

Published in final edited form as:

Genomics. 2021 June 30; 113(5): 2953–2964. doi:10.1016/j.ygeno.2021.06.037.

Epigenetic mapping of the somatotrophic axis in Nile tilapia reveals differential DNA hydroxymethylation marks associated with growth

Ioannis Konstantinidis^a, Dafni Anastasiadi^{b,c}, Pål Sætrom^{a,d,e,f,g}, Artem V. Nedoluzhko^a, Robin Mjelle^{a,d}, Tomasz Podgorniak^a, Francesc Piferrer^c, Jorge M.O. Fernandes^{a,*}

^aFaculty of Biosciences and Aquaculture, Nord University, Bodø, Norway ^bThe New Zealand Institute for Plant and Food Research, Nelson, New Zealand ^cInstitut de Ciències del Mar, Spanish National Research Council (CSIC), Barcelona, Spain ^dDepartment of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway ^eDepartment of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway ^fBioinformatics core facility-BioCore, Norwegian University of Science and Technology, Trondheim, Norway ^gK.G. Jebsen Center for Genetic Epidemiology, Norwegian University of Science and Technology, Trondheim, Norway

Abstract

In vertebrates, the somatotrophic axis comprising the pituitary gland, liver and muscle plays a major role in myogenesis. Its output in terms of muscle growth is highly affected by nutritional and environmental cues, and thus likely epigenetically regulated. Hydroxymethylation is emerging as a DNA modification that modulates gene expression but a holistic characterization of the hydroxymethylome of the somatotrophic axis has not been investigated to date. Using reduced representation 5-hydroxymethylcytosine profiling we demonstrate tissue-specific localization of 5-hydroxymethylcytosines at single nucleotide resolution. Their abundance within gene bodies and promoters of several growth-related genes supports their pertinent role in gene regulation. We propose that cytosine hydroxymethylation may contribute to the phenotypic plasticity of growth through epigenetic regulation of the somatotrophic axis.

This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

*Corresponding author. jorge.m.fernandes@nord.no (J.M.O. Fernandes).

Ethics statement

All animal handling procedures comply with the EU Directive 2010/ 63 and were approved by the Nord University ethics committee and the Norwegian Animal Research Authority (FOTS ID 1042).

Author's contributions

I.K. carried out the sampling, wet lab, bioinformatic analysis of the RRHP dataset, interpreted the results and wrote the article. D.A. contributed to the bioinformatic analysis of the RRHP and RNA-Seq data sets and revised the article. P.S. contributed to the establishment of the RRHP pipeline, the interpretation of the results and revised the article. A.V.N. contributed to the wet lab experiments, prepared the RNA-Seq libraries and revised the article. R.M. participated in the bioinformatic analysis of the RRHP and RNA-Seq datasets and revised the article. T.P. contributed to sampling, wet lab experiments and revised the article. F. P. revised the article. J.M.O.F. conceived the study, provided reagents and consumables, contributed to the interpretation of the results and revised the manuscript.

Declaration of Competing Interest

The authors declare no competing interests.

Keywords

DNA hydroxymethylation; Epigenetics; Somatotrophic axis; Growth; Teleosts

1 Introduction

Somatic growth requires a well-tuned transcriptional synchronization of three main tissue types that compose the somatotrophic axis: the pituitary gland, liver and skeletal muscle. From the production of growth hormone (GH) in the pituitary gland to the stimulation and regulation of hundreds of genes in the liver and muscle, the somatotrophic axis plays a critical role across several biological pathways involved in somatic growth, carbohydrate and lipid metabolism, energy equilibrium, normal development and reproduction [1–3]. The insulinlike growth factors (IGF-I and -II), their receptors (IGFRS) and binding proteins (IGFBPS) that are present in substantial amounts in liver and fast muscle, are directly associated with GH and collectively regulate somatic growth. However, the interaction of these molecules is greatly affected by both endogenous and exogenous factors, i.e. sex, developmental stage, nutrition, genetic background, temperature, photoperiod, salinity, dissolved oxygen, immunocompetence and stress [4–7]. For example, previous research has shown that seasonal differences [8], photoperiod, and sexual maturity [9] affect insulin-like growth factor 1 (*igf1*) plasma levels and consequently somatic growth. In addition, the liver is the main source of *igf1* production; however, in cases of malnutrition the liver is predominantly unresponsive to GH stimulation [10]. Such interactions have been widely demonstrated and both endocrine and paracrine *igf1* have been linked to bone, cartilage and muscle growth across vertebrates [11].

Because growth rate varies considerably with environmental factors [12], epigenetic mechanisms, which are known to integrate genomic and environmental information to bring about the phenotype [13], have the potential to be major regulators of growth both in early development (e.g., thermal imprinting) [14–16], and adulthood. Indeed, intrauterine growth restriction in rats induces DNA methylation changes around growth hormone response elements of the IGF-1 gene, resulting in developmental re-programming and persisting long-term effects [17]. Additionally, epigenetic mechanisms are likely responsible for splicing events leading to the translation of tissue-specific IGF-1 peptides through the usage of alternative promoters and exons [18,19]. In teleosts, recent studies have shown that epigenetic modifications regulate the expression of important growth genes during temperature fluctuations such as the DNA methylation of *myogenin* in Atlantic salmon (*Salmo salar*) [20] and Senegalese sole (*Solea senegalensis*) [14,21]. We have recently reported that growth-related methylation differences in *igf2bp2* are likely linked to phenotypic differences between small and large Nile tilapia full-sibs [22]. We identified sex-specific DNA methylation marks within the *map3k5*, *akt3*, *gadd45g* and *ppargc1a* genes. Their epigenetic regulation could potentially explain size-related differences among sexes because of their involvement in processes such as cell cycle, proliferation, cellular response to glucose and muscle tissue development [22]. The epigenetic function of non-coding RNAs, such as microRNAs (miRNAs), has also been shown to play critical roles in somatic growth [15]. For example, miR-206 targets the 3'UTR of IGF-1 and directly affects its

expression in Nile tilapia [23], while miR-1 and miR-133 regulate the sarcomeric actin organization in zebrafish [24]. Similar results have been shown in mammalian cells, such as hepatocellular carcinoma cells, where miR-190b directly regulates IGF-I expression [25]. DNA methylation plays also an important role within the GH/IGF-I axis. Recent studies have identified the P2 promoter of IGF-I to be epigenetically mediated by DNA methylation with significant consequences in IGF-I circulation and human growth [26,27]. Although DNA methylation is an epigenetic modification that has received much attention in response to several biological contexts, including the function and regulation of major components of the somatotrophic axis, little is known for its oxidized derivative. DNA hydroxymethylation is an epigenetic modification that occurs via ten-eleven-translocation enzymes, which catalyse the oxidation of 5-methyl (-CH₃) into 5-hydroxymethyl (-CH₂OH) groups in cytosines (5hmCs) [28]. Additionally, 5hmC has been identified as a stable DNA modification with a pertinent role in gene regulation and transcription [29,30], tissue-specific patterns and developmental stage-dependent profiles [31]. In human, 5hmCs are not exclusively correlated with DNA methylcytosine (5mC) content and when found in transcribed gene bodies they have been positively correlated with transcription [32]. In spite of 5hmC's significance in regulating gene expression in humans [31–33] and mice [34,35], little is known about its role on growth, especially taking into account that it has been associated with abnormal growth, cancers and oncogenic pathways such as the PI3K-Akt [36,37].

Growth hormone signaling and the overall function of the somatotrophic axis is largely conserved through evolution between fishes and higher vertebrates [38]. In contrast to mammals, most teleosts are poikilothermic and thus their growth rate is influenced to a greater extent by variations in environmental factors such as temperature. Furthermore, mammals have a determinate growth in comparison to many teleost species, which exhibit indeterminate growth. Together, the functional conservation of the somatotrophic axis and the peculiarities described above, make teleosts excellent models for the investigation of epigenetic mechanisms that modulate the somatotrophic axis. In this study, we compared genome-wide and tissue-specific 5hmC profiles at single nucleotide resolution between pituitary, liver and muscle using Nile tilapia (*Oreochromis niloticus*) as our model organism. Finally, we investigated the link between 5hmC and gene expression in skeletal muscle, since it constitutes the majority of the body mass and is directly related to growth.

2 Results

2.1 Tissue-specific 5hmC levels across the somatotrophic axis

Pituitary, liver and muscle samples were collected from five Nile tilapia male siblings that were reared for 8 months in common garden. Among the three main tissues that comprise the somatotrophic axis, 5hmC levels were highest in liver, followed by muscle and pituitary. The overall 5hmC levels of each tissue followed the sequencing depth of the corresponding reduced representation 5hmC profiling (RRHP) libraries (see Methods). Pituitary libraries had an average depth of 21.7 M raw reads compared to 28.1 M and 22.6 M raw reads in liver and muscle libraries, respectively. After quality control, adapter trimming and alignment of the sequenced reads to the Nile tilapia genome, an average of 10.6, 14.9, and 11.5 M reads

were obtained from the libraries prepared from pituitary, liver and muscle, respectively (Fig. S1).

The Nile tilapia genome contains 1,613,446 CCGG sites in both strands across its 22 linkage groups and the mitochondrial genome. RRHP captured 75% (1,211,876 CCGG) of all possible sites. The reduced coverage was mostly attributed to the size selection during library preparation. Compared to the total number of genomic CCGGs, 15.7% was significantly hydroxymethylated based on the median filtering of 5hmC counts (see Methods). The remaining 59.3% was dominated by 5hmC sites with very low counts across all tissues. A total of 16,812 5hmCs, which was approximately 1% of the total genomic CCGGs, were identified as differentially hydroxymethylated among the three tissues (Fig. 1A).

The highest difference was detected between the muscle and liver (MvL), with 15,757 cytosines being differentially hydroxymethylated, out of which 3119 were hyper-hydroxymethylated in muscle ($q < 0.01$). By comparing the hydroxymethylation profiles between muscle-pituitary (MvP) and liver-pituitary (LvP) we identified 1728 ($q < 0.05$) and 1467 ($q < 0.01$) additional DhmcS, respectively (Fig. 1B). An overlap of almost 2000 DhmcS was observed between pairs (MvL-LvP, MvL-MvP and LvP-MvP); however, we identified only four cytosines with significantly different hydroxymethylation levels among all pairwise comparisons (Table S1). Two of these four DhmcS were found within predicted long non-coding RNAs (*LOC102076435* and *LOC112847764*), one Dhmc located within the predicted gene gap junction delta-3 protein-like (*LOC100690349*) and one Dhmc within the par-3 family cell polarity regulator alpha, b (*pard3ab*).

2.2 Tissue-specific DNA hydroxymethylation abundance within genomic features

Although DNA hydroxymethylation levels were not significantly different among tissues due to high variation among individuals, we identified significant differences in Dhmc enrichment within the annotated features of the Nile tilapia genome. Compared to liver, the muscle was characterized by an overall hypo-hydroxymethylation across all the annotated features ($p < 0.001$), including promoter regions ($p < 0.05$; Fig. 2A). Striking differences were also observed between pituitary and muscle, with exons and intergenic regions being mostly hypo-hydroxymethylated in muscle ($p < 0.001$; Fig. 2B). By comparing liver to pituitary, we observed that the liver genome was hyper-hydroxymethylated across all the annotated features with significant enrichment within introns and exons ($p < 0.001$) and transcription termination sites ($p < 0.05$; Fig. 2C). The relative enrichment of DhmcS was calculated based on the number of 5hmCs with the highest 5hmC levels in each annotated feature per tissue. Compared to liver and pituitary, DhmcS with the highest levels of hydroxymethylation in muscle were relatively enriched in introns (12%) and promoters (104%), whereas exons were depleted by 58%.

In contrast, DhmcS with highest levels of hydroxymethylation in liver were relatively depleted within promoters and enriched within exons, compared to muscle and pituitary (Fig. S2). The highest percentage of DhmcS among all tissues was detected within introns (40% on average), while DhmcS in intragenic regions including promoters represented 70%, 74.6% and 71.8% of the entire dataset in muscle, liver, and pituitary, respectively.

2.3 Differentially hydroxymethylated cytosines are found in genes with distinct molecular functions in each tissue

To better understand whether DhmcCs were linked to tissue-specific genes and potentially their functions, we performed a gene ontology enrichment analysis based on the genes containing or being closely associated to DhmcCs. In total, we identified 2102 hyper-hydroxymethylated genes in muscle compared to liver and pituitary, 4993 genes in liver compared to muscle and pituitary and 1006 genes in the pituitary compared to muscle and liver. Interestingly, we identified 773 genes containing hyper-hydroxymethylcytosines in both the muscle and liver, 40 common genes between muscle and pituitary and 23 genes shared between liver and pituitary (Fig. 1C).

In muscle compared to liver, genes containing hypo-5hmCs (lower 5hmC levels in muscle) were associated with molecular functions related to transcription regulator and GTPase activator activity, GTPase binding and Ras GTPase binding. Gene ontologies related to development as well as anatomical structure and system development were among the most enriched biological processes. Furthermore, we identified 12 biological processes linked to metabolism, including regulation of primary, cellular, macromolecule, RNA and nitrogen compound metabolic process. Interestingly, hypo-5hmCs were assigned to gene ontologies such as skeletal system, muscle structure and cartilage development (Table S2). On the other hand, hyper-5hmCs (higher 5hmC levels in the muscle) were primarily associated to ionotropic glutamate receptor activity, signaling receptor activity and cell surface receptor signaling pathway, as well as system development and developmental process (Table S3). Notably, 773 genes were identified containing both hyper- and hypo-5hmCs that occurred in different intragenic loci. These genes were primarily linked to the regulation of biological and cellular processes, as well as the signal transduction and response to stimulus (Table S4).

In muscle compared to pituitary, hyper-hydroxymethylated genes (higher 5hmC levels in the muscle) were associated with tissue morphogenesis and the morphogenesis of an epithelium, as well as the developmental process and signaling among other gene ontologies (Table S5). Hypo-hydroxymethylated genes (lower 5hmC levels in muscle) were enriched for functions related to transcription regulator activity and the regulation of metabolic and biosynthetic process of macromolecules and RNA (Table S6). We identified 40 shared genes between muscle and pituitary containing hyper- and hypo-5hmCs in different loci. However, due to the relatively low number of genes we did not identify any significant enrichment for molecular functions or biological processes.

Compared to pituitary, the liver contained hypo-hydroxymethylated genes (lower 5hmC levels in liver) that were enriched for a cell surface receptor signaling pathway and axon target recognition (Table S7). In contrast, hyper-hydroxymethylated genes (higher 5hmC levels in the liver) were enriched for molecular functions related to protein and DNA binding. The most enriched biological processes were the regulation of cellular and biological processes, signaling and signal transduction (Table S8). In total, 23 genes containing hyper- and hypo-5hmCs were found within the LvP pairwise comparison. Among them, we identified the gene coding for arginine-glutamic acid dipeptide repeats protein (rere), which is involved in cell survival and apoptosis [39], and the protein phosphatase 1

regulatory subunit 3C (*ppp1r3c*) involved in glycogen metabolism as well as the metabolic process of macromolecules [40].

2.4 Differentially hydroxymethylated genes involved in somatic growth

Genes that play a critical role in development and somatic growth were identified across all pairwise comparisons. Most genes involved in somatic growth were hypo-hydroxymethylated in muscle compared to liver. These included several fibroblast growth factors (*fgf5*, *fgf6*, *fgf8*, *fgf12*, *fgf14* and *fgf22*), the receptor *fgfr3* and the fibroblast growth factor receptor substrate 2 (*frs2*). Only *fgf10* and *fgfr4* were hyper-hydroxymethylated in muscle. Additionally, growth arrest-specific proteins (*gas2l3* and *gas7*) and growth differentiation factors (*gdf3*, *gdf5*, *gdf6*, *gdf6a* and *gdf10*) were also hypo-hydroxymethylated in muscle. Hypo-5hmCs were detected within the last exon of growth hormone secretagogue receptor type 1 (*ghsr*), the second intron of insulin-like growth factor 2 mRNA binding protein 2 (*igf2bp2*), the first and third introns of insulin-like growth factor binding protein 5 (*igfbp5*) and the 32nd intron of IGF like family receptor 1 (*igflr1*). On the other hand, the promoter of the growth hormone releasing hormone (*ghrh*), the first intron of the insulin-like growth factor binding protein 2-B and the 29th exon of *igflr1* contained a single hyper-5hmC. Several genes involved in somatic growth contained both hyper- and hypo-5hmCs in different intragenic loci. These included the myoD family inhibitor domain-containing protein (*mdfc*), the multiple epidermal growth factor-like domains protein 10 (*mefl10*) and the myosin heavy chain fast skeletal muscle (*myh*) (Fig. 3A; Table S9A). In muscle compared to pituitary, we identified two hyper-5hmCs located 7549 and 7367 bp, from the transcription start site (TSS) of the myostatin (*mstn*) gene, and a single hyper-5hmC at 6945 bp from the TSS of the growth arrest and DNA damage inducible gamma (*gadd45g*) gene. Similar to the comparison between muscle and liver, fibroblast growth factor genes *fgf5*, *fgf12* and *frs2* were hypo-hydroxymethylated in muscle compared to pituitary; however, hyper-5hmCs were detected within the *fgfr2* and *igfbp3* genes. Additionally, two hyper-5hmCs were identified within the 5th exon and 7th intron of the gene coding for myoD family inhibitor domain-containing protein 2 (*mdfic2*) (Fig. 3B; Table S9B). Several of the above-mentioned genes were also differentially hydroxymethylated in liver compared to pituitary (Fig. 3C; Table S9C).

2.5 Gene expression profiling in Nile tilapia fast muscle

We investigated the correlation between 5hmC levels and gene expression in muscle, since it constitutes the majority of the total body mass in Nile tilapia and it closely reflects growth rate. The average sequencing depth of the RNA-Seq libraries was 26 M reads (Table S10). Approximately 1 M reads were excluded from the analysis after quality control ($Q > 20$) and trimming. On average, 2.8 M reads were mapped multiple times and 19.2 M reads were mapped uniquely to the reference Nile tilapia genome. In total, we identified 80,417 transcripts that were ranked based on their normalized counts. The gene with the highest expression in Nile tilapia fast muscle was the actin alpha skeletal muscle A (*LOC100534413*), which is involved in several biological processes including muscle contraction, myosin binding, skeletal muscle fibre development and thin filament assembly. Among the genes with the highest expression, we identified several fast skeletal muscle myosin light and heavy chains (*mylpf*, *LOC100712344*, *LOC100707599*, *LOC100698429*),

creatine kinases muscle a and b (*ckma* and *ckmb*), ATPase sarcoplasmic/endoplasmic reticulum Ca²⁺ transporting 1 like (*atp2a1l*) and several isoforms of troponin T (*LOC100707421*) and tropomyosin 1 (*tpm1*) (Table S11).

2.6 Association of muscle-specific gene expression and DNA hydroxymethylation

To investigate the links between DNA hydroxymethylation and gene expression levels in muscle, we performed a principal component analysis using three variables: the 5hmC levels, the distance to the transcription start site (TSS) and gene expression. The first principal component (PC1) was strongly related to variation in both the distance to TSS ($\rho = 0.63$) and gene expression (GE; $\rho = 0.70$), indicating that these variables were strongly positively correlated in our data. On the other hand, 5hmC levels were strongly represented in the second principal component (PC2; $\rho = 0.84$) revealing no correlation with GE ($\rho = 0.10$; Fig. 4). Since the first kilobases up- and downstream of promoter regions have been previously correlated with gene transcription and alternative splicing [41–43], we also investigated the correlation between gene expression and 5hmC levels in promoters, first introns and first exons.

Consistent with our findings above, we identified a negative correlation between 5hmC levels and gene expression in all three comparisons (Fig. 5). Among the most expressed genes in fast muscle, we identified 4 differentially hydroxymethylated genes between muscle and liver. Two DhmcS were located 19,650 and 19,551 bp upstream of the myosin heavy chain fast skeletal muscle (*LOC100706261*) promoter, two DhmcS were located 10,176 and 10,040 bp upstream of the nucleoside diphosphate kinase B (*LOC100697135*) promoter, two DhmcS were found within the 16th exon of *atp2a1l*, one Dhmc within the 5th exon of myosin light chain 3 skeletal muscle isoform (*LOC100698429*) and one Dhmc within the 1st intron of troponin T fast skeletal muscle isoforms (*LOC100707421*). Surprisingly, all 5hmCs were hypo-hydroxymethylated in muscle compared to liver, except the one that was located closest to the TSS (1318 bp) and within the 1st intron of troponin T fast skeletal muscle isoform (Table 1).

3 Discussion

This study presents the first DNA hydroxymethylation profiling of the somatotrophic axis at single nucleotide resolution in a teleost species. We compared the 5hmC levels among tissues and identified the liver as the tissue with the highest, the muscle with intermediate and the pituitary with the lowest 5hmC levels (Fig. 6). In contrast to these results, global 5hmC levels of the muscle and liver measured by liquid chromatography mass spectrometry (LC/MS) in male zebrafish (*Danio rerio*) revealed that muscle had higher 5hmC levels than liver [44]. Similarly, 5hmC quantification by LC/MS in muscle and liver of mice, showed higher 5hmC levels in muscle than liver [45]. These results can potentially be explained either by species-specific 5hmC patterns or by the methods used, i.e. reduced representation of the genome using RRHP, compared to global 5hmC quantification using LC/MS. The low 5hmC content in pituitary gland in Nile tilapia is in accordance with previous findings in mice [45]. Indeed, reports have demonstrated that high 5hmC levels in brain are associated

with tissue-specific neuronal functions [46,47], compared to pituitary, which is responsible for hormone secretion.

Based on the functional enrichment analysis of hyper-hydroxymethylated genes in muscle compared to liver, we identified several genes involved in ionotropic and metabotropic glutamate receptor activity. These genes were associated with all three types of ionotropic glutamate receptors, namely, the *N*-methyl-D-aspartate (NMDA), the kainate and the α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors while their functions were mostly associated with fast synaptic transmission and the modulation of synaptic responses [48]. A previous study using C2C12 mouse skeletal myoblasts has shown that NMDA receptors play a major role in calcium influx and myoblast fusion, which are fundamental processes for muscle growth [49]. Furthermore, we also identified several differentially hydroxymethylated metabotropic glutamate receptors (*grm1*, *grm2*, *grm4*, *grm7* and *grm8*) between muscle and liver, which play critical roles in cell survival, metabolism and proliferation through their association with the PI3 kinase and the initiation of the PI3K-Akt-mTOR signaling pathway [50]. Glutamate receptors were found to be under positive selection across several domesticates [51], including fish [52]. Particularly in sea bass (*Dicentrarchus labrax*), the identification of genetic mutations after many generations in captivity was found to be overlapping with early changes in DNA methylation profiles during domestication, suggesting that epimutations potentially drive genome evolution [53]. Here, we identified significantly hydroxymethylated ionotropic and metabotropic glutamate receptors within the somatotrophic axis, which supports the functional importance of these genes in somatic growth.

Finally, we identified several Dhmc within the receptor-type tyrosine-protein phosphatase N2 (*ptprm2*) gene, which is expressed in nervous and endocrine cells and it is associated with lipid metabolism and insulin secretion [54,55]. Previous studies have shown that *ptprm2* is largely associated with somatic growth, and regulated by epigenetic modifications in various biological processes including hepatocellular carcinomas [56], breast cancer [57], as well as prenatal growth patterns and birthweight in humans [58]. Therefore, it is likely that DNA hydroxymethylation may modulate somatic growth indirectly, through the regulation of genes related to neuromuscular and neurohepatic signals.

Growth hormone release depends on three hypothalamic factors, namely, growth hormone releasing hormone (GHRH), somatostatin (SST) and ghrelin multifaceted hormone (GHRL) [59,60]. The latter is stimulated by receptors, such as the growth hormone secretagogue receptor (*ghsr*), which was found to be hyper-hydroxymethylated in the liver, compared to the muscle and pituitary. *Ghsr* encodes for a tissue-specific hormone receptor that is commonly found across the somatotrophic axis and regulates energy homeostasis, fat distribution and the expression of genes involved in lipid metabolism and total body weight [61,62]. Notably, among the differentially hydroxymethylated genes, we also identified a hyper-5hmC (higher 5hmC level in muscle than liver) within the promoter of the growth hormone *ghrh* as well as several hyper-5hmCs (higher 5hmC levels in muscle and pituitary than liver) within predicted genes coding for somatostatin in Nile tilapia (*loc102077970* – somatostatin-1; *loc100698045* – somatostatin-1B; *loc100694069* – somatostatin-2), known also as growth hormone-inhibiting hormone. Additionally, we identified differentially

hydroxymethylated genes that play critical roles in growth hormone signaling and are downstream regulators of somatic growth, such as the insulin-like growth factor binding proteins *igfbp2*, *igfbp2* and *igfbp5* and the receptor *igflr1*. In particular, hyper-5hmCs (higher 5hmC levels in muscle than liver) were located within *igfbp2* and the 29th exon of *igflr1* while hypo-5hmCs (lower 5hmC levels in muscle than liver) were located within *igfbp5*, *igfbp2* and the 32nd intron *igflr1* (Table S9A). The same genes were also differentially hydroxymethylated between liver and pituitary. A hyper-5hmC (higher 5hmC levels in liver than pituitary) was found within *igfbp2*, and a hypo-5hmC (lower 5hmC levels in liver than pituitary) was located within *igflr1*. Considering that epigenetic modifications drive tissue-specific gene expression, it is likely that differentially hydroxymethylated cytosines can contribute towards the expression of insulin-like growth factor receptor and binding protein genes. These results suggest a link between DNA hydroxymethylation and the function of the somatotrophic axis (Fig. 6).

Interestingly, several genes were both hyper- and hypo-hydroxymethylated between tissues due to the localization of unique DhmcCs across their gene body or their proximal regions. For example, the above-mentioned *ptprn2* gene was highly hydroxymethylated in liver (compared to muscle) across the 1st, 3rd and 8th introns as well as the 5th exon, but a hypo-hydroxymethylated cytosine (low 5hmC levels in liver compared to muscle) was also present in the 10th intron. Similar patterns were observed in other genes, including the multiple epidermal growth factor-like domains protein 10 (*mefg10*) and insulin receptor (*insr*), methyltransferases such as *mettl21c* and *mettl4*, adhesion G protein-coupled receptors (*adgrf4*, *adgrl2*, *adgrl3*, *adgrb1*) and several transcription factors (*atf2*, *atf3*, *ebf3*, *grhl1*, *gtf3c1*, *gtf3c3*, *hsf2*, *sox4* and *runx2*). Taking into consideration the enrichment of 5hmCs within introns and their positional dynamics between tissues, these findings could potentially explain tissue-specific gene expression through the regulation of alternative promoters or splicing junctions. Indeed, recent studies have shown that DNA hydroxymethylation is involved in alternative splicing [63,64] and that positional dynamics of 5hmCs within or in proximal distance of genes result in expression changes [35].

Additionally, studies in mammals and mammalian cells have shown that DNA hydroxymethylation is positively correlated to gene expression and is present in gene bodies of actively transcribed genes [65–67]. Here, even though the correlations of gene expression in the muscle and 5hmC levels were generally moderate ($\rho = 0.47$ – 0.63), we found that 5hmCs display a positive correlation with distance from TSSs; the farther the 5hmCs from a TSS, the higher the expression of genes. Although, this correlation included both upstream and downstream 5hmCs from the TSS, 80% of those 5hmCs were located within gene bodies. Therefore, this correlation was primarily affected by 5hmCs within the downstream region of TSSs. Frequently, first introns and exons are positively associated with gene regulation [41–43]. By comparing muscle gene expression and 5hmC levels of promoters, first introns and first exons, we determined that 5hmCs in close proximity to TSSs had an overall negative correlation with gene transcription. However, a close examination of the top 30 most expressed genes in muscle and their hydroxymethylation status revealed that the only hyper-5hmC in muscle compared to liver was also the closest one to the TSS of *tropoin T fast skeletal muscle isoform*. Several hypo-5hmCs (lower 5hmC levels in muscle than liver) were found at a distance of 10–19 kbp upstream from the promoter as well as

within the open reading frames of highly expressed genes in the muscle but none of them were located within the first intron or exon. These findings suggest that the position of 5hmCs within gene features plays indeed a crucial role in gene repression or activation. The overall negative correlation between 5hmCs and gene expression in our study is likely attributed to the presence of other epigenetic mechanisms that were not profiled here, such as cytosine methylation, histone modifications and non-coding RNAs. Although, high 5hmC levels are usually correlated with gene activation, there is also evidence of 5hmCs being involved in the repression of genes such as the insulin-degrading enzyme (IDE) in depressed individuals [68] and polycomb groups of developmental regulators [66]. In human, there is a high overlap (85.6%) between 5mCs and 5hmCs [69]. Previously, high 5mC levels within the first intron of genes have been strongly associated to low levels of expression across multiple species and tissues [43]. Considering the potential overlap between these two cytosine modifications, their overall contrasting roles in gene regulation, their relative abundance (5mCs are estimated to be 14-fold more abundant than 5hmCs in mammalian embryonic stem cells [28]), and the moderate negative correlation between substantially hydroxymethylated genes and their expression, we propose that the presence of 5hmCs throughout the genome plays a balancing role in gene regulation.

4 Conclusions

Somatic growth and normal development are regulated by a complex molecular cascade that is affected by several mechanisms, including epigenetic modifications. Here, we show that DNA hydroxymethylation is present throughout the genome, is characterized by tissue-specific profiles and is highly enriched within gene bodies (Fig. 6).

Several genes that are both directly or indirectly associated with functions such as growth hormone release, insulin-like growth factor binding, liver metabolism and somatic growth were found to be differentially hydroxymethylated between pituitary, liver and muscle. We identified a moderate but significant negative correlation between gene expression and 5hmC levels in muscle. However, one of the most expressed genes in the muscle (*troponin T fast skeletal muscle*) contained a single hyper-5hmC (higher 5hmC levels in the muscle than liver) within the first intron, revealing a positive correlation between DNA hydroxymethylation and gene expression. The high overlap between 5hmCs and 5mCs and their overall contrasting roles in gene transcription pinpoints towards a dynamic interrelationship with DNA hydroxymethylation acting as a balancing factor in gene regulation within tissues. The identification of tissue-specific epigenetic modifications within the somatotrophic axis adds an additional layer of complexity to our current knowledge about the regulation of growth.

5 Materials and methods

5.1 Experimental design and sampling

Female wild Nile tilapia mouthbrooding fertilized eggs were captured using traditional fishing nets and traps at the river Nile in Luxor, Egypt. At 9 days post-hatching, the larvae were transported to our research facilities in Bodø, Norway. A total of 14 batches of juveniles originating from 14 different wild females were tagged and used as the base

population (F0) for our Nile tilapia breeding program. These Nile tilapia were reared in a recirculating aquaculture system (pH = 7.6, oxygen saturation = 100%, temperature = 28 °C and photoperiod adjusted at 11:13 dark:light) and fed ad libitum with 0.15–0.8 mm Amber Neptun pellets (Skretting, Norway) as previously described [70]. After a successful reproduction cycle in captivity, we obtained the F1 generation. These fish were PIT-tagged and reared for 8 months in common garden to minimize the effect of environmental factors. For the samples used in this study, 5 full-sib F1 males were randomly selected from the tank. We focused on Nile tilapia males because they are preferred for fish farming, since they grow faster and larger than females. The fish were euthanized by immersion for 3 min in 10 L of water containing 15 mL clove oil pre-mix, consisting of pure clove oil (Sigma Aldrich, USA) and 96% ethanol at a 1:9 ratio. Fast (white) muscle, liver and pituitary gland were carefully dissected, snap-frozen in liquid nitrogen and stored at – 80 °C. In particular, fast muscle was extracted from a 4 mm cross-section at 0.7 standard length. The samples were taken above the lateral line from the left upper white muscle mass. Liver samples were collected from the left lobe and around the entry point of the portal vein. The pituitary gland was extracted by cutting open the skull, removing the brain from the dorsal side and picking the pituitary out with forceps. Information regarding the measurements (weight, total and standard length) of the fish can be found in Table 2.

5.2 DNA extraction and RRHP library preparation

DNA extraction was carried out using the Quick-DNA miniprep plus kit (Zymo Research, USA) according to the manufacturer's instructions. Quantification and integrity control of the extracted DNA was performed using Qubit 3.0 fluorometer and the double-stranded DNA high sensitivity assay kit (ThermoFisher Scientific, USA) and Tapestation 2200 (Agilent Technologies, USA), respectively.

Reduced representation of 5-hydroxymethylcytosine profiling (RRHP) was performed according to manufacturer's protocol (Zymo Research). The preparation of RRHP libraries was performed synchronously for all tissues and replicates. The starting DNA concentration was 150 ng diluted in 10 µL DNase and RNase free water. Sequencing was performed using Nord University's inhouse NextSeq500 (Illumina, USA) and all libraries were pooled with equal volumes. Finally, they were distributed among three High-Output kit v2.5 150 cycle flow cells, yielding approximately 360 million single-end reads.

5.3 RRHP bioinformatic pipeline

Raw reads were trimmed for adapters using trim_galore v0.4.4 and quality check was performed using MultiQC. The software Bowtie v0.12.8 was used for the alignment of high quality and trimmed reads to the latest Nile tilapia reference genome (NCBI assembly GCA_001858045.3) [71] with the following parameters: -S -v 1 -n 1 -m 3 -strata -best. The chromosome, position and strand information from reads that start with CCGG were extracted in text files for each sample. These files were used as input into R, where a count matrix was created by summing up overlapping chromosomal positions. To minimize the false discovery of 5hmCs the data set was filtered twice, as reported [70]. Briefly, the first filter removed reads for which more than two groups of samples (2 out of 3 tissues) had less than 1 count. For the second filter, the median of the entire dataset was calculated and 5hmC

sites were removed when at least one tissue among the three (5 or more samples) had counts equal or less than the median (21 counts). For the comparison of 5hmC levels at single nucleotide resolution between the three tissues, we used the R package *limma* [72] and the adjusted p -values were calculated using the Benjamini-Hochberg correction. Three contrasts were performed comparing the 5hmC levels of i) the muscle against the liver (MvL), ii) the muscle against the pituitary gland (MvP) and iii) the liver against the pituitary gland (LvP).

5.4 Functional enrichment analysis

Differentially hydroxymethylated cytosines (DhmCs) were assigned to genes based on their proximal distance to transcription start sites, using the all-in-one Perl script *annotatepeaks.pl* from HOMER (<http://homer.ucsd.edu>). Genes associated with DhmCs were considered to be differentially hydroxymethylated. DhmCs located within uncharacterized genes were excluded from further analysis, since they lack information regarding their function. In total, 4590 DhmCs in MvL, 463 DhmCs in MvP and 367 DhmCs in LvP comparisons were annotated within uncharacterized genes. Functional enrichment analysis was performed using *gProfiler* (version: e99_eg46_p14_f929183) with default parameters for *Oreochromis niloticus*. Multiple testing correction was performed using the suggested tailor-made algorithm *g:SCS* and the significance threshold was set at 0.05 [73]. The analysis was separated in 6 stages using hyper- and hypo-hydroxymethylated genes per pairwise tissue comparison as well as 3 additional stages for genes that were both hypo- and hyper-hydroxymethylated in a single tissue per pairwise tissue comparison.

5.5 RNA extraction, library preparation and sequencing

The same fast muscle samples that were used for the RRHP library preparation were homogenised in DNA/RNA Shield (Zymo Research) using ZR bashing beads lysis tubes (Zymo Research). Homogenization was carried out using a Precellys 24 homogenizer (Bertin Instruments, France) with 2 cycles at 5000 rpm and 20 s duration. RNA extraction was performed using the Quick-RNA miniprep kit (Zymo Research) following the manufacturer's protocol for samples stored in DNA/RNA Shield. The Ribo-Zero Gold rRNA removal kit (Illumina) was used to ensure the removal of ribosomal RNA. RNA quality was determined using the spectrophotometer NanoDrop ND 1000 (ThermoFisher Scientific), while its concentration and integrity was measured with an RNA High-sensitivity screentape on a TapeStation 2200 (Agilent Technologies). For the library preparation, the NEBNext Ultra II Directional RNA library prep kit for Illumina (NEB, USA) was used following the manufacturer's instructions and recommendations. Each sample was tagged with a unique barcode and the libraries were pooled at equimolar concentrations. The pool was sequenced at the Norwegian Sequencing Centre on a HiSeq 4000 lane (Illumina) in paired-end 150 bp mode.

5.6 Bioinformatic analysis of RNA-seq data

The sequencing data obtained above (section 5.5) were trimmed using *BBDuk* of the *BBTools Suite* (v. 38.22-0, Joint Genome Institute), which decontaminates using *Kmers*. Adapters were trimmed as per default ($ktrim = r$, $k = 23$, $mink = 11$, $hdist = 1$, tpe , tbo) and quality trimmed for both sides of the read with the *Phred* threshold set to 20, while reads less than 23 bp after trimming to Q20 were discarded. An index was built with the *HISAT2* (v.

2.1.0) program [74] using the Nile tilapia reference genome above and trimmed reads were aligned by the core function of HISAT2 with RNA strandness set to FR and reported alignments tailored for downstream transcript assemblers. Alignments were sorted by samtools (v. 1.9) and a summarized report for all samples was generated using the MultiQC software [75]. Calculation of count numbers for each gene was performed by featureCounts (v 1.6.2) [76] of the Subread package [77] using the gff3 genome annotation (NCBI, ref_O_niloticus_UMD_NMBU_top_level.gff3). Chimeric fragments were excluded, duplicates were ignored, only primary alignments were counted and fragments were allowed to match more than one metafeature. Further analysis were performed in R (v. 3.6.1) [78] and Rstudio (1.2.1335) [79]. The DESeq2 package [80] was used for applying a regularized log transformation to the data in order to minimize differences between small counts and normalize according to the library size in an unbiased manner by prior information.

For each cytosine with 5hmC information, the mean of 5hmC counts across all five biological replicates was used together with the distance of each C from the TSS. Each C corresponded to a single expressed gene, the value of which was estimated based on the mean of gene expression counts across all five biological replicates. In other words, a value of gene expression could be attributed to more than one C. Principal component analysis and visualizations were performed with the Facto-MineR (v. 1.42) and factoextra (v. 1.0.5) packages.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are thankful to Hilde Ribe (Nord University) for her assistance in establishing our Nile tilapia breeding program. We are also thankful to Øivind Torslett, Steinar Johnsen and Kaspar Klaudiussen (Nord University) for their contributions in rearing, breeding and welfare of our fish.

Funding

This work was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreements nos. 683210 and 812986) and by the Research Council of Norway under the Toppforsk programme (grant agreement no 250548/F20).

Data availability

The DNA hydroxymethylation dataset of this study is available in the SRA (NCBI) repository, under the accession number PRJNA665628. The RNA sequencing dataset of this study is available in the GEO (NCBI) repository, under the accession number GSE158910.

References

- [1]. Whittington CM, Wilson AB. The role of prolactin in fish reproduction. *Gen Comp Endocrinol.* 2013; 191:123–136. [PubMed: 23791758]
- [2]. Company R, Astola A, Pendon C, Valdivia MM, Perez-Sanchez J. Somatotropic regulation of fish growth and adiposity: growth hormone (GH) and somatolactin (SL) relationship. *Comp Biochem Physiol C Toxicol Pharmacol.* 2001; 130:435–445. [PubMed: 11738631]

- [3]. Renaville R, Hammadi M, Portetelle D. Role of the somatotrophic axis in the mammalian metabolism. *Domest Anim Endocrinol.* 2002; 23:351–360. [PubMed: 12142251]
- [4]. Beckman BR. Perspectives on concordant and discordant relations between insulin-like growth factor 1 (IGF1) and growth in fishes. *Gen Comp Endocrinol.* 2011; 170:233–252. [PubMed: 20800595]
- [5]. Katan T, Nash GW, Rise ML, Hall JA, Fernandes JMO, Boyce D, Johnsen CA, Gamperl AK. A little goes long way: improved growth in Atlantic cod (*Gadus morhua*) fed small amounts of wild zooplankton. *Aquaculture.* 2016; 451:271–282.
- [6]. Campos C, Valente LM, Conceicao LE, Engrola S, Sousa V, Rocha E, Fernandes JM. Incubation temperature induces changes in muscle cellularity and gene expression in Senegalese sole (*Solea senegalensis*). *Gene.* 2013; 516:209–217. [PubMed: 23291415]
- [7]. Campos C, Valente LM, Borges P, Bizuayehu T, Fernandes JM. Dietary lipid levels have a remarkable impact on the expression of growth-related genes in Senegalese sole (*Solea senegalensis* Kaup). *J Exp Biol.* 2010; 213:200–209. [PubMed: 20038653]
- [8]. Metcalfe KM, Bull CD, Mangel M. Seasonal variation in catch-up growth reveals state-dependent somatic allocations in salmon. *Evol Ecol Res.* 2002; 4:871–881.
- [9]. Taylor JF, Porter MJ, Bromage NR, Migaud H. Relationships between environmental changes, maturity, growth rate and plasma insulin-like growth factor-I (IGF-I) in female rainbow trout. *Gen Comp Endocrinol.* 2008; 155:257–270. [PubMed: 17592733]
- [10]. Perez-Sanchez J, Marti-Palanca H, Kaushik SJ. Ration size and protein intake affect circulating growth hormone concentration, hepatic growth hormone binding and plasma insulin-like growth factor-I immunoreactivity in a marine teleost, the gilthead sea bream (*Sparus aurata*). *J Nutr.* 1995; 125:546–552. [PubMed: 7876930]
- [11]. Valente LMP, Moutou KA, Conceicao L, Engrola S, Fernandes JMO, Johnston IA. What determines growth potential and juvenile quality of farmed fish species. *Rev Aquac.* 2013; 5:168–193.
- [12]. Perez-Sanchez J, Simo-Mirabet P, Naya-Catala F, Martos-Sitcha JA, Perera E, Bermejo-Nogales A, Benedito-Palos L, Caldach-Giner JA. Somatotrophic axis regulation unravels the differential effects of nutritional and environmental factors in growth performance of marine farmed fishes. *Front Endocrinol (Lausanne).* 2018; 9:687. [PubMed: 30538673]
- [13]. Turner BM. Epigenetic responses to environmental change and their evolutionary implications. *Philos Trans R Soc Lond Ser B Biol Sci.* 2009; 364:3403–3418. [PubMed: 19833651]
- [14]. Campos C, Valente LM, Conceicao LE, Engrola S, Fernandes JM. Temperature affects methylation of the myogenin putative promoter, its expression and muscle cellularity in Senegalese sole larvae. *Epigenetics.* 2013; 8:389–397. [PubMed: 23538611]
- [15]. Campos C, Sundaram AY, Valente LM, Conceicao LE, Engrola S, Fernandes JM. Thermal plasticity of the miRNA transcriptome during Senegalese sole development. *BMC Genomics.* 2014; 15:525. [PubMed: 24966054]
- [16]. Campos C, Valente LMP, Conceicao LEC, Engrola S, Fernandes JMO. Molecular regulation of muscle development and growth in Senegalese sole larvae exposed to temperature fluctuations. *Aquaculture.* 2014; 432:418–425.
- [17]. Fu Q, McKnight RA, Callaway CW, Yu X, Lane RH, Majnik AV. Intrauterine growth restriction disrupts developmental epigenetics around distal growth hormone response elements on the rat hepatic IGF-1 gene. *FASEB J.* 2015; 29:1176–1184. [PubMed: 25466885]
- [18]. Schober ME, Ke X, Xing B, Block BP, Requena DF, McKnight R, Lane RH. Traumatic brain injury increased IGF-1B mRNA and altered IGF-1 exon 5 and promoter region epigenetic characteristics in the rat pup hippocampus. *J Neurotrauma.* 2012; 29:2075–2085. [PubMed: 22413999]
- [19]. Beresewicz M, Majewska M, Makarewicz D, Vayro S, Zablocka B, Gorecki DC. Changes in the expression of insulin-like growth factor 1 variants in the postnatal brain development and in neonatal hypoxia-ischaemia. *Int J Dev Neurosci.* 2010; 28:91–97. [PubMed: 19766709]
- [20]. Burgerhout E, Mommens M, Johnsen H, Aunsmo A, Santi N, Andersen O. Genetic background and embryonic temperature affect DNA methylation and expression of myogenin and muscle

- development in Atlantic salmon (*Salmo salar*). *PLoS One*. 2017; 12:e0179918 [PubMed: 28662198]
- [21]. Campos C, Valente LM, Conceicao LE, Engrola S, Fernandes JM. Epigenetic regulation of muscle development and growth in senegalese sole larvae. *Commun Agric Appl Biol Sci*. 2013; 78:60–61. [PubMed: 25141624]
- [22]. Podgorniak T, Brockmann S, Konstantinidis I, Fernandes JMO. Differences in the fast muscle methylome provide insight into sex-specific epigenetic regulation of growth in Nile tilapia during early stages of domestication. *Epigenetics*. 2019; 14:818–836. [PubMed: 31131688]
- [23]. Yan B, Zhu CD, Guo JT, Zhao LH, Zhao JL. miR-206 regulates the growth of the teleost tilapia (*Oreochromis mloticus*) through the modulation of IGF-1 gene expression. *J Exp Biol*. 2013; 216:1265–1269. [PubMed: 23197102]
- [24]. Mishima Y, Abreu-Goodger C, Staton AA, Stahlhut C, Shou C, Cheng C, Gerstein M, Enright AJ, Giraldez AJ. Zebrafish miR-1 and miR-133 shape muscle gene expression and regulate sarcomeric actin organization. *Genes Dev*. 2009; 23:619–632. [PubMed: 19240126]
- [25]. Hung TM, Ho CM, Liu YC, Lee JL, Liao YR, Wu YM, Ho MC, Chen CH, Lai HS, Lee PH. Up-regulation of microRNA-190b plays a role for decreased IGF-1 that induces insulin resistance in human hepatocellular carcinoma. *PLoS One*. 2014; 9:e89446 [PubMed: 24586785]
- [26]. Ouni M, Gunes Y, Belot MP, Castell AL, Fradin D, Bougneres P. The IGF1 P2 promoter is an epigenetic QTL for circulating IGF1 and human growth. *Clin Epigenetics*. 2015; 7:22. [PubMed: 25789079]
- [27]. Ouni M, Belot MP, Castell AL, Fradin D, Bougneres P. The P2 promoter of the IGF1 gene is a major epigenetic locus for GH responsiveness. *Pharm J*. 2016; 16:102–106.
- [28]. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009; 324:930–935. [PubMed: 19372391]
- [29]. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet*. 2011; 13:7–13. [PubMed: 22083101]
- [30]. Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem*. 2014; 6:1049–1055. [PubMed: 25411882]
- [31]. Ponnaluri VK, Ehrlich KC, Zhang G, Lacey M, Johnston D, Pradhan S, Ehrlich M. Association of 5-hydroxymethylation and 5-methylation of DNA cytosine with tissue-specific gene expression. *Epigenetics*. 2017; 12:123–138. [PubMed: 27911668]
- [32]. Nestor CE, Ottaviano R, Reddington J, Sproul D, Reinhardt D, Dunican D, Katz E, Dixon JM, Harrison DJ, Meehan RR. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res*. 2012; 22:467–477. [PubMed: 22106369]
- [33]. Rohde K, Keller M, Stumvoll M, Dietrich A, Bluher M, Bottcher Y. DNA 5-hydroxymethylation in human adipose tissue differs between subcutaneous and visceral adipose tissue depots. *Epigenomics*. 2015; 7:911–920. [PubMed: 26418625]
- [34]. Ma Q, Xu Z, Lu H, Xu Z, Zhou Y, Yuan B, Ci W. Distal regulatory elements identified by methylation and hydroxymethylation haplotype blocks from mouse brain. *Epigenetics Chromatin*. 2018; 11:75. [PubMed: 30594220]
- [35]. Greco CM, Kunderfranco P, Rubino M, Larcher V, Carullo P, Anselmo A, Kurz K, Carell T, Angius A, Latronico MV, Papait R, et al. DNA hydroxymethylation controls cardiomyocyte gene expression in development and hypertrophy. *Nat Commun*. 2016; 7:12418 [PubMed: 27489048]
- [36]. Bhattacharyya S, Pradhan K, Campbell N, Mazdo J, Vasantkumar A, Maqbool S, Bhagat TD, Gupta S, Suzuki M, Yu Y, Grealley JM, et al. Altered hydroxymethylation is seen at regulatory regions in pancreatic cancer and regulates oncogenic pathways. *Genome Res*. 2017; 27:1830–1842. [PubMed: 28986391]
- [37]. Wu Y, Chen X, Zhao Y, Wang Y, Li Y, Xiang C. Genome-wide DNA methylation and hydroxymethylation analysis reveal human menstrual blood-derived stem cells inhibit hepatocellular carcinoma growth through oncogenic pathway suppression via regulating 5-hmC in enhancer elements. *Stem Cell Res Ther*. 2019; 10:151. [PubMed: 31151404]

- [38]. Duan C. Nutritional and developmental regulation of insulin-like growth factors in fish. *J Nutr.* 1998; 128 306S–314S
- [39]. Waerner T, Gardellin P, Pfizenmaier K, Weith A, Kraut N. Human RERE is localized to nuclear promyelocytic leukemia oncogenic domains and enhances apoptosis. *Cell Growth Differ.* 2001; 12:201–210. [PubMed: 11331249]
- [40]. Doherty MJ, Young PR, Cohen PT. Amino acid sequence of a novel protein phosphatase 1 binding protein (R5) which is related to the liver- and muscle-specific glycogen binding subunits of protein phosphatase 1. *FEBS Lett.* 1996; 399:339–343. [PubMed: 8985175]
- [41]. Bieberstein NI, Carrillo Oesterreich F, Straube K, Neugebauer KM. First exon length controls active chromatin signatures and transcription. *Cell Rep.* 2012; 2:62–68. [PubMed: 22840397]
- [42]. Rose AB. Introns as gene regulators: a brick on the accelerator. *Front Genet.* 2018; 9:672. [PubMed: 30792737]
- [43]. Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. *Epigenetics Chromatin.* 2018; 11:37. [PubMed: 29958539]
- [44]. Kamstra JH, Loken M, Alestrom P, Legler J. Dynamics of DNA hydroxymethylation in zebrafish. *Zebrafish.* 2015; 12:230–237. [PubMed: 25751297]
- [45]. Globisch D, Munzel M, Muller M, Michalakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One.* 2010; 5 e15367 [PubMed: 21203455]
- [46]. Santiago M, Antunes C, Guedes M, Sousa N, Marques CJ. TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they? *Genomics.* 2014; 104:334–340. [PubMed: 25200796]
- [47]. Kinde B, Gabel HW, Gilbert CS, Griffith EC, Greenberg ME. Reading the unique DNA methylation landscape of the brain: non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A.* 2015; 112:6800–6806. [PubMed: 25739960]
- [48]. Colombo MN, Francolini M. Glutamate at the vertebrate neuromuscular junction: from modulation to neurotransmission. *Cells.* 2019; 8
- [49]. Lee KH, Park JY, Kim K. NMDA receptor-mediated calcium influx plays an essential role in myoblast fusion. *FEBS Lett.* 2004; 578:47–52. [PubMed: 15581614]
- [50]. Rong R, Ahn JY, Huang H, Nagata E, Kalman D, Kapp JA, Tu J, Worley PF, Snyder SH, Ye K. PI3 kinase enhancer-Homer complex couples mGluRI to PI3 kinase, preventing neuronal apoptosis. *Nat Neurosci.* 2003; 6:1153–1161. [PubMed: 14528310]
- [51]. O'Rourke T, Boeckx C. Glutamate receptors in domestication and modern human evolution. *Neurosci Biobehav Rev.* 2020; 108:341–357. [PubMed: 31743726]
- [52]. Bertolini F, Geraci C, Schiavo G, Sardina MT, Chiofalo V, Fontanesi L. Whole genome semiconductor based sequencing of farmed European sea bass (*Dicentrarchus labrax*) Mediterranean genetic stocks using a DNA pooling approach. *Mar Genomics.* 2016; 28:63–70. [PubMed: 27020381]
- [53]. Anastasiadi D, Piferrer F. Epimutations in developmental genes underlie the onset of domestication in farmed European sea bass. *Mol Biol Evol.* 2019; 36:2252–2264. [PubMed: 31289822]
- [54]. Cui L, Yu WP, DeAizpurua HJ, Schmidli RS, Pallen CJ. Cloning and characterization of islet cell antigen-related protein-tyrosine phosphatase (PTP), a novel receptor-like PTP and autoantigen in insulin-dependent diabetes. *J Biol Chem.* 1996; 271:24817–24823. [PubMed: 8798755]
- [55]. Bence, KK. *Protein Tyrosine Phosphatase Control of Metabolism.* Springer; New York: 2013.
- [56]. Gentilini D, Scala S, Gaudenzi G, Garagnani P, Capri M, Cescon M, Grazi GL, Bacalini MG, Pisoni S, Dicitore A, Circelli L, et al. Epigenome-wide association study in hepatocellular carcinoma: identification of stochastic epigenetic mutations through an innovative statistical approach. *Oncotarget.* 2017; 8:41890–41902. [PubMed: 28514750]
- [57]. Sengelaub CA, Navrazhina K, Ross JB, Halberg N, Tavazoie SF. PTPRN2 and PLCbeta1 promote metastatic breast cancer cell migration through PI(4,5)P2-dependent actin remodeling. *EMBO J.* 2016; 35:62–76. [PubMed: 26620550]

- [58]. Chen PY, Chu A, Liao WW, Rubbi L, Janzen C, Hsu FM, Thamotharan S, Ganguly A, Lam L, Montoya D, Pellegrini M, et al. Prenatal growth patterns and birthweight are associated with differential DNA methylation and gene expression of cardiometabolic risk genes in human placentas: a discovery-based approach. *Reprod Sci.* 2018; 25:523–539. [PubMed: 28693373]
- [59]. Howard AD, Feighner SD, Cully DF, Arena JP, Liberatore PA, Rosenblum CI, Hamelin M, Hreniuk DL, Palyha OC, Anderson J, Paresse PS, et al. A receptor in pituitary and hypothalamus that functions in growth hormone release. *Science.* 1996; 273:974–977. [PubMed: 8688086]
- [60]. Yin Y, Li Y, Zhang W. The growth hormone secretagogue receptor: its intracellular signaling and regulation. *Int J Mol Sci.* 2014; 15:4837–4855. [PubMed: 24651458]
- [61]. Barazzoni R, Bosutti A, Stebel M, Cattin MR, Roder E, Visintin L, Cattin L, Biolo G, Zanetti M, Guarnieri G. Ghrelin regulates mitochondrial-lipid metabolism gene expression and tissue fat distribution in liver and skeletal muscle. *Am J Physiol Endocrinol Metab.* 2005; 288:E228–E235. [PubMed: 15328073]
- [62]. Pfluger PT, Kirchner H, Gunnel S, Schrott B, Perez-Tilve D, Fu S, Benoit SC, Horvath T, Joost HG, Wortley KE, Sleeman MW, et al. Simultaneous deletion of ghrelin and its receptor increases motor activity and energy expenditure. *Am J Physiol Gastrointest Liver Physiol.* 2008; 294:G610–G618. [PubMed: 18048479]
- [63]. Cingolani P, Cao X, Khetani RS, Chen CC, Coon M, Sammak A, Bollig-Fischer A, Land S, Huang Y, Hudson ME, Garfinkel MD, et al. Intronic non-CG DNA hydroxymethylation and alternative mRNA splicing in honey bees. *BMC Genomics.* 2013; 14:666. [PubMed: 24079845]
- [64]. Gao D, Pinello N, Nguyen TV, Thoeng A, Nagarajah R, Holst J, Rasko JE, Wong JJ. DNA methylation/hydroxymethylation regulate gene expression and alternative splicing during terminal granulopoiesis. *Epigenomics.* 2019; 11:95–109. [PubMed: 30427216]
- [65]. Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature.* 2011; 473:394–397. [PubMed: 21552279]
- [66]. Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. Genomewide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* 2011; 25:679–684. [PubMed: 21460036]
- [67]. Szulwach KE, Li X, Li Y, Song CX, Wu H, Dai Q, Irier H, Upadhyay AK, Gearing M, Levey AI, Vasanthakumar A, et al. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat Neurosci.* 2011; 14:1607–1616. [PubMed: 22037496]
- [68]. Gross JA, Pacis A, Chen GG, Drupals M, Lutz PE, Barreiro LB, Turecki G. Gene-body 5-hydroxymethylation is associated with gene expression changes in the prefrontal cortex of depressed individuals. *Transl Psychiatry.* 2017; 7 e1119 [PubMed: 28485726]
- [69]. Petterson A, Chung TH, Tan D, Sun X, Jia XY. RRHP: a tag-based approach for 5-hydroxymethylcytosine mapping at single-site resolution. *Genome Biol.* 2014; 15:456. [PubMed: 25248841]
- [70]. Konstantinidis I, Saetrom P, Mjelle R, Nedoluzhko AV, Robledo D, Fernandes JMO. Major gene expression changes and epigenetic remodelling in Nile tilapia muscle after just one generation of domestication. *Epigenetics.* 2020; 15:1052–1067. [PubMed: 32264748]
- [71]. Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics.* 2017; 18:341. [PubMed: 28464822]
- [72]. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015; 43 e47 [PubMed: 25605792]
- [73]. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 2019; 47:W191–W198. [PubMed: 31066453]
- [74]. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019; 37:907–915. [PubMed: 31375807]

- [75]. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016; 32:3047–3048. [PubMed: 27312411]
- [76]. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014; 30:923–930. [PubMed: 24227677]
- [77]. Liao Y, Smyth GK, Shi W. The subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013; 41 e108 [PubMed: 23558742]
- [78]. R.D.C. Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2015.
- [79]. R. Team. RStudio: Integrated Development for R, RStudio. PBC; Boston, MA: 2020.
- [80]. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15:550. [PubMed: 25516281]

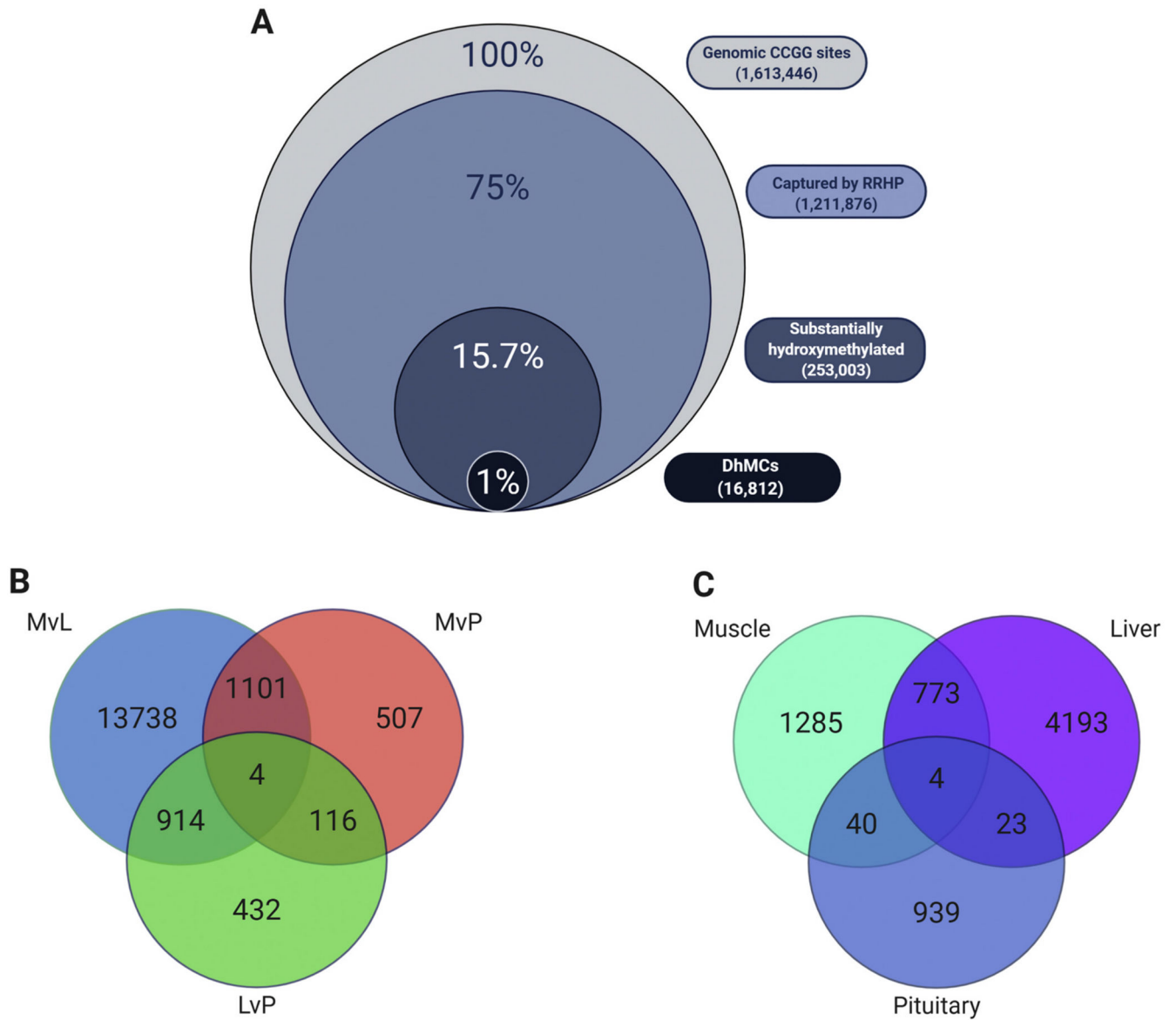


Fig. 1. Venn diagrams representing the RRHP dataset.

A) Stacked Venn diagram showing the distribution of total CCGG sites captured by RRHP, substantially hydroxymethylated and differentially hydroxymethylated among tissues compare to the total genomic CCGG sites of the Nile tilapia genome. B) Venn diagram depicting total differentially hydroxymethylated sites between and among tissue comparisons (MvL – Muscle vs Liver, MvP – Muscle vs Pituitary, and LvP – Liver vs Pituitary). C) Venn diagram showing the total and shared number of hyper-hydroxymethylated genes among tissues. Genes containing or associated with a DhMC in close proximity based on their distance to TSS were considered differentially hydroxymethylated.

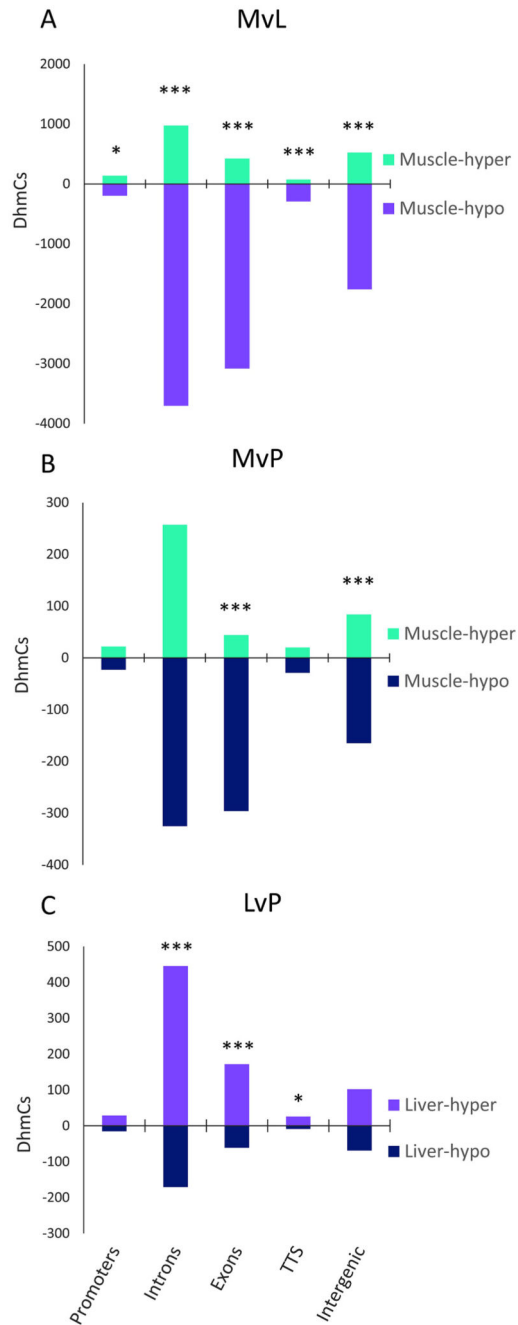
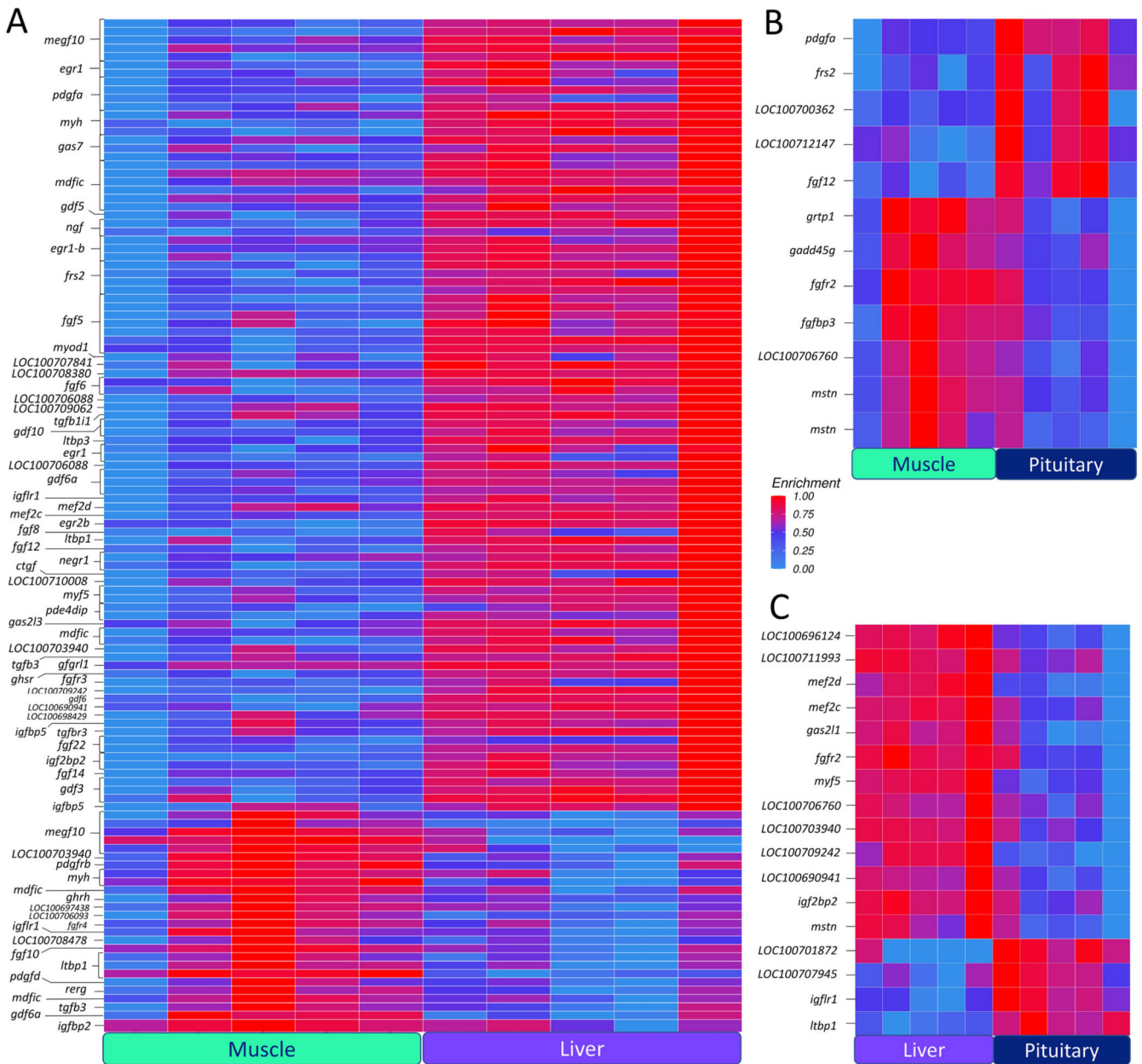


Fig. 2. Histograms depicting the number of differentially hydroxymethylated cytosines (DhmCs) per pairwise comparison and annotated feature. Pairwise tissue comparisons are presented as (A) for muscle versus liver, (B) muscle versus pituitary, and (C) liver versus pituitary. The number of hyper- and hypo-hydroxymethylated sites are shown separately for promoters, introns, exons, transcription termination sites (TTS) and intergenic regions. *P*-values were calculated using Fisher's test with 95% confidence interval and asterisks depict the *p*-value [*** - ($0.01 < p < 0.05$); ***** - ($p < 0.001$)].

**Fig. 3.**

Heatmaps of growth-related genes containing differentially hydroxymethylated cytosines (DhmcCs) among the three tissues. The level of 5hmC enrichment is mapped based on 5hmC counts across samples. The three heatmaps represent the comparisons between (A) muscle and liver, (B) muscle and pituitary, and (C) liver and pituitary. The 5hmC enrichment is represented as a gradient from light blue (low) to red (high). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

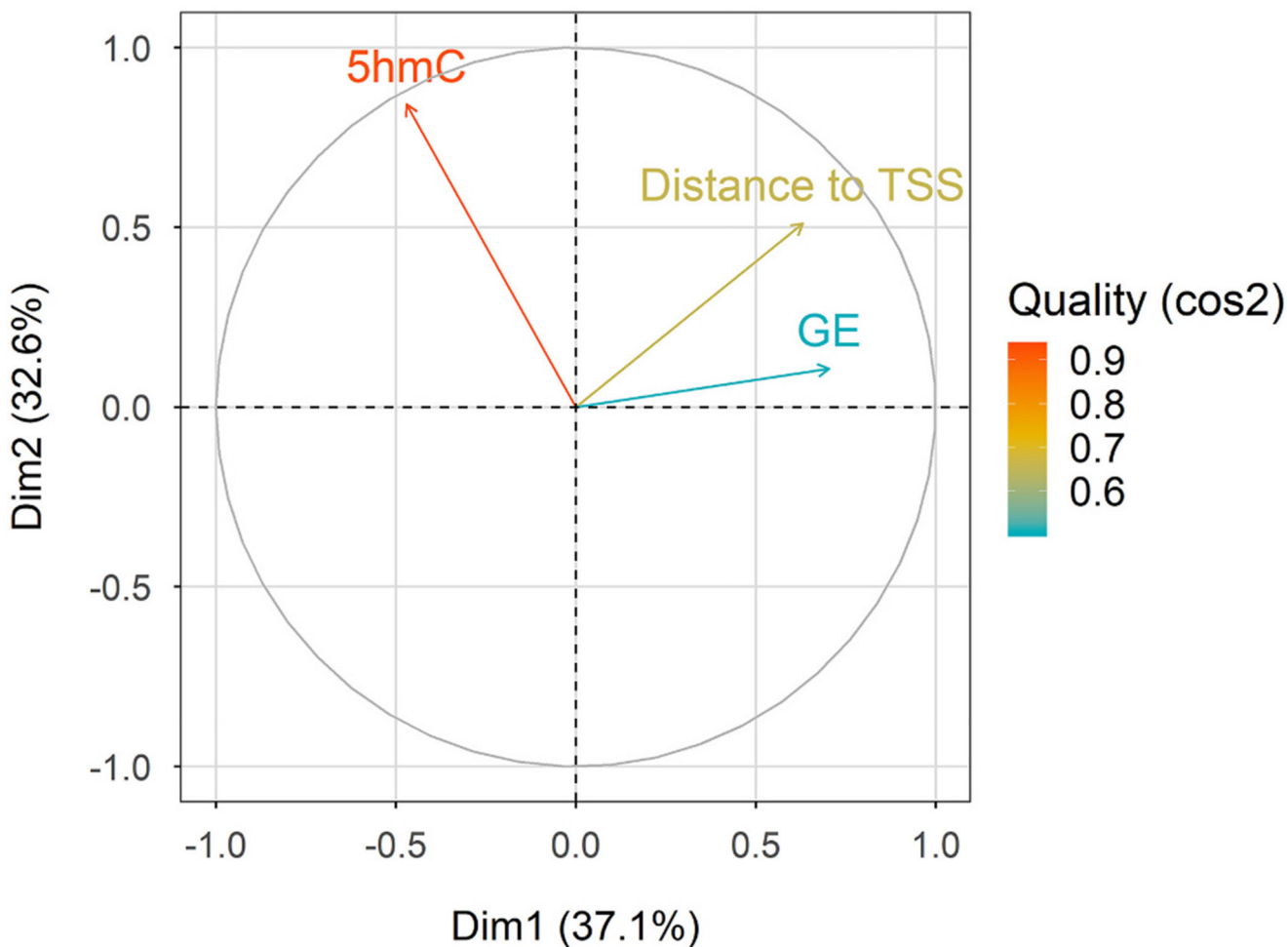


Fig. 4.

Principal component analysis using three variables. The three variables are the 5hmC level, gene expression (GE) corresponding to each 5hmC and their distance to the closest transcription start site, across the first two dimensions that explain 67.5% of the total dataset variation. Colour coding depicts the quality (\cos^2) of representation for the three variables, with lower, intermediate and higher quality represented in blue, yellow and orange, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

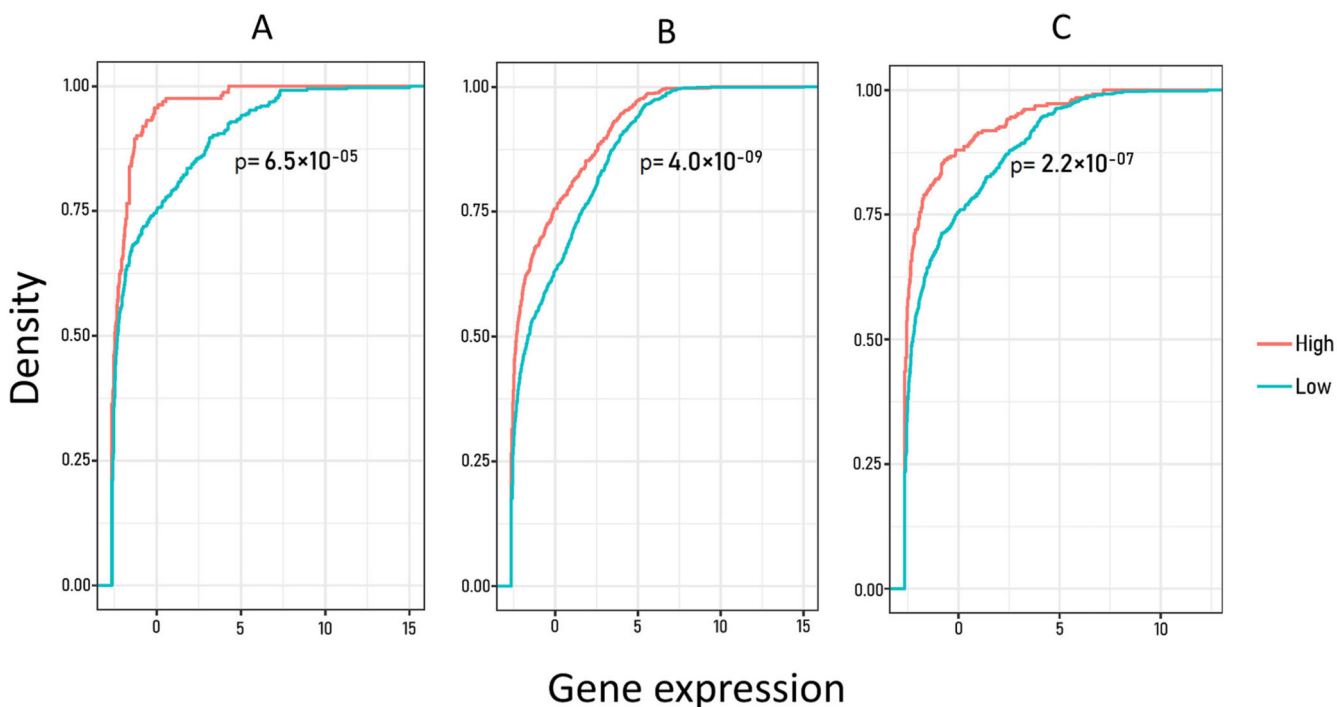
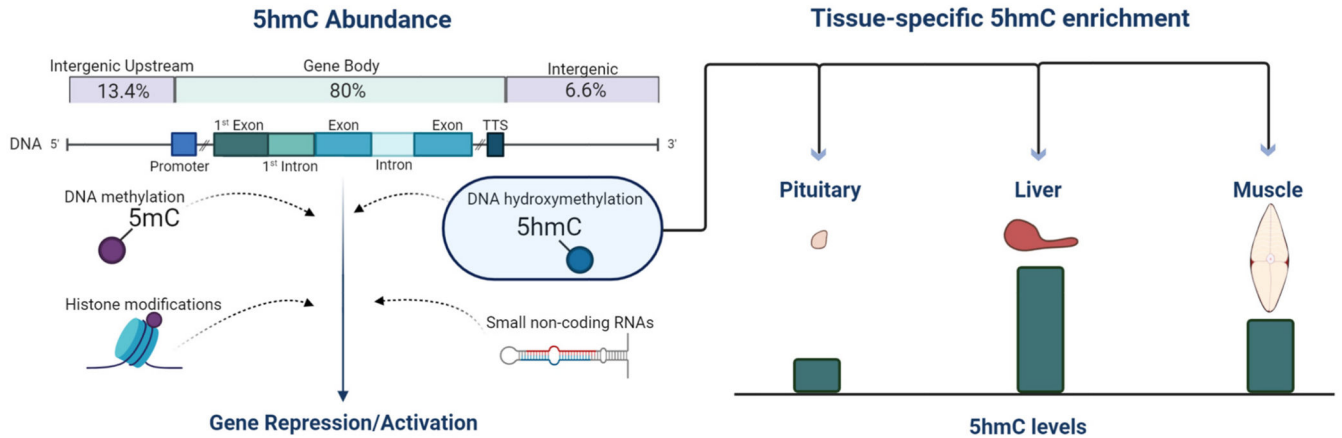


Fig. 5.

Line graph depicting the correlation between highly and lowly hydroxymethylated genes with their expression. Genes were categorized to highly (red) and lowly (light blue) hydroxymethylated based on the mean 5hmC counts across all muscle samples. The total number of genes containing Dhmc is depicted on the y-axis in four density quantiles from 0 to 100%, while the x-axis shows mean gene expression. The three graphs from left to right represent the correlations between gene expression and 5hmC levels within first exons (A), first introns (B) and promoters (C). *P*-values were calculated based on two-sample, two-sided Kolmogorov-Smirnov test. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



DhM genes within the somatotropic axis

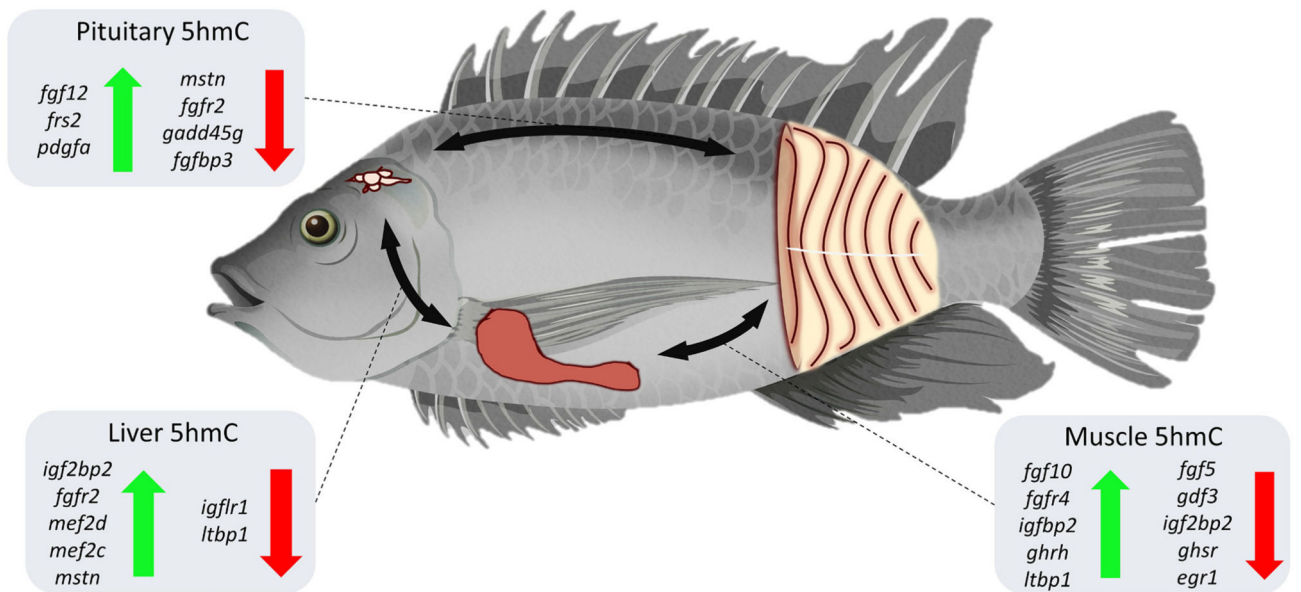


Fig. 6.

Graphical summary of the main findings of the study. DNA hydroxymethylation was enriched within gene bodies and promoters (80%) compared to intergenic regions (20%). Tissues of the somatotropic axis were differentially hydroxymethylated. The liver had the highest, muscle intermediate, and pituitary the lowest 5hmC levels. Several genes involved in somatic growth were found to be differentially hydroxymethylated among all three comparisons (pituitary – muscle, muscle – liver and liver – pituitary).

Table 1

List of DhmcCs between muscle and liver, found within the most expressed genes in muscle. Positive and negative logFC values correspond to hyper- and hypo-hydroxymethylation in muscle compared to liver, respectively.

Chromosome	Position	LogFC	Annotation	Dist. to TSS	Gene Symbol	Description
NC_031972.2	3,582,578	0.70	Intron (1/11)	1318	<i>LOC100707421</i>	troponin T fast skeletal muscle isoforms
NC_031973.2	16,443,052	-1.51	Exon (16/24)	7879	<i>LOC100706607</i>	sarcoplasmic/endoplasmic reticulum calcium ATPase 1
NC_031986.2	22,143,050	-1.53	Exon (5/6)	2333	<i>LOC100698429</i>	myosin light chain 3 skeletal muscle isoform
NC_031969.2	34,504,991	-1.74	Intergenic	10,176	<i>LOC100697135</i>	nucleoside diphosphate kinase B
NC_031969.2	34,505,127	-1.94	Intergenic	10,040	<i>LOC100697135</i>	nucleoside diphosphate kinase B
NC_031973.2	16,442,970	-2.01	Exon (16/24)	7797	<i>LOC100706607</i>	sarcoplasmic/endoplasmic reticulum calcium ATPase 1
NC_031969.2	23,538,241	-2.72	Intergenic	19,650	<i>LOC100698429</i>	myosin heavy chain fast skeletal muscle
NC_031969.2	23,538,340	-3.18	Intergenic	19,551	<i>LOC100698429</i>	myosin heavy chain fast skeletal muscle

Table 2

Sampling information. The list includes measurements of weight, total and standard length of all the individuals used for the current study ($n = 5$) as well as their family and passive integrated transponder tag information.

Sample ID	Family	PIT	Weight (g)	Total length (cm)	Standard length (cm)
M1	flc9	439E4	1001	37.1	29.7
M2	flc9	440D7	683	33.9	26.8
M3	flc9	4296F	1298	37.9	29.1
M4	flc9	4485C	915	35.9	28
M5	flc9	44BE5	708	34.2	26.4