramatic switch from a salt-conserving (FW) to a salt-excreting (SW) function, the gill has become the focus of efforts to optimize commercial smolt production (Takvam et al., 2024). Within the gill, mitochondria-rich cells (MRCs) are considered the primary drivers of ionic regulation (Evans, 2008), and PST includes a

pronounced shift in the location and phenotypic attributes of MRCs in the gill (McCormick, 2013; West et al., 2021). During PST, the gill complement of MRCs shifts from an ion-absorbing FW type to an ion-secreting SW type, and the distribution of MRCs shifts from the lamellae to the gill filament itself (Madsen et al., 2009; Pisam et al., 1988). One of the most studied molecular changes in the gill MRC during PST is the shift in expression of the genes encoding the ion regulatory Na^+ , K^+ -ATPase (NKA) pump and concurrent increase in NKA protein activity (reviewed in Takvam et al., 2024). Associated with this, relative levels of expression of two isoforms of the NKA α -subunit shifts during PST, with $\alpha 1a$ showing higher expression in parr and $\alpha 1b$ showing higher expression in smolts (McCormick et al., 2009).

From an applied perspective, characterization of gill transcriptomic changes during PST provides a potential route to designing marker-based predictive strategies for optimizing smolt production. Current approaches focus on relative expression of NKA alpha subunit isoforms (Takvam et al., 2024), but there is clearly potential to exploit other aspects of transcriptomic change to optimize predictive power. We have previously hypothesized that genes with photoperiod-history dependent regulation in the developing smolt gill would be interesting candidates for novel smolt status markers (Iversen et al., 2020). In this study we aim to test this, using a large-scale experiment in which gene expression prior to entering SW is compared with individual fish performance in the SW phase, including survival after SW transfer and stunted growth (i.e., loser fish (Noble et al., 2018)). We raised 3000 tagged fish on three different light regimes, produced RNAseq from gill biopsies 1 week prior to SW transfer, and subsequently recorded individual survival and growth in SW. We then applied a machine learning approach to this unique dataset, attempting to link SW performance to FW gill gene expression profiles. Our results demonstrate the potential for the development of novel and improved markers for smolt status in salmon aquaculture.

2 | MATERIALS AND METHODS

2.1 | Fish rearing experiment

For detailed description of the fish rearing see Gjerde et al. (2024). In brief, three groups of about 1000 juvenile Atlantic salmon (*Salmo salar*) were exposed to three different photoperiods; LL (24 h of light), 12:12 (12 h of light/24 h) and 8:16 (8 h of light/24 h). The LL group was kept from hatching until SW transfer on continuous light. The 8:16 group was reared on LL from hatching until the fish reached a body mass of ~50 g, after which they were exposed to a 6-week period of short photoperiod exposure (8 h of day light and 16 h of darkness [8L:16D]). After these 6 weeks followed another 8 weeks on LL before SW transfer when the fish reached a body mass of ~100 g. The 12:12 group was exposed to 12 h of day light and 12 h of darkness for 6 weeks, following 8 weeks in LL before SW transfer. One week prior to SW transfer, gill biopsies were sampled and smolt index

(i.e., skin coloration), and body weight and length were measured. The smolt index score is a categorical scoring of skin coloration; parr marks (score = 1), a mix of silvery skin with residual weak parr marks (score = 2), and completely silvery skin (score = 3).

Approximately 1 week after gill biopsies were taken, the fish were moved to one common SW cage where they were kept for 10 months and daily recording of mortality was carried out.

2.2 | Ethics statement

The experiment was performed according to EU regulations concerning the protection of experimental animals (Directive 2010/63/EU). Appropriate measures were taken to minimize pain and discomfort. The experiment was approved by the Norwegian Food and Safety Authority (FOTS id number 25658).

2.3 | RNAseq data generation

Gill biopsies were flash frozen on dry ice and sent to Qiagen for RNA isolation using the RNeasy Fibrous Tissue Mini kit. Isolated RNA samples were then shipped to Novogene for RNA-sequencing library construction which were sequenced using 2×150 bp pair end illumina sequencing.

2.4 | RNAseq read mapping

Prior to transcript quantification, adaptor sequences were trimmed off using fastp version 0.23.2 (Chen et al., 2018). Read mapping and transcript quantification were done using Salmon version 1.1.0 (Patro et al., 2017) with the Atlantic salmon transcriptome annotation from ENSEMBL from the Ssal_v3.1 genome assembly (Ssal_v3.1, GCA_905237065.2) in the selective mapping-mode. A transcriptome index file was used as a decoy sequence to avoid the incorrect mapping of reads between annotated and unannotated yet highly similar genomic regions (Srivastava et al., 2020). The options --keepDuplicates and --gcBias were used for transcript quantification to account for the high sequence similarity between some of the duplicated genes from the salmonid genome duplication (Lien et al., 2016) and correct for fragment-level GC biases of aligned reads, respectively. To calculate mRNA expression at the gene-level from output files from salmon the sum of the raw read counts and normalized gene expression values (transcripts per million) were done with the R package tximport (Soneson et al., 2015).

2.5 | Differential gene expression analyses between samples with different photoperiod history

We carried out three different analyses of differential gene expression using gill RNAseq data to define a set of core genes whose

expression during PST in the gill consistently depends on photoperiodic history in the FW phase. Two of these analyses used previously published datasets (Iversen et al., 2020, 2021) (experiments 1 and 2) and the third analysis used data from this study (experiment 3) (Figure S1). In experiment 1 (Figure S1A) we contrasted 7-month-old salmon (reared in FW) either raised on continuous light (LL group, $n = 6$) or exposed to 60 days of short days (8L:16D) followed by 50 days in LL (SPLL group, $n = 6$). In experiment 2 (Figure S1B) we contrasted 11-month-old salmon (reared in FW) exposed to 2 weeks of short photoperiod (2-week group, 8L:16D, $n = 6$) with fish exposed to 8 weeks of short photoperiod (8-week group, $n = 6$). Both groups experienced an 8-week period on continuous light after the short photoperiod. In experiment 3 (Figure S1C) we contrasted smolt raised on continuous light (LL, $n = 981$) with smolt exposed to 6 weeks of short photoperiod treatments (8L:16D, $n = 927$).

Differential expression in experiments 1 and 2 was analyzed with EdgeR version 3.42.4. Genes with low expression were filtered (the minimum number of reads summed across all fish per group was 10) and normalized using trimmed means of M values. A quasi-likelihood negative binomial generalized log-linear model with 0 as intercept was fitted to the data. Differentially expressed genes (DEGs) were identified in pairwise contrasts using empirical Bayes F tests. The p values were adjusted for multiple testing using the Benjamini & Hochberg (BH) (Benjamini & Hockberg, 1995) method implemented in the `p.adjust()` function in R. Since large sample numbers have been shown to exaggerate false-positive rates when applying standard differential gene expression methods (Li et al., 2022), DEGs from experiment three were identified using the Wilcoxon rank-sum test. False-discovery rate thresholds were defined as the square root of the minimum sample size per group, resulting in adjusted p value cut-offs of ≤ 0.05 for experiments 1 and 2, and ≤ 0.002 for experiment 3. Finally, upset plots of shared significant DEGs were made with the R package upsetR (Conway et al., 2017) to define a core set of photoperiodic history-dependent smolt development genes identified as DEGs in all three analyses.

2.6 | Differential gene expression between fish with different SW performance

DEG analyses was also used to identify gene expression differences in gill samples from FW between smolts with different growth and mortality outcomes in SW. First, we contrasted gene expression in fish from the LL group which died within 40 days post SW transfer versus fish that were still alive in November, approximately 5 months after transfer (i.e., excluding fish that died of disease or were otherwise lost). Second, we contrasted gene expression differences between “loser” fish with extremely low growth and “winners” (i.e., fish with normal growth r). Classification of losers was based on the growth rate (i.e., gained weight divided by initial weight) where the threshold of 0.5 was decided based on the minimum between the two peaks in the growth rate distribution (see Figure 2a). Read counts were normalized across samples using EdgeR's TMM method and then converted

to Counts Per Million (CPM). These values were subsequently log-transformed to $\log_2(\text{CPM} + 1)$. Genes with zero reads in more than 90% of the samples were excluded, resulting in the removal of approximately 9000 out of 47000 genes. The final dataset of log-transformed expression values was used for Wilcoxon rank-sum tests. Fold change values were calculated as the difference of the mean log-transformed expression values of the contrasted samples. The p values were corrected using the BH method (Benjamini & Hochberg, 1995) implemented in the `p.adjust()` function in R.

2.7 | Random forest prediction

To build prediction models for SW survival and growth using FW gill gene expression data from this study (experiment 3) we used the `RandomForestClassifier` from the `sklearn` python module (Pedregosa et al., 2011) with parameters `n_estimators = 1000` and `class_weight = "balanced"` (otherwise defaults). We included the log-transformed gene expression as well as the commonly used smolt marker NKA ratio as features. NKA ratio was calculated by subtracting the log transformed gene expression of NKA $\alpha 1a$ (ENSSSAG00000088099) from NKA $\alpha 1b$ (ENSSSAG00000041746). This is equivalent to dividing $\alpha 1b/\alpha 1a$ in the non-transformed expression values. Samples were randomly split into 80% training samples and 20% test samples. This was repeated 100 times, that is, random split and model training, recording the importance scores of the genes and the prediction probabilities of each sample. The mean of the importance scores and the prediction probabilities of the test samples are reported in the analyses. We performed random forest (RF) analyses for both “early death” and “loser fish” outcomes. Both early death and loser fish were defined as outlined for the DEG analyses.

2.8 | Identification of gill cell-type specific genes using single nuclei RNAseq

To identify genes specific to different cell types in gill tissues, we utilized the single nuclei RNA sequencing (snRNAseq) data provided by West et al. (2021). This dataset comprises snRNAseq analyses performed on gill tissues sampled at various stages of smoltification under different photoperiod regimes. Nuclei were assigned to 20 clusters based on their expression patterns using Seurat (Hao et al., 2024). Each cluster was subsequently annotated with a cell type using known marker genes and gene ontology (GO) analysis.

The expression levels of 2000 variable genes were averaged across the nuclei in each cell-type cluster and each gene was assigned to the cell type of the cluster with the highest average expression level. Since the original gene annotations from West et al. (2021) were based on NCBI identifiers, we converted these to Ensembl gene IDs using the `g:Profiler's g:Convert` utility (Kolberg et al., 2023). This conversion resulted in 1354 cell-type specific Ensembl genes mapped from 1271 NCBI genes (i.e., not all genes could be mapped and some had multiple matches).

2.9 | Cell-type enrichment analysis

Fisher's exact test using the `fisher.test()` in R (R Core Team, 2021) was used to assess whether the DEGs in early death are associated with certain cell types. DEGs were identified based on their importance scores, with those exceeding $5e-5$ (top ~10%) being classified as DEGs. These DEGs were further divided into upregulated or downregulated based on the fold change. For each cell type, we evaluated the independence between a gene's classification as a DEG and its association with that particular cell type. An odds ratio greater than 1 from this test suggests an enrichment of DEGs that are specifically expressed in that cell type. Specifically, for upregulated DEGs, it suggests a higher abundance of the corresponding cell type in the gill samples from fish that died early and, conversely, for downregulated DEGs, it indicates a greater abundance in the surviving fish. The p values were adjusted for multiple comparisons using the `p.adjust()` function in R (R Core Team, 2021) with the Benjamini & Hochberg method (Benjamini and Hochberg, 1995).

3 | RESULTS

3.1 | Defining a core set of photoperiodic history-dependent smolt gill development genes

Our reanalysis of the datasets from Iversen et al. (2020, 2021) (experiments 1 and 2) identified 781 and 6923 genes with significantly different expression levels between gill samples from fish with different photoperiodic history during the PST (Figure S1D). DEG analyses between gills from fish with different photoperiodic histories in this study (i.e., LL vs. 8L:16D and LL vs. 12L:12D) identified a total of 14209 genes which were DEGs in one or both contrasts (Figure S1D and Table S1). Finally, we intersected these three analyses to define a core set of 219 genes whose expression during PST gill consistently depends on photoperiodic history in the FW phase (Table S2).

3.2 | Random forest prediction models for mortality and growth stagnation

The main aim of this study was to test if smolt gill gene expression profiles in fresh water could be used to predict SW performance at the individual fish level, and secondarily to determine whether candidate markers defined by their sensitivity to photoperiodic history constituted useful SW performance predictors. We made RF prediction models for two SW outcomes; early mortality (<40 days after SW transfer) and stagnation of growth, referred to as loser fish.

Early mortality was substantial (>8%; Figure 1a) in the LL group, while extremely low in both groups of fish exposed to reduced photoperiods during the FW phase (12L:12D, $n = 8$ fish = 0.8% mortality) and (8L:16D, $n = 6$ fish = 0.6% mortality). Hence, we only used the LL group to develop the RF prediction model for early mortality.

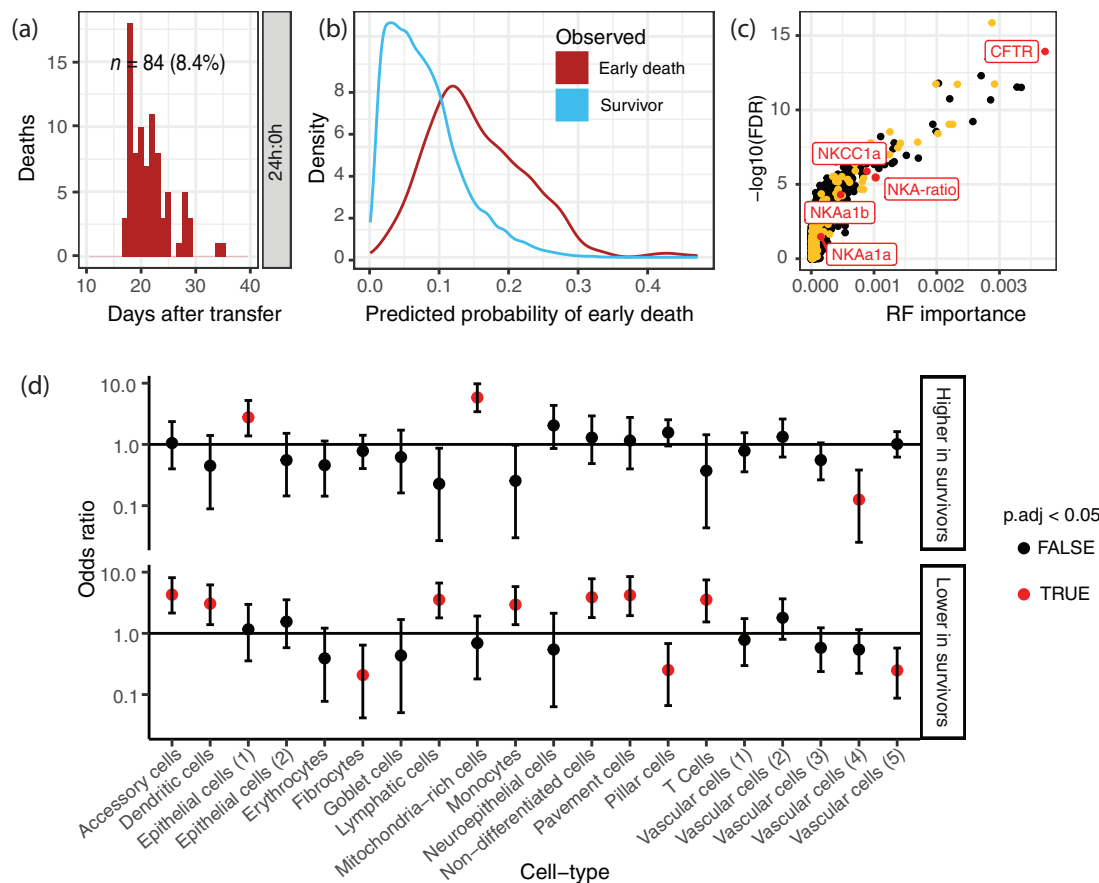


FIGURE 1 Random forest predictions of early mortality in seawater. (a) Early mortality cases in LL group. (b) Random forest model predictions for individual fish probability for mortality in an early phase (<40 days) following seawater transfer. Predictions are based on gill tissue gene expression profiles from smolts in fresh water prior to seawater transfer. (c) Scatterplot of random forest model importance scores versus differential gene expression adjusted p values from a contrast between fish dying early after seawater transfer and surviving fish. Yellow coloured dots represents genes with photoperiod history dependent regulation. Red dots represents classical smolt physiology states associated genes. (d) Fisher test odds-ratios (i.e., enrichment) for cell type biased/specific genes which are expressed higher and lower in surviving fish compared to fish dying early (<40 days) after seawater transfer.

We tested the RF model on random subsets of LL fish not used in model training. This demonstrated good prediction capability (Figure 1b) with an average area under the receiver operating characteristic (auROC) of 0.82. In other words, in 82% of the predictions the model will correctly assign a higher absolute risk of mortality to a randomly selected fish that will die compared to a randomly selected fish which will not die.

Although an auROC >0.80 is considered to be clinically useful in medicine (Çorbacıoğlu & Aksel, 2023), this prediction model will have relatively high false-positive rate. Nevertheless, a ranked list of the importance scores of the RF model features (i.e., gene expression levels) can also be used to gain insights into biological processes which contribute to SW performance. We therefore performed an in-depth analysis of the gene importance scores. Interestingly, the most used molecular marker for smolt status, the two individual genes forming the NKA-subunits (NKA α 1a and NKA α 1b), had low importance scores in the prediction model (Figure 1c). We also included a proxy for the commercially available SmoltVision test (Pharmaq Analytic) which is the ratio of mRNA levels for the two NKA-subunits

(NKA α 1b:NKA α 1a). Surprisingly, this NKA-ratio feature was only the 49th most important feature in the RF model (Figure 1c and Table S1, S3). However, another well-known smolt-gill gene, the cystic fibrosis transmembrane regulator (CFTR) (Hiroi & McCormick, 2012), had the highest importance in our RF model (Figure 1c). In addition, the top predictive genes (importance >0.001, 37 genes; Table S3) also contained other genes linked to ion transport such as the sodium/potassium-transporting ATPase subunit alpha-3 and the *patj* gene, which is involved in tight junction formation in mammals (Shin et al., 2005).

We then explored the predictive importance of genes with photoperiod-history dependent regulation (intersection between DEGs in experiments 1 and 2, and this study; see Figure S1) for initial SW survival. Among the 219 photoperiodic history-dependent genes, 48% were in the top 10% most important genes in the RF model (yellow points in Figure 1c). This was a significantly higher percentage than expected by chance (Fisher test p value = 9.53×10^{-47}).

The development of the smolt gill involves a pronounced remodeling of its cellular composition, notably in terms of its complement of chloride cells and of immune cell types (McCormick, 2013; West

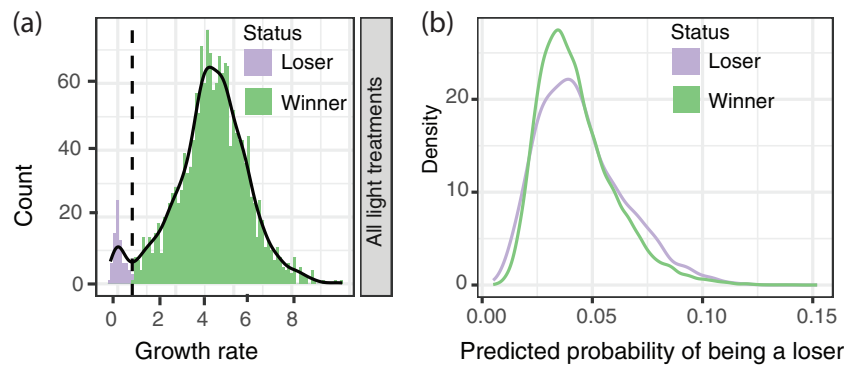


FIGURE 2 Random forest predictions of loser fish status in seawater. (a) Classification of loser fish and winner fish based on growth rates in seawater. (b) Random forest model predictions for individual fish classification of loser fish (low or no growth in seawater) and winner fish (normal growth rates) following seawater transfer. Scatterplot of random forest model importance scores versus differential gene expression adjusted p values from a contrast between fish classified as loser fish (low or no growth in seawater) and winner fish (normal growth rates) following seawater transfer.

et al., 2021). Leveraging recently published data on gill cell type-specific markers (West et al., 2021), we linked gene expression differences between surviving fish and those which died early after SW transfer to the cell types these DEGs were associated with. Genes associated with mitochondria-rich cells and one subtype of epithelial cells (1) were significantly more likely ($p < 0.05$) to be expressed more strongly in survivors compared to fish dying within 40 days of SW transfer (Figure 1d). Genes associated with seven cell types, including three immune-system related cells, were significantly more likely ($p < 0.05$) to be expressed at lower levels in survivors compared to fish that died within 40 days of SW transfer (Figure 1d).

Finally, we used a similar RF approach to predict loser fish status, which is a common challenge in aquaculture. Short photoperiod exposure protocols were not associated with loser fish status, hence we combined all loser fish in one group (Figure 2a). Our results clearly demonstrate that it is not possible to use gill gene expression in the FW phase to train an RF model to predict loser fish status in SW (Figure 2b).

3.3 | Including external smolt characteristics in the RF model

External characteristics such as size, condition factor, and skin coloration are used to evaluate smolt status and development in salmon aquaculture (Ytrestøl et al., 2023). Although our study focused on transcriptome-based markers, we also build an RF model including size and a smolt index score in June prior to SW transfer.

Our results showed that body weight had the highest model importance score for early SW mortality and that skin coloration had higher model importance than the commercially used NKA ratio (Figure 3a). Condition factor ranked below the NKA ratio in model importance (Figure 3a). It is interesting to note that even though the LL group had 10 times higher mortality in early SW phase, the

proportion of silvery colored fish among fish that died was similar across all smolt production groups (Figure 3b).

4 | DISCUSSION

In the present study we aimed to identify and validate gene expression markers for smolt status by using gill gene expression at the end of smolt rearing in FW to predict survival and performance in the SW phase. We could not predict loser fish outcomes based on FW gill transcriptomes, indicating that this phenomenon is not mechanistically linked to smolt gill development and ion regulation ability. Conversely, the prediction model for SW survival had reasonable discriminatory ability (auROC = 0.84), and results from the RF model (Figure 2c) provide novel biological insights into the smolt gill and development of new biomarkers for smolt status.

The most common molecular biomarker for smolt status is the NKA-subunit gene expression ratio (NKA α 1b:NKA α 1a) (reviewed in Takvam et al., 2024). Surprisingly, in this study (Figure 1c) the NKA ratio was only ranked as the 35th most important feature in the prediction model for early mortality in SW. This result is in line with a recent experiment by Kahw et al. (2021), who found large variation in NKA ratio related to fish genetics and smolt production protocols (0+ vs. 1+) and concluded that NKA ratio was not a robust biomarker for smolt status. Hence, Khaw et al. (2021) and our findings in this study question the reliability of the NKA ratio as a good predictor for SW survival and performance in aquaculture smolt production.

Our large-scale transcriptomic approach using gene expression profiles from individual fish represents an important step towards identifying new and improved biomarkers for smolt status. One interesting aspect of our list of putative smolt biomarkers is the highly significant overrepresentation of core photoperiod-history sensitive genes (Fisher test, p value = $9.529355e-47$; Figure 1c). The most predictive gene CFTR (Figure 1c) was not in the core photoperiod-history sensitive gene set, however CFTR was deemed photoperiodic-history sensitive in two out of three experiments (Table S1). Among the top

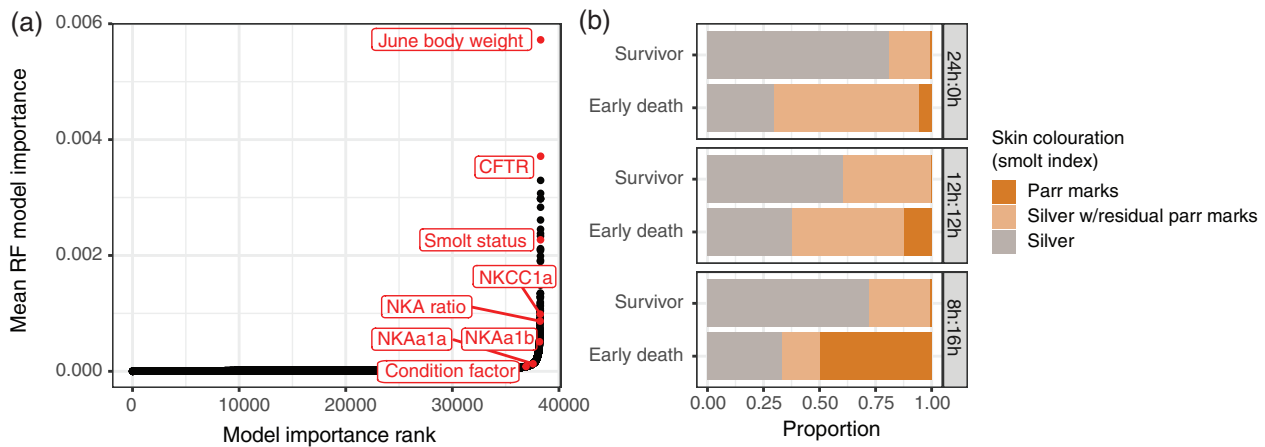


FIGURE 3 Random forest predictions including external smolt characteristics. (a) Feature importance plotted for random forest model for early mortality in sea water. This model included three external smolt characteristics: body size, condition factor, and skin coloration (i.e., smolt index). (b) Proportions of smolt index scores in the three smolt groups reared under different photoperiod histories.

37' features with highest prediction model importance (>0.001 , Table S3), 12 were in the core photoperiod-sensitive gene list, including calpain 2, *egln2*, and *TEFa*. Interestingly, *calpain2* was ranked as the gene with the strongest photoperiod-history sensitive regulation during PST in Iversen et al. (2020). *Egln2* is known to interact with hypoxia inducible factors (HIFs) in human studies according to the Molecular Interaction Database (<https://www.ebi.ac.uk/intact/>) and in zebrafish larvae there is a bidirectional crosstalk between HIFs and glucocorticoid signaling (Marchi et al., 2020). Furthermore, the D-element transcription factor *TEFa*, which is implicated in light-circadian interactions in zebrafish (Vatine et al., 2009), is a core part of photoperiod transduction in mammals (Dardente et al., 2010) and is directly responsive to cortisol in salmon (West et al., 2020). The presence of *egln2* and *TEFa* among the most predictive genes thus suggests a glucocorticoid- signaling link to PST gill development and initial SW survival. In conclusion, our results point to a mechanistic link between poor smolt development in the LL group and dysregulation of molecular developmental processes that are naturally regulated by photoperiod signals. Focusing on these “winter signal” affected genes is therefore an exciting avenue in future research and validation of novel biomarkers for smolt physiology.

In the models including external features, smolt size were ranked higher in model feature importance compared to any of the gene expression phenotypes (Figure 3a). This is not surprising as the ability to tolerate and survive osmotic stress increases with fish size, irrespective of smolt development (Bjerknes et al., 1992; Parry, 1960). Skin coloration and condition factor did, however, rank lower than many gene expression phenotypes in their model importance score, and the increased mortality in the LL group was not associated with a lower proportion of fish with silvery skin color (Figure 3b). This aligns with the notion that development of smolt physiology happens through a suite of parallel and independent processes, and that not all are causally linked to the molecular basis for SW tolerance.

One caveat related to our prediction model results is that the data we used to build the model were limited to data from fish raised on

LL. This was because initial survival was $>99\%$ in smolts raised using exposure to short photoperiods. We did try to apply our model to the limited number of fish in the 8L:16D and 12L:12D groups that suffered early SW mortality, which showed that the fish with highest probability of early SW mortality did in fact die (Figure S2). However there was not a significant difference in the distribution of probability of mortality for the few fish that did die compared to the vast majority that survived in these groups. Even though an LL protocol can compromise the development of hypo-osmoregulatory ability as well as other smolt development characteristics, our results are still of high relevance to the industry due to extensive use of LL in smolt production.

AUTHOR CONTRIBUTIONS

Mattis J. van Dalum: Investigation, formal analysis, visualization, writing. Bjarne Gjerde: Conceptualization, experimental design, manuscript editing. Solomon Boison: Conceptualization, experimental design, manuscript editing. Lars Grønvald: Investigation, formal analysis, writing, manuscript editing. David G. Hazlerigg: Conceptualization, funding acquisition, writing and editing, supervision. Even Jørgensen: Fish rearing and sampling, manuscript editing. Turid Mørkøre: Fish phenotype recording, manuscript editing. Simen R. Sandve: Conceptualization, funding acquisition, writing and editing, supervision. Anja Striberny: Fish rearing and sampling, writing. Domniki Manousi: Bioinformatic data analyses, manuscript editing. Trine Ytrestøy: Fish phenotype recording, manuscript editing.

ACKNOWLEDGEMENTS

Thanks to MOWI for providing fish for this experiment and Sofie Robertson for exploring the use of RF prediction models using gene expression data.

FUNDING INFORMATION

This study was funded by the Norwegian Seafood Research Fund (project no. 901589).

DATA AVAILABILITY STATEMENT

All the reads from transcriptome data from the new experiments presented in this paper are deposited at ENA under the project number ERP131717. All scripts and data to reproduce figures and statistical tests can be deposited at zenodo (<https://doi.org/10.5281/zenodo.14177709>).

ORCID

Simen R. Sandve  <https://orcid.org/0000-0003-4989-5311>

REFERENCES

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.
- Bjerknes, V., Duston, J., Knox, D., & Harmon, P. (1992). Importance of body size for acclimation of underyearling Atlantic salmon parr (*Salmo salar* L.) to seawater. *Aquaculture*, 104, 357–366.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34, i884–i890.
- Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, 33, 2938–2940.
- Çorbacıoğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*, 23, 195–198.
- Dardente, H., Wyse, C. A., Birnie, M. J., Dupré, S. M., Loudon, A. S. I., Lincoln, G. A., & Hazlerigg, D. G. (2010). A molecular switch for photoperiod responsiveness in mammals. *Current Biology*, 20, 2193–2198.
- Duncan, N. J., & Bromage, N. (1998). The effect of different periods of constant short days on smoltification in juvenile Atlantic salmon (*Salmo salar*). *Aquaculture*, 168, 369–386.
- Evans, D. H. (2008). Teleost fish osmoregulation: What have we learned since August Krogh, Homer Smith, and Ancel Keys. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, 295, R704–R713.
- Furey, N. B., Hinch, S. G., Bass, A. L., Middleton, C. T., Minke-Martin, V., & Lotto, A. G. (2016). Predator swamping reduces predation risk during nocturnal migration of juvenile salmon in a high-mortality landscape. *The Journal of Animal Ecology*, 85, 948–959.
- Gjerde, B., Boison, S. A., Hazlerigg, D., Ytrestoyl, T., Morkore, T., Jørgensen, E., Striberny, A., & Sandve, S. R. (2024). Impact of three light smoltification regimes on performance and genetic parameters of traits in Atlantic salmon. *BioRxiv*.
- Handeland, S. O., Imsland, A. K., Björnsson, B. T., & Stefansson, S. O. (2013). Long-term effects of photoperiod, temperature and their interaction on growth, gill Na⁺, K⁺-ATPase activity, seawater tolerance and plasma growth-hormone levels in Atlantic salmon *Salmo salar*. *Journal of Fish Biology*, 83, 1197–1209.
- Handeland, S. O., Järvi, T., Fernö, A., & Stefansson, S. O. (1996). Osmotic stress, antipredatory behaviour, and mortality of Atlantic salmon (*Salmo salar*) smolts. *Canadian Journal of Fisheries and Aquatic Sciences*, 53, 2673–2680.
- Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., & Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*, 42, 293–304.
- Hiroi, J., & McCormick, S. D. (2012). New insights into gill ionocyte and ion transporter function in euryhaline and diadromous fish. *Respiratory Physiology & Neurobiology*, 184, 257–268.
- Imsland, A. K., Handeland, S. O., & Stefansson, S. O. (2014). Photoperiod and temperature effects on growth and maturation of pre- and post-smolt Atlantic salmon. *Aquaculture International*, 22, 1331–1345.
- 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.
- Iversen, M., Mulugeta, T., West, A., Jørgensen, E. H., Martin, S. A. M., Sandve, S. R., & Hazlerigg, D. (2021). Photoperiod-dependent developmental reprogramming of the transcriptional response to seawater entry in Atlantic salmon (*Salmo salar*). *G3 (Bethesda, Md.)*, 11(4), jkab072.
- Khaw, H. L., Gjerde, B., Boison, S. A., Hjelle, E., & Difford, G. F. (2021). Quantitative genetics of smoltification status at the time of seawater transfer in Atlantic salmon (*salmo salar*). *Frontiers in genetics*, 12, 696893.
- Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). G:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research*, 51, W207–W212.
- Li, Y., Ge, X., Peng, F., Li, W., & Li, J. J. (2022). Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biology*, 23, 79.
- Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., Leong, J. S., Minkley, D. R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R. A., von Schalburg, K., Rondeau, E. B., Di Genova, A., Samy, J. K., ... Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533, 200–205.
- Madsen, S. S., Killeirich, P., & Tipsmark, C. K. (2009). Multiplicity of expression of Na⁺,K⁺-ATPase α -subunit isoforms in the gill of Atlantic salmon (*Salmo salar*): Cellular localisation and absolute quantification in response to salinity change. *Journal of Experimental Biology*, 212, 78–88.
- Marchi, D., Santhakumar, K., Markham, E., Li, N., Storbeck, K.-H., Krone, N., Cunliffe, V. T., & van Eeden, F. J. M. (2020). Bidirectional crosstalk between hypoxia-inducible factor and glucocorticoid signaling in zebrafish larvae. *PLoS Genetics*, 16, e1008757.
- McCormick, S. D. (2013). *Smolt physiology and Endocrinology*. Fish Physiology.
- McCormick, S. D., Regish, A. M., & Christensen, A. K. (2009). Distinct freshwater and seawater isoforms of Na⁺/K⁺-ATPase in gill chloride cells of Atlantic salmon. *Journal of Experimental Biology*, 212, 3994–4001.
- Noble, C., Gismervik, K., Iversen, M. H., & Kolarevic, J. (2018). *Welfare indicators for farmed Atlantic salmon: Tools for assessing fish welfare*. Nofima report, Nofima. <https://nordopen.nord.no/nord-xmliui/handle/11250/2575780>
- Parry, G. (1960). The development of salinity tolerance in the salmon, *Salmo salar* (L.) and some related species. *Journal of Experimental Biology*, 37, 425–434.
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14, 417–419.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85), 2825–2830.
- Pisam, M., Prunet, P., Boeuf, G., & Rambourg, A. (1988). Ultrastructural features of chloride cells in the gill epithelium of the Atlantic salmon, *Salmo salar*, and their modifications during smoltification. *The American Journal of Anatomy*, 183, 235–244.

- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Shin, K., Straight, S., & Margolis, B. (2005). PATJ regulates tight junction formation and polarity in mammalian epithelial cells. *The Journal of Cell Biology*, *168*, 705–711.
- Sigholt, T., Åsgård, T., & Staurnes, M. (1998). Timing of parr-smolt transformation in Atlantic salmon (*Salmo salar*): Effects of changes in temperature and photoperiod. *Aquaculture*, *160*, 129–144.
- Sigholt, T., Staurnes, M., Jakobsen, H. J., & Åsgård, T. (1995). Effects of continuous light and short-day photoperiod on smolting, seawater survival and growth in Atlantic salmon (*Salmo salar*). *Aquaculture*, *130*, 373–388.
- Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521.
- Srivastava, A., Malik, L., Sarkar, H., Zakeri, M., Almodaresi, F., Soneson, C., Love, M. I., Kingsford, C., & Patro, R. (2020). Alignment and mapping methodology influence transcript abundance estimation. *Genome Biology*, *21*, 239.
- Strand, J. E. T., Hazlerigg, D., & Jørgensen, E. H. (2018). Photoperiod revisited: Is there a critical day length for triggering a complete parr-smolt transformation in Atlantic salmon *Salmo salar*? *Journal of Fish Biology*, *93*, 440–448.
- Striberny, A., Lauritzen, D. E., Fuentes, J., Campinho, M. A., Gaetano, P., Duarte, V., Hazlerigg, D. G., & Jørgensen, E. H. (2021). More than one way to smoltify a salmon? Effects of dietary and light treatment on smolt development and seawater growth performance in Atlantic salmon. *Aquaculture*, *532*, 736044.
- Takvam, M., Sundell, K., Sundh, H., Gharbi, N., Kryvi, H., & Nilsen, T. O. (2024). New wine in old bottles: Modification of the Na⁺/K⁺-ATPase enzyme activity assay and its application in salmonid aquaculture. *Reviews in Aquaculture*, *16*, 1087–1098.
- Thorpe, J. E. (1977). Bimodal distribution of length of juvenile Atlantic salmon (*Salmo salar* L.) under artificial rearing conditions. *Journal of Fish Biology*, *11*, 175–184.
- Vatine, G., Vallone, D., Appelbaum, L., Mracek, P., Ben-Moshe, Z., Lahiri, K., Gothilf, Y., & Foulkes, N. S. (2009). Light directs zebrafish period2 expression via conserved D and E boxes. *PLoS Biology*, *7*, e1000223.
- West, A. C., Iversen, M., Jørgensen, E. H., Sandve, S. R., Hazlerigg, D. G., & Wood, S. H. (2020). Diversified regulation of circadian clock gene expression following whole genome duplication. *PLoS Genetics*, *16*, e1009097.
- West, A. C., Mizoro, Y., Wood, S. H., Ince, L. M., Iversen, M., Jørgensen, E. H., Nome, T., Sandve, S. R., Martin, S. A. M., Loudon, A. S. I., & Hazlerigg, D. G. (2021). Immunologic profiling of the Atlantic salmon gill by single nuclei transcriptomics. *Frontiers in Immunology*, *12*, 669889.
- Ytrestøyl, T., Jenssen, I., Wetterwald, V. E., Striberny, A., Alvestad, R., Dam, R., Mortensen, H. S., Johansen, E., Espmark, Å. M. O., Johansen, L. H., Kolarevic, J., Larsen, J. S., Sandberg, M. G., & Nilsen, T. O. (2023). *Kunnskapskartlegging-produksjon av stor laksesmolt*. Nofima AS.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Grønvold, L., van Dalum, M. J., Striberny, A., Manousi, D., Ytrestøyl, T., Mørkøre, T., Boison, S., Gjerde, B., Jørgensen, E., Sandve, S. R., & Hazlerigg, D. G. (2024). Transcriptomic profiling of gill biopsies to define predictive markers for seawater survival in farmed Atlantic salmon. *Journal of Fish Biology*, 1–9. <https://doi.org/10.1111/jfb.16025>