Received: 19 January 2021 | Revised: 14 December 2022 | Accepted: 17 January 2023

DOI: 10.1111/1755-0998.13757

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

Influence of RNA-Seq library construction, sampling methods, and tissue harvesting time on gene expression estimation

Ylenia Chiari[1] | Leif Howard[2,3] | Nickolas Moreno[1] | Scott Relyea[4] | James Dunnigan[4] | Matthew C. Boyer[5] | Marty Kardos[6] | Scott Glaberman[7] | Gordon Luikart[2,3]

[1]Department of Biology, George Mason University, Fairfax, Virginia, USA

[2]Flathead Lake Biological Station, Montana Conservation Genomics Laboratory, Division of Biological Science, University of Montana, Missoula, Montana, USA

[3]Wildlife Biology Program, College of Forestry and Conservation, University of Montana, Missoula, Montana, USA

[4]Sekokini Springs Hatchery, Montana Fish Wildlife and Parks, Bozeman, Montana, USA

[5]Montana Fish Wildlife and Parks, Kalispell, Montana, USA

[6]Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, Washington, USA

[7]Department of Environmental Science and Policy, George Mason University, Fairfax, Virginia, USA

Correspondence
Ylenia Chiari, Department of Biology, George Mason University, Fairfax, VA 22030, USA.
Email: ychiari@gmu.edu

Handling Editor: Andrew P. Kinziger

Abstract

RNA sequencing (RNA-Seq) is popular for measuring gene expression in non-model organisms, including wild populations. While RNA-Seq can detect gene expression variation among wild-caught individuals and yield important insights into biological function, sampling methods can also affect gene expression estimates. We examined the influence of multiple technical variables on estimated gene expression in a non-model fish, the westslope cutthroat trout (*Oncorhynchus clarkii lewisi*), using two RNA-Seq library types: 3′ RNA-Seq (QuantSeq) and whole mRNA-Seq (NEB). We evaluated effects of dip netting versus electrofishing, and of harvesting tissue immediately versus 5 min after euthanasia on estimated gene expression in blood, gill, and muscle. We found no significant differences in gene expression between sampling methods or tissue collection times with either library type. When library types were compared using the same blood samples, 58% of genes detected by both NEB and QuantSeq showed significantly different expression between library types, and NEB detected 31% more genes than QuantSeq. Although the two library types recovered different numbers of genes and expression levels, results with NEB and QuantSeq were consistent in that neither library type showed differences in gene expression between sampling methods and tissue harvesting times. Our study suggests that researchers can safely rely on different fish sampling strategies in the field. In addition, while QuantSeq is more cost effective, NEB detects more expressed genes. Therefore, when it is crucial to detect as many genes as possible (especially low expressed genes), when alternative splicing is of interest, or when working with an organism lacking good genomic resources, whole mRNA-Seq is more powerful.

KEYWORDS
3′ RNA-Seq, dip netting, electrofishing, tissue collection time, transcript length, westslope cutthroat trout, whole mRNA sequencing

1 | INTRODUCTION

RNA sequencing (RNA-Seq) is increasingly common in ecological and evolutionary studies focusing on variation in gene expression (Alvarez et al., 2015; Conesa et al., 2016; Ekblom & Galindo, 2011). It has been used in research on physiology and conservation, and to assess organismal responses to environmental variables (Corlett, 2017; Rey et al., 2020; Todd et al., 2016). RNA-Seq is highly accurate for quantifying expression levels, requires less RNA sample compared to microarrays, does not necessarily require a reference genome (e.g., Cahais et al., 2012), can uncover sequence variation in transcribed regions, and shows high reproducibility (Wang et al., 2009). However, gene expression data can be strongly influenced by biological and nonbiological factors such as experimental and stochastic variation (Auer & Doerge, 2010; Qian et al., 2014; Todd et al., 2016). Given the recent surge in RNA-based studies, it is critical to identify and quantify biological and nonbiological sources of variation in gene expression estimates.

Tissue sampling methods can be an important experimental cause of variation in estimated gene expression (Gayral et al., 2011; Mutch et al., 2008; Passow et al., 2019; Romero et al., 2014). This is a consequence of mRNAs being produced in relatively rapid bursts in response to internal or external stimuli and having short half-lives (Ross, 1995; Staton et al., 2000). Similarly, the use of different anaesthetics, methods of tissue preservation, different RNA extraction methods, and timeframe between sample collection and RNA isolation can all impact RNA quality and gene expression (e.g., Debey et al., 2004; Huitink et al., 2010; Jeffries et al., 2014; Mutter et al., 2004; Olsvik et al., 2007; Passow et al., 2019).

Variation in gene expression due to stochastic variation in cellular and molecular processes can result in random differences among individuals of the same population for the same genes without necessarily being a consequence of microenvironmental variation or other biological factors (e.g., maternal effects and potentially heritable variation). For studies with few biological replicates, this variation may be misinterpreted as biologically relevant (Hansen et al., 2011; Kaern et al., 2005). Mitigation of stochastic variation in gene expression may be achieved through careful sampling design (e.g., individuals vary at only one treatment) and by increasing the number of sampled individuals (Kim et al., 2015; Liu et al., 2014) to gain statistical power (Ching et al., 2014). However, RNA-Seq experiments are often limited in the number of sampled individuals due to cost, with a consequent loss of statistical power and potentially misleading results (Bi & Liu, 2016; Li et al., 2013).

Higher sequencing cost has led to the development of RNA library construction protocols that allow processing and sequencing of a larger number of samples in a more cost-effective manner (Meyer et al., 2011; Morrissy et al., 2009; Wu et al., 2010). 3′ RNA-Seq methods only prime the 3′ poly-A tail, thus reducing the sequencing effort and cost (Lohman et al., 2016; Ma et al., 2019). However, whole mRNA-Seq libraries have been reported to have higher power in detecting transcripts, especially for genes with low expression (Jarvis et al., 2020) and that are differentially expressed (see Crow et al., 2022 for a review on pros- and cons- of each library type; Jarvis et al., 2020; Ma et al., 2019; Mandelboum et al., 2019; Tandonnet & Torres, 2017). Whole mRNA-Seq libraries also permit identification of alternative splicing at a single gene, as library and sequencing with this approach captures different fragments and transcripts for the same locus (Crow et al., 2022).

In many species including fish, RNA-Seq data are commonly used to investigate the effects of environmental variables (e.g., temperature, hypoxia) on gene expression (e.g., Jeffries et al., 2021; Krishnan et al., 2020; Long et al., 2015; Meyer et al., 2011; Smith et al., 2013; Wang et al., 2015). However, there is little known about the influence of different methods used to sample individuals on gene expression, especially under field conditions. Conditions in the field may limit the use of optimal sampling protocols or storage methods (Mutter et al., 2004; Pérez-Portela & Riesgo, 2013). Handling time of individuals before tissue sampling may also be longer than in the laboratory and affect gene expression differently depending on the field sampling technique and tissue used.

The impacts of stress from handling on fish physiology are well understood (Sopinka et al., 2016). Although most studies focus on glucocorticoid and blood chemistry responses following capture (Barton, 2002; Milla et al., 2010; Milligan, 1996; Ruane et al., 2001; Wiseman et al., 2007; Wood et al., 1983; see also Romero & Reed, 2005 for influence on handling time of nonfish species), gene expression responses to handling stress indicate that the magnitude, intensity, and duration of changes vary across genes, species, and tissue types (Krasnov et al., 2005; Lopez-Patino et al., 2014). While there is evidence that blood cortisol and glucose levels are affected by capture method (e.g., electrofishing), to our knowledge (Barton & Dwyer, 1997; Barton & Grosh, 1996; Bracewell et al., 2004), it is still unclear whether gene expression is affected by capture method or the amount of handling or processing time prior to sample collection.

Here, we test whether sampling method (electrofishing vs. dip netting), processing time, and RNA-Seq library type (3′ RNA-Seq—hereafter called QuantSeq—vs. whole mRNA-Seq—hereafter called NEB) influence gene expression data in multiple tissue types from westslope cutthroat trout (*Oncorhynchus clarkii lewisi*), a species of conservation concern native to western North America (Allendorf & Leary, 1988; Shepard et al., 2003). Electrofishing, which consists of a backpack mounted electrofishing unit that applies an electrical current in the water to momentarily stun the fish, is one of the most common fisheries sampling methods. This method may cause the fish to express genes in response to the electric current, and may affect individual fish and tissue types differently, increasing variation among biological replicates. An alternative to electrofishing is dip netting. Nets may potentially have less of an effect on gene expression and a lower risk of inadvertently killing both target and nontarget organisms. However, sampling by netting is more laborious, time consuming, and less effective in the field where circumstances may not allow for long sampling periods or aquatic systems may have obstacles that prevent effective capture with nets (e.g., fallen tree limbs and rocks). Capturing fish by dip netting may still influence gene expression through stress, as the fish tries to escape capture.

The results of this study will provide a foundation for improving future RNA-based study designs for field sampling of wild caught non-model fish and other species.

2 | MATERIALS AND METHODS

2.1 | Sample collection, group assignment, and tissue harvesting

All samples of westslope cutthroat trout were collected on a single day in May 2019 at the Montana Fish, Wildlife, & Parks Sekokini Springs Hatchery in West Glacier, MT (USA). We collected 30 fish divided into three treatment groups (10 fish per group), each with three tissues, for a total of 90 samples as follows (Tables 1 and 2, Table S1 available on Dryad): group 1 = dip netting sampling with immediate tissue harvest (*n* samples = 30); group 2 = electrofishing with immediate tissue harvest (*n* samples = 30); group 3 = electrofishing, tissue harvested from fish 5 min after death by pithing (see below, *n* samples = 30). All fish were fry (1 year old; fish were non-sexually mature as they cannot be sexed until 2–4 years old) from the same breeding stock and were offspring (F1) from wild parents from Danaher Creek (MT). Average size (as total body length) of fish was 108 mm ± 11 and average weight was 10 g ± 3. Total body length of the fish was measured from the most forward point of the head, with the mouth closed, to the farthest tip of the tail, with the tail compressed or squeezed, while the fish was on its side.

Fish were gently netted five at a time from the raceway into a 100 gallon holding tank containing hatchery system water. Fish were then either captured from the tank by net or electrofished with a backpack electrofishing unit set to 150 volts with a standard pulse for a duration of 3 s. Duration of time in this tank was maintained consistent independently of whether the fish were sampled by electrofishing or netting and was approximately 1–2 min. Captured fish were immediately transferred to 5 gallon holding buckets with hatchery system water and euthanized by pithing and processed for tissue harvesting, except for group 3. Fish were not sedated at any point during the sampling steps. Average time in the 5 gallon bucket was approximately 2 min before euthanasia and average time of tissue collection after pithing was approximately 3 min, except for group 3, for which tissue harvesting began 2–3 min after the 5 min from pithing. Fish from group 3 were sampled in the same way as fish from group 2, except that after pithing they were placed in a separate holding bucket of water for 5 min before tissue harvesting to test for the influence of delayed tissue harvesting. Total body length and weight data were collected for each fish. Sample information, including times of tissue harvest after euthanasia for each sample, can be found in Table S1 on Dryad.

Tissue removal was performed using single use scalpels on a nylon cutting board. Tissue samples from each fish were collected in the following order: blood, dorsal muscle, and gills. We first collected the blood immediately before euthanasia as coagulated blood may affect RNA quality (Chiari & Galtier, 2011). To obtain the blood sample, the tail was removed by a diagonal cut made through the caudal peduncle from dorsally anterior of the anal fin to ventrally posterior of the anal fin to avoid intersecting the gastrointestinal tract. Slight pressure was applied to the body of the fish and blood was allowed to drip out of the cut directly into the 2 mL tube. Muscle tissue was sliced into smaller pieces to allow penetration of the preservative (Gayral et al., 2011). Sampling tools and the cutting board were thoroughly cleaned with 10% bleach first and then purified water between fish to avoid sample and tissue contamination. Tissue samples were placed in 2 mL sterile tubes filled with RNAlater (Qiagen) for preservation. Tubes were left at room temperature overnight and then stored at −80°C (or in dry ice for transportation) until the RNA extraction was carried out. All sampling was carried out according to the Institutional Animal Care and Use Committee (IACUC) approved permit no. AUP 007-19GLFLBS-062819 to GL.

2.2 | RNA extraction

RNA extractions and the following laboratory procedures described below were carried out by a private company (Admera Health). The same extraction protocol was used for each of the different tissues and generally followed the manufacturer's instructions for QIazol (Qiagen) extraction. Briefly, up to 10 mg of tissue was mechanically homogenized in 500 μL of QIazol. After homogenization, QIazol was added to reach 1 mL and then 200 μL of chloroform was added and mixed. Blood samples were centrifuged at 2000 *g* for 5 min; the supernatant was discarded and 1 mL of QIazol was added to the tube. Tubes with blood samples were then left at room temperature for 5 min and vortexed to ensure homogenization of the sample. Then, 200 μL of chloroform was added and mixed. All samples (blood or other tissues, all containing 1 mL of QIazol and 200 μL of chloroform) were then incubated at room temperature for 3–5 min and centrifuged at 4°C and 12,000 *g* for 15 min. The upper aqueous RNA containing phase was transferred to a new tube. An equal volume of 70% ethanol was added and mixed. The mixture was loaded into a RNeasy mini prep column (Qiagen RNeasy Mini Plus Kit) and RNA eluted following the manufacturer's protocol.

The quality of RNA was determined by Qubit HS RNA assay (ThermoFisher), and the integrity of RNA was evaluated based on RIN (RNA integrity number), which varies between 1–10 with 10 indicating no degradation. RIN values were acquired via capillary gel electrophoresis performed using a Bioanalyser 2100 (Agilent Technologies). ANOVA was run in R using the *F*-test to compare RIN numbers among samples belonging to different treatment groups and to compare RIN numbers among samples belonging to different tissues in each group.

2.3 | RNA library preparation and sequencing

Since variation in RNA quality may affect downstream results (Passow et al., 2019), library construction and sequencing

TABLE 1 Samples and RNA quality.

Sample ID	QuantSeq	NEB ID	Group no.	Tissue	Sampling method	RIN	Concentration (ng/μL)
1_1_1	19101XR-01-41		1	Blood	dip net	9	60.3
1_1_3	19101XR-01-42		1	Muscle	dip net	9.2	188.5
1_1_5	19101XR-01-44		1	Gill	dip net	9.3	68.9
1_10_1	19101XR-01-77	19101XR-01-77NEB	1	Blood	dip net	9.8	166.1
1_10_3	19101XR-01-78		1	Muscle	dip net	8.9	586.1
1_10_5	19101XR-01-80		1	Gill	dip net	9.2	784
1_2_1	19101XR-01-45		1	Blood	dip net	9.4	51.8
1_2_3	19101XR-01-46		1	Muscle	dip net	10	150.7
1_2_5	19101XR-01-48		1	Gill	dip net	9.3	361.9
1_3_1	19101XR-01-49		1	Blood	dip net	9.2	75.3
1_3_3	19101XR-01-50		1	Muscle	dip net	9.6	307.3
1_3_5	19101XR-01-52		1	Gill	dip net	9.5	69.3
1_4_1	19101XR-01-53	19101XR-01-53NEB	1	Blood	dip net	9.5	80.8
1_4_3	19101XR-01-54		1	Muscle	dip net	8.9	224.4
1_4_5	19101XR-01-56		1	Gill	dip net	9.2	201.8
1_5_1	19101XR-01-57	19101XR-01-57NEB	1	Blood	dip net	9.6	195.2
1_5_3	19101XR-01-58		1	Muscle	dip net	9	90.7
1_5_5	19101XR-01-60		1	Gill	dip net	8.8	44.5
1_6_1	19101XR-01-61	19101XR-01-61NEB	1	Blood	dip net	9.6	64.8
1_6_3	19101XR-01-62		1	Muscle	dip net	8.8	122.5
1_6_5	19101XR-01-64		1	Gill	dip net	9.2	91
1_7_1	19101XR-01-65		1	Blood	dip net	9.4	88
1_7_3	19101XR-01-66		1	Muscle	dip net	9.4	248.8
1_8_1	19101XR-01-69	19101XR-01-69NEB	1	Blood	dip net	9.4	154.3
1_8_3	19101XR-01-70		1	Muscle	dip net	8.8	128.5
1_8_5	19101XR-01-72		1	Gill	dip net	9.2	575.4
1_9_1	19101XR-01-73		1	Blood	dip net	9.7	40.5
1_9_5	19101XR-01-76		1	Gill	dip net	9.3	483.3
3_1_1	19101XR-01-121	19101XR-01-121NEB	2	Blood	efishing	9.6	105.5
3_1_3	19101XR-01-122		2	Muscle	efishing	8.8	384.8
3_1_5	19101XR-01-124		2	Gill	efishing	9.5	489.9
3_10_1	19101XR-01-157	19101XR-01-157NEB	2	Blood	efishing	9.8	92.7
3_10_3	19101XR-01-158		2	Muscle	efishing	9.4	234.3
3_10_5	19101XR-01-160		2	Gill	efishing	9.1	526.7
3_2_3	19101XR-01-126		2	Muscle	efishing	9.1	388.6
3_2_5	19101XR-01-128		2	Gill	efishing	9.2	325
3_3_3	19101XR-01-130		2	Muscle	efishing	9.3	241.3
3_3_5	19101XR-01-132		2	Gill	efishing	9.2	124.6
3_4_1	19101XR-01-133	19101XR-01-133NEB	2	Blood	efishing	9.6	151.3
3_4_3	19101XR-01-134		2	Muscle	efishing	8.8	294.2
3_4_5	19101XR-01-136		2	Gill	efishing	9.1	158.1
3_5_1	19101XR-01-137	19101XR-01-137NEB	2	Blood	efishing	9.8	122
3_5_3	19101XR-01-138		2	Muscle	efishing	9.5	225.2
3_5_5	19101XR-01-140		2	Gill	efishing	9.2	238.1
3_6_3	19101XR-01-142		2	Muscle	efishing	9.4	204.8

TABLE 1 (Continued)

Sample ID	QuantSeq	NEB ID	Group no.	Tissue	Sampling method	RIN	Concentration (ng/μL)
3_6_5	19101XR-01-144		2	Gill	efishing	8.8	172.2
3_7_1	19101XR-01-145		2	Blood	efishing	9.8	195.7
3_7_3	19101XR-01-146		2	Muscle	efishing	9.2	329.3
3_7_5	19101XR-01-148		2	Gill	efishing	9.2	270.4
3_8_1	19101XR-01-149		2	Blood	efishing	9.8	139.5
3_8_3	19101XR-01-150		2	Muscle	efishing	9.4	206.8
3_8_5	19101XR-01-152		2	Gill	efishing	9.2	299.4
3_9_1	19101XR-01-153	19101XR-01-153NEB	2	Blood	efishing	9.6	199.7
3_9_3	19101XR-01-154		2	Muscle	efishing	9.5	161.6
5_1_1	19101XR-01-201		3	Blood	efishing	9.1	20
5_1_3	19101XR-01-202		3	Muscle	efishing	8.9	167.3
5_1_5	19101XR-01-204		3	Gill	efishing	9.4	129.7
5_10_1	19101XR-01-237	19101XR-01-237NEB	3	Blood	efishing	9.6	133.3
5_10_3	19101XR-01-238		3	Muscle	efishing	9.4	642.9
5_10_5	19101XR-01-240		3	Gill	efishing	9.4	147.7
5_2_1	19101XR-01-205		3	Blood	efishing	9.7	66.6
5_2_3	19101XR-01-206		3	Muscle	efishing	9	506.6
5_3_1	19101XR-01-209	19101XR-01-209NEB	3	Blood	efishing	9.6	222.1
5_3_3	19101XR-01-210		3	Muscle	efishing	9.4	275
5_3_5	19101XR-01-212		3	Gill	efishing	8.9	935.2
5_4_1	19101XR-01-213		3	Blood	efishing	9.7	71
5_4_3	19101XR-01-214		3	Muscle	efishing	9.1	280
5_4_5	19101XR-01-216		3	Gill	efishing	9.2	570.5
5_5_5	19101XR-01-220		3	Gill	efishing	9.4	370.9
5_6_1	19101XR-01-221	19101XR-01-221NEB	3	Blood	efishing	9.7	73.7
5_6_3	19101XR-01-222		3	Muscle	efishing	9.3	277.7
5_6_5	19101XR-01-224		3	Gill	efishing	9.2	233
5_7_1	19101XR-01-225		3	Blood	efishing	9.4	112.2
5_7_3	19101XR-01-226		3	Muscle	efishing	9.6	31.7
5_7_5	19101XR-01-228		3	Gill	efishing	9.1	96.9
5_8_1	19101XR-01-229		3	Blood	efishing	9.7	68.9
5_8_3	19101XR-01-230		3	Muscle	efishing	9.6	156.1
5_8_5	19101XR-01-232		3	Gill	efishing	9.5	68.7
5_9_1	19101XR-01-233	19101XR-01-233NEB	3	Blood	efishing	9.7	212.3
5_9_3	19101XR-01-234		3	Muscle	efishing	9.8	60.5
5_9_5	19101XR-01-236		3	Gill	efishing	9	115.5

Note: List of all the samples on which transcriptomic data were obtained in this study (sample ID and Admera Health ID for QuantSeq and NEB) with information about the treatment group they belong to, tissue type, sampling method, RIN value, and RNA concentration (ng/μL). A full list of samples and relative information for which RNA was extracted can be found in Table S1 on Dryad.

were carried out for 81 tissue samples with a RIN value ≥ 8.8 for QuantSeq and a subset of 14 blood samples (for which we also had QuantSeq data) with RIN ≥ 9.4 for NEB (Tables 1 and 2, Table S1 on Dryad). None of these samples showed signs of RNA degradation based on the BioAnalyser profile (Tables 1 and 2, Table S1 on Dryad). Whole mRNA libraries (NEB) were made for 14 selected blood samples with similar RIN and concentrations among compared groups (Tables 1 and 2). Library preparation was performed with the NEB Ultra II RNA library prep kit with NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs). For 3′ RNA-Seq, library preparation was performed with QuantSeq 3′ mRNA-Seq Library Preparation Kit FWD for Illumina (Lexogen). All procedures were performed according to the manufacturer's suggested protocols. The quantity and molecular size of the libraries were confirmed by

TABLE 2 Sample size used for each comparison.

(a)			
Comparisons	**Blood**	**Gills**	**Muscle**
Group 1 vs. group 2 QuantSeq			
$N_{tot} = 54$	$N_{tot} = 17$	$N_{tot} = 18$	Ntot = 19
$N_1 = 28$	$N_1 = 10$	$N_1 = 9$	$N1 = 9$
$N_2 = 26$	$N_2 = 7$	$N_2 = 9$	$N2 = 10$
Group 2 vs. group 3 QuantSeq			
$N_{tot} = 53$	$N_{tot} = 16$	$N_{tot} = 18$	Ntot = 19
$N_2 = 26$	$N_2 = 7$	$N_2 = 9$	$N2 = 10$
$N_3 = 27$	$N_3 = 9$	$N_3 = 9$	$N3 = 9$

(b)	
NEB	**Comparison for blood only**
$N_{tot} = 14$	Group 1 vs. group 2
	$N_{tot} = 10$
	$N_1 = 5$
	$N_2 = 5$
	Group 2 vs. group 3
	$N_{tot} = 9$
	$N_2 = 5$
	$N_3 = 4$

(c)
Comparison of QuantSeq vs. NEB only for blood samples
$N_{tot} = 28$
$N_{QuantSeq} = 14$
$N_{NEB} = 14$

Note: Sample size for the samples for which transcriptomic data were obtained divided by tissue type, treatment group, and library type (Table 2a QuantSeq, Table 2b NEB, Table 2c QuantSeq vs. NEB) is also indicated. Group number is as in Table 1: Group 1, dip net with tissue harvested immediately after death; Group 2, efishing with tissue harvested immediately after death; Group 3, efishing with tissue harvested 5 min after death.

Qubit HS DNA assay (ThermoFisher) and Tapestation 2200 system coupled with High Sensitivity D1000 ScreenTapes (Agilent). Sequencing was performed on Illumina Hiseq X with 150 bp pair-end reads for all QuantSeq samples (Lexogen) and four NEB samples, while the remaining 10 NEB samples were sequenced on a NovaSeq machine. Raw reads were deposited on NCBI (SRA PRJNA691889).

2.4 | RNASeq reads check and genome coverage

Quality checks of the raw RNA-Seq reads were performed using FASTQC (Andrews, 2014). Reads were trimmed with TRIMMOMATIC using the default parameters (version 0.38, Bolger et al., 2014; see specific code file on Dryad). Raw reads were mapped to an *Oncorhynchus mykiss* reference genome from NCBI (Omyk_1.0, https://www.ncbi.nlm.nih.gov/assembly/GCF_002163495.1/, Annotation release ID:100) using STAR (version 2.7.1a; Dobin et al., 2013; Dobin & Gingeras, 2015; see specific code file on Dryad) to obtain the number of genes detected by each library type, QuantSeq versus NEB.

In order to perform bioinformatic analyses on samples with an equal percentage of uniquely mapped reads across samples and between library types (as previously done by others; see Ma et al., 2019), we randomly selected 11 million and 40 million reads per sample (step done prior mapping of reads using STAR) for all analyses performed on QuantSeq and NEB, respectively. Previous studies have shown that using >10 M reads does not increase the percentage of uniquely mapped reads, after which the ability to detect differently expressed genes becomes independent of sequencing depth (Jarvis et al., 2020; Liu et al., 2014; Ma et al., 2019; see also Crow et al., 2022 for an in-depth discussion on the issue of redundancy of reads). However, as detection of differentially expressed genes increases with the number of reads between 10 and 20 M reads in whole mRNA-Seq (Jarvis et al., 2020), we randomly selected more reads for NEB (40 M) than QuantSeq (11 M) in our work. Transcripts are randomly sheared into fragments with NEB but not QuantSeq. Consequently, the number of reads with NEB are proportional to the number of fragments not transcripts, whereas the number of reads with QuantSeq is proportional to the number of transcripts. Because of this, more reads may be needed for NEB than for QuantSeq to have a similar percentage of uniquely mapped reads.

Reads were mapped to the *O. mykiss* reference genome. HT-SEQ (version 0.11.1; Anders et al., 2015; see specific code file on Dryad) was then used to quantify the number of reads uniquely (unambiguously) mapped to a single gene of the *O. mykiss* reference genome. Finally, a python script provided with STRINGTIE (prepDE.py) was used to generate a gene counts matrix (Pertea et al., 2016).

2.5 | Similarity in gene expression among samples

To assess the variation and direction of variation among samples based on their gene expression, we calculated the correlation of gene expression levels among samples and the Euclidean distances among samples in DESEQ 2 (version 1.22.2; Love et al., 2014). These measures are especially useful to assess the similarity of biological replicates (Koch et al., 2018) and therefore to detect anomalies among samples belonging to the same treatment group. The sample correlation matrix was calculated by computing the Pearson's correlation of the normalized matrix after the variance stabilizing transformation (VST was performed on the most variable 2000 genes based on the HTSeq data produced). VST allows taking into account the sample variability of low counts. The most variable 2000 genes were used only for the sample-to-sample distances and the heatmap, while differential expression analyses were run taking into account all genes, not just these 2000 most variable genes.

Following DESeq 2, Pearson's correlation was calculated in pairwise comparison between samples. Pearson's correlation ranges from −1 to 1, where a value of 0 indicates no correlation (gene

expression is completely dissimilar between the two samples), while values of 1 indicate that samples have identical expression level (and −1 corresponds to negative correlation). The Euclidean distance between sample expression profiles was calculated by the equation $dist = sqrt(1 - cor^2)$, where *cor* stands for the correlation coefficient of two samples. The smaller the distance, the higher the correlation between samples. These distances were then used to build the heatmaps of sample distance of each normalized matrix, which allows the data to be shrunken towards the genes' average expression across all samples. Gene heatmaps were instead based on VST transformation to normalize the raw count. After this, the mean expression in each sample is then normalized to zero. Finally, differences in gene expression among the studied groups were visualized by a PCA plot using the gene count matrix after applying the VST to normalize the raw counts. PCA plots are useful to assess the effect of covariates and batch effects (nonbiological variation due to experimental artefacts [Reese et al., 2013]).

2.6 | Differential gene expression analysis

Differences in gene expression among groups were identified by differential expression analysis using DESeq2 on raw read counts (the input data were non-normalized, as suggested by DESeq2). The false discovery rate (FDR) was adjusted to 0.05, corresponding to a recovery at most of 5% of false positives following the DESeq2 manual. We used the default options for all other parameters. We looked at differences in gene expression between sampling methods, harvest tissue time, tissue type, and QuantSeq versus NEB in Table 2 (see Tables 1 and 2, Table S1 on Dryad for detailed information about comparisons and sample sizes for each comparison; minimum $N = 4$). The log2 fold changes obtained from DESeq2 were used as a measure of how much more (or less) genes are expressed in one group versus the other. Genes had a different expression if the adjusted *p*-value (using the adjusted *p*-value results in less false positives) was <.05. For all differential gene expression comparisons among library types, sampling methods, and tissue types, we only analysed genes with baseMean values >0 for both groups being compared, as a baseMean of 0 could either mean that there was truly no expression or that the transcript could not be detected in that sample. In DESeq2, baseMean is the average of the normalized count values, dividing by size factors, taken over all samples. We also eliminated all genes that returned an adjusted *p*-value of NA by DESeq2 for a gene comparison. DESeq2 returns NA for expressed genes when there are low mean normalized counts of mapped reads or outlier values.

Finally, previous work has indicated an increase in read count for longer transcripts using NEB compared to QuantSeq (Ma et al., 2019; but see Crow et al., 2022). To further evaluate this potential difference among library types, we compared transcript length and gene expression between QuantSeq and NEB using the longest known transcript length for each gene from *O. mykiss* in Ensembl 108 (Yates et al., 2020). For this analysis, we used all data collected on the same 14 blood samples, independently of treatment group, for which we had QuantSeq and NEB data. We also used the same approach to specifically assess if transcript length influences the absence of gene expression or non-detection (baseMean = 0 in DESeq2 output) in one library type but not the other.

3 | RESULTS

3.1 | RNA and raw sequencing data quality statistics

Out of the 90 samples for which RNA was extracted, 86 had a RIN value equal or above 8.8. Little variation in RIN scores was observed among sampled tissues and sampling methods (Table 1 and Table S1 on Dryad). Mean and standard deviation for RIN values for the three tissues were 9.6 ± 0.22 (blood), 9.2 ± 0.40 (muscle), 9.0 ± 1 (gill). Mean and standard deviation for RIN values for the three treatment groups were 9.2 ± 0.43 (dip netting), 9.3 ± 0.34 (electrofishing), and 9.2 ± 1.06 (tissue harvesting after 5 min). We found no differences in RIN values among groups ($F = 0.299$, df = 2, $p = .74$) or among tissues within each group ($F = 0.595$, df = 4, $p = .67$, Table S1 on Dryad).

The final number of reads per individual for QuantSeq libraries ranged from 11 million to 15.6 million (mean = 12.88 million ± 0.67). On average, of the 11 million reads randomly selected for each sample, around 77% of QuantSeq reads were uniquely mapped onto the rainbow trout (*O. mykiss*) genome independently of the sampling method used (range: 67.7%–86.3%; Table S1 on Dryad). The final number of reads per individual for NEB ranged from 77.8 to 148.8 million reads (mean = 105.6 million ± 19.1). On average, of the 40 million reads randomly selected for each sample, 75% of NEB reads were uniquely mapped onto the *O. mykiss* genome (Table S1 on Dryad).

Based on DESeq2 output for each of the two RNA-Seq library types built from the same 14 blood samples, we found that NEB detected 31% more expressed genes than QuantSeq (Table S2 on Dryad). A gene was considered to be detected/expressed when baseMean was >0. Specifically, we found that NEB and QuantSeq detected 34,715 and 26,465 expressed genes, respectively, which mapped onto the annotated *O. mykiss* genome. A total of 25,396 expressed genes were detected by both library types. However, 9319 expressed genes were detected by NEB but not QuantSeq, and 1069 expressed genes were detected by QuantSeq but not NEB. Presence/absence of expressed genes detected by one or the other library type was not dependent on gene transcript length (Figure 1).

Comparing the number of genes detected among tissue types (i.e., genes with a baseMean > 0) with QuantSeq, we found that blood recovered the fewest genes (23,731 to 25,770 genes across treatment groups), followed by muscle (31,553 to 33,071 genes across treatment groups), with the largest number of genes detected in gill (35,193 to 36,036 genes across treatment groups).

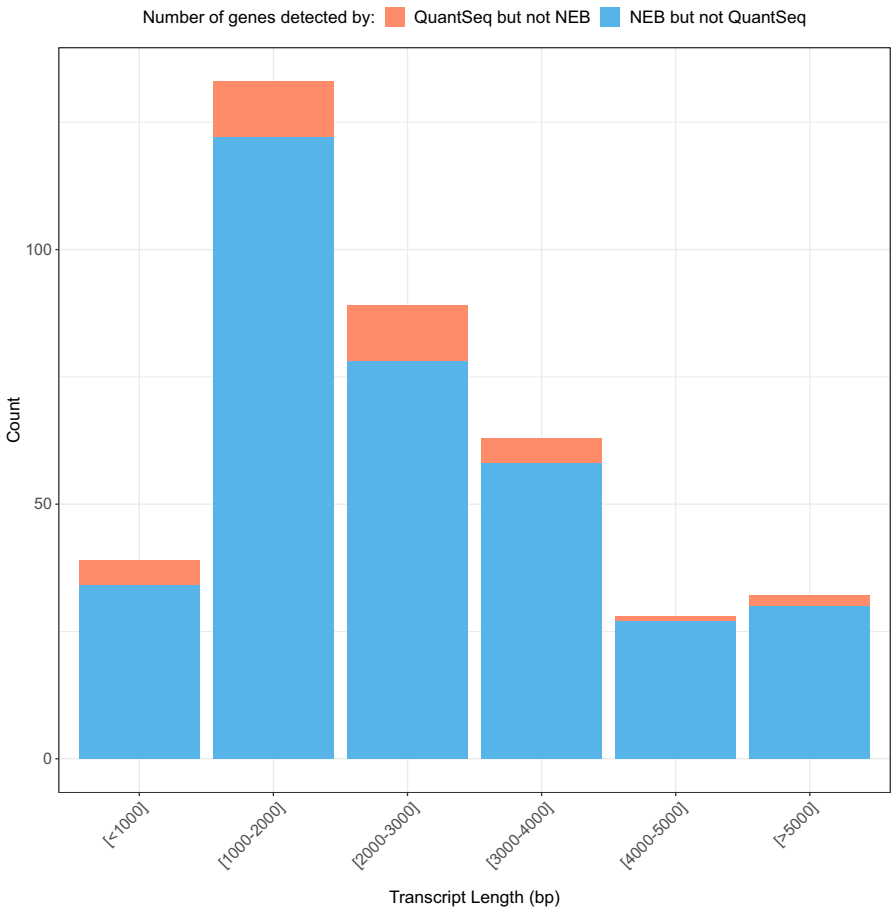

FIGURE 1 Bar plot of transcript length versus number of nonexpressed/nondetected genes for each RNA-Seq library type. Data based on the 14 blood samples processed and sequenced using both library types. The plot only depicts gene that have been annotated with a known transcript length. Transcript length is based on the longest transcript for a gene. Bar plot made using R (2019)

3.2 | Similarity in gene expression among samples

Similarity in gene expression among biological replicates (i.e., individuals belonging to the same treatment group) gives an idea of reproducibility of our data and of the overall variation among samples. Similarity in gene expression within and among groups can be estimated using the sample correlation or Euclidean distances (see Section 2). Pearson's correlation coefficients (r) for biological replicates were ≥0.9 for 97% of comparisons (same tissue within a treatment group) (Table S3 on Dryad). This indicates that although variation in gene expression occurs among individuals, biological replicates were generally very similar. Pearson's r values between the two sequencing platforms for NEB were all >0.9 for blood samples belonging to the same group (Table S3), indicating that different sequencing methods did not influence the number of uniquely mapped reads. Pearson's r among different tissues (for QuantSeq) were generally <0.5 and sometimes even negative, indicating different levels of gene expression among tissues. Finally, r values among QuantSeq versus NEB (same blood samples only) were typically <0.5 indicating that the two library types do not always recover equivalent gene expression for the same genes.

Heatmaps of the distance matrices for the different treatment group comparisons provide hierarchical clustering based on sample distances. When heatmaps were made using combined data from all three tissues for QuantSeq, we found three clusters corresponding to the three different tissue types (Figure 2a,b). However, within each cluster, as also shown by the heatmaps built with data from each tissue separately, samples belonging to different treatment groups are clustered together, indicating no clear difference in gene expression among the tested groups (Figure S1). Lack of difference in gene expression among the different treatment groups was also found using NEB data (Figure 2c,d).

Finally, comparison of QuantSeq versus NEB found differences in gene expression between the two library types; however, this difference was not associated with any of the treatment groups (Figure 2e,f). Principal component analysis (PCA)–another way to visualize variation in gene expression among samples–further supports the lack of differences among sampling methods and time of tissue harvesting and the differentiation between QuantSeq versus NEB and among the three sampled tissues (Figures 3 and 4, Figure S2).

3.3 | Differential gene expression

3.3.1 | Dip netting versus electrofishing sampling method

QuantSeq data identified only three out of 35,772 genes (0.008%) that were significantly (adjusted p-value <.05) differentially expressed between dip netting and electrofishing across all tissue

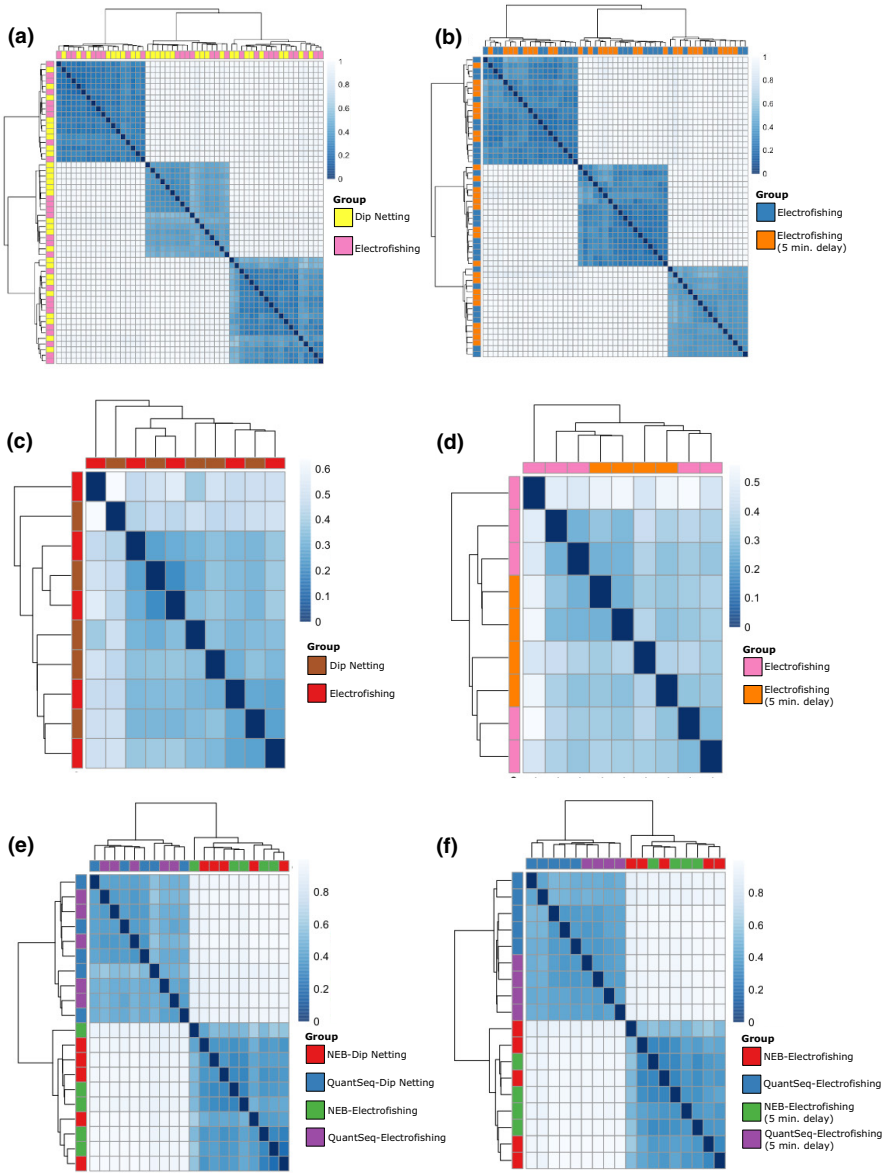

FIGURE 2 Sample-to-sample distance heatmaps for the comparison between different sampling techniques, different tissue harvesting time, and different mRNA-Seq library types. The rows and columns are arranged based on hierarchical clustering, so that samples with similar expression profiles are positioned near to each other. The colour scale represents the distance between samples. A value of distance 0 indicates that two samples have identical gene expression. The smaller the distance, the higher the correlation between two samples. Treatment groups are indicated in different colours next to each heatmap. (a) 3′ RNA-Seq (QuantSeq) dip netting versus electrofishing for all tissues combined. (b) QuantSeq electrofishing with immediate sampling versus electrofishing with delayed sampling for all tissues combined. (c) Whole mRNA-Seq (NEB) dip netting versus electrofishing only for blood samples. (d) NEB electrofishing with immediate sampling versus electrofishing with delayed sampling only for blood samples. (e) NEB versus QuantSeq comparisons for dip netting versus electrofishing only for blood samples. (f) NEB versus QuantSeq comparisons for electrofishing with immediate sampling versus electrofishing with delayed sampling only for blood samples.

types (treatment groups 1 vs. 2). When gene expression between dip netting and electrofishing was analysed separately for different tissue types, 0 out of 21,859 genes were differentially expressed in blood, 16 out of 20,465 genes (0.08%) in gills, and 151 out of 9201 genes (1.6%) in muscle. NEB data (available for blood samples only) confirmed QuantSeq data, with no genes showing differential expression between the two sampling methods (0 out of 28,506 genes) (Table S2 on Dryad).

3.3.2 | Immediate versus delayed postmortem tissue harvesting

We found no significant difference (adjusted p-value <.05) in gene expression between samples for which tissues were harvested immediately versus ~5 min after euthanasia (treatment groups 2 vs. 3). QuantSeq data identified only one out of 35,468 genes (0.003%) that was significantly differentially expressed between tissue harvesting times across all tissue types. Similar to what was observed between sampling techniques, when tissues were analysed separately for harvesting times, 0 out of 22,071 genes were differentially expressed in blood, 18 out of 15,310 genes (0.1%) in gills, and three out of 30,058 genes (0.01%) in muscle (Table S2 on Dryad). NEB data (available for blood samples only) was similar to QuantSeq data, with only 6 out of 23,258 genes (0.03%) showing differential expression between the two sampling methods (Table S2 on Dryad). While using QuantSeq did not detect any genes that were differently expressed in blood between treatment groups 2 and 3, NEB found six loci that were differentially expressed; the only gene that could be identified was cingulin 1 (CGNL1).

3.3.3 | QuantSeq versus whole mRNA-Seq

We compared gene expression of the same 14 blood samples (N_{tot} = 28) among RNA-Seq library types built using QuantSeq and

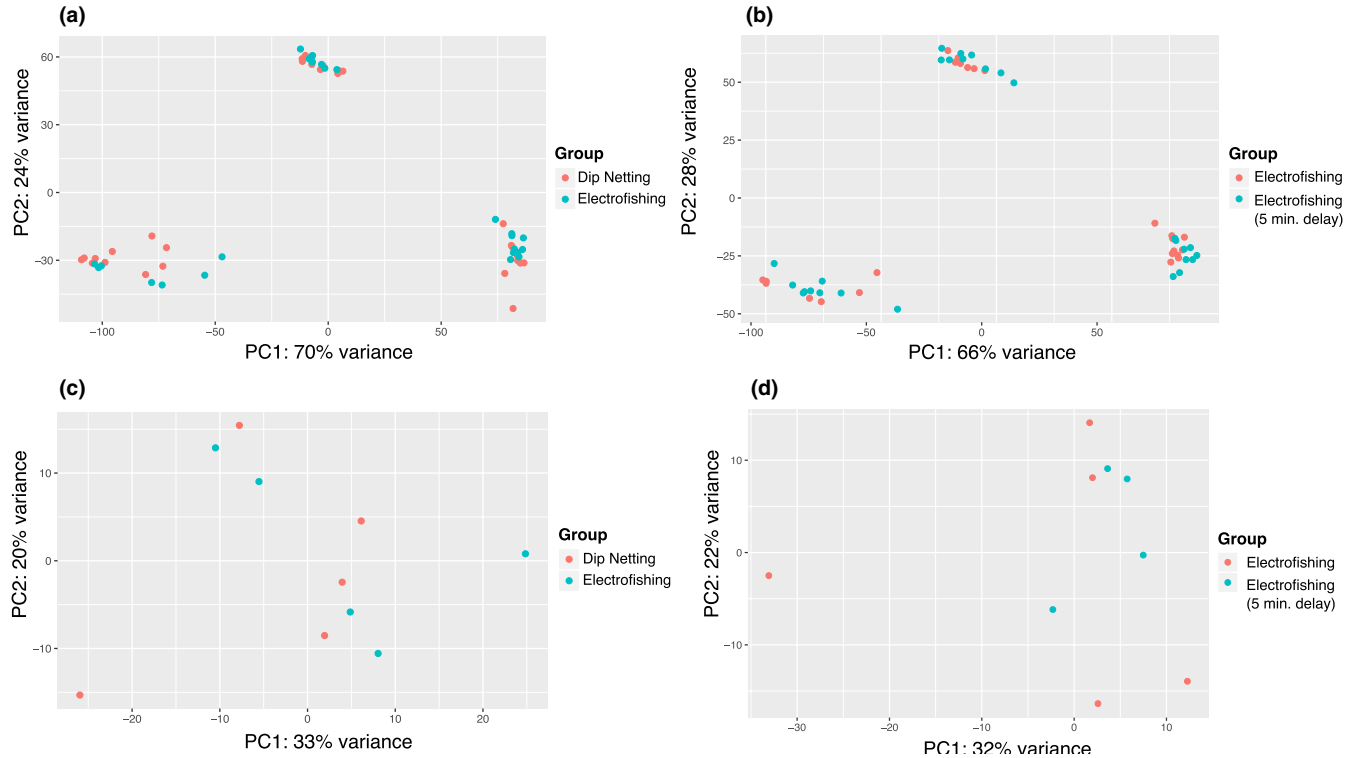

FIGURE 3 Principal component analysis (PCA) plots showing PC1 and PC2 for samples that are differentially expressed among sampling techniques, tissue harvesting time, and library preparation methods. Treatment groups compared are indicated by different coloured symbols next to each PCA plot. (a) 3′ RNA-Seq (QuantSeq) dip netting versus electrofishing for all tissues combined. (b) QuantSeq electrofishing with immediate sampling versus electrofishing with delayed sampling for all tissues combined. (c) NEB dip netting versus electrofishing only for blood samples. (d) Whole mRNA-Seq (NEB) electrofishing with immediate sampling versus electrofishing with delayed sampling only for blood samples.

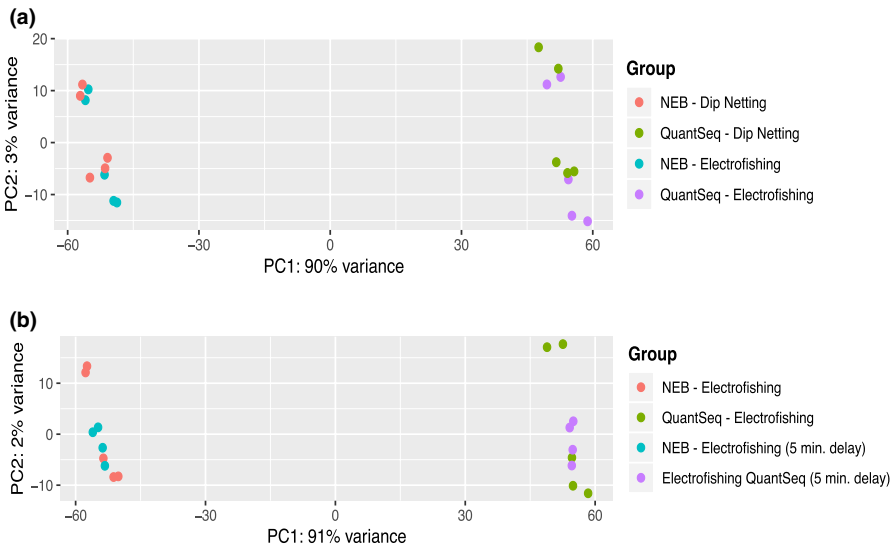

FIGURE 4 Principal component analysis (PCA) plots showing PC1 and PC2 for samples that are differentially expressed among sampling techniques, tissue harvesting time, and library preparation methods. Treatment groups compared are indicated by different coloured symbols next to each PCA plot. (a) Whole mRNA-Seq (NEB) versus 3′ RNA-Seq (QuantSeq) comparisons for dip netting versus electrofishing only for blood samples. (b) NEB versus QuantSeq comparisons for electrofishing with immediate sampling versus electrofishing with delayed sampling only for blood samples.

NEB. We found 13,597 out of 23,460 genes (58%) that were significantly (adjusted *p*-value <.05) differentially expressed between library types (Table S2 on Dryad). Specifically, we found 8170 transcripts with higher estimated expression in NEB than QuantSeq and 5427 transcripts with higher expression in QuantSeq versus NEB, which translates into 51% more genes with higher expression for NEB relative to QuantSeq. Within sampling methods, the percentage of differentially expressed genes between QuantSeq and NEB was 44%, 40%, and 42% for dip netting, electrofishing sampled immediately, and electrofishing sampled after 5 min, respectively. For each treatment group (groups 1–3), NEB showed a higher proportion of genes with significantly greater expression

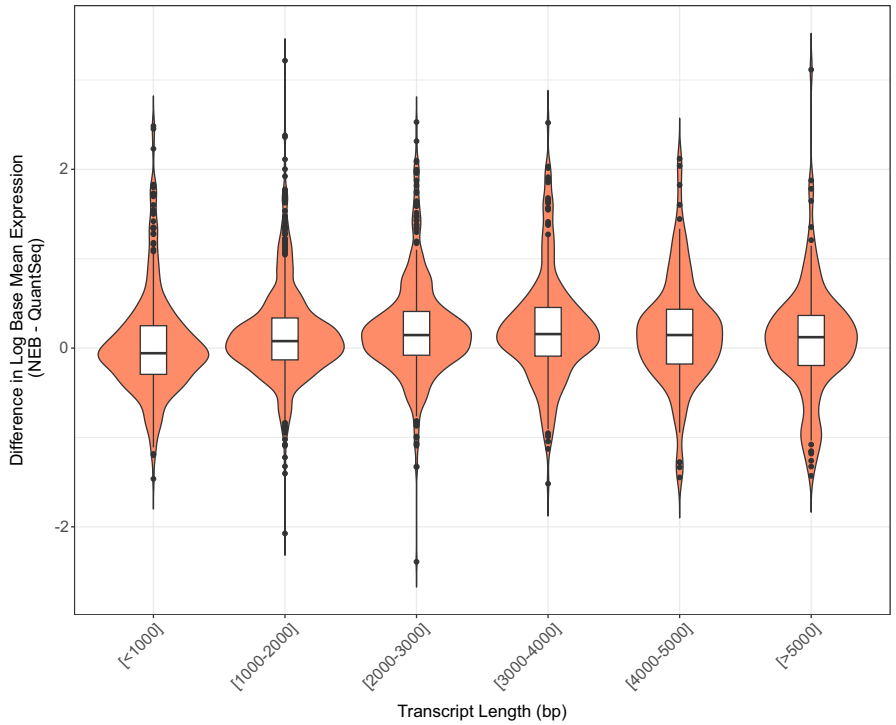

FIGURE 5 Violin and box plots comparing gene expression versus gene length for whole mRNA-Seq (NEB) and 3′ RNA-Seq (QuantSeq) library types independently of the treatment group. Each individual plot shows the difference in base mean expression for NEB versus QuantSeq, calculated as log baseMean NEB – log baseMean QuantSeq. Genes with equal expression fall on the zero line of the y-axis; genes with higher expression for NEB versus QuantSeq have positive numeric values above 0, while genes with higher expression for QuantSeq versus NEB have negative numeric values below 0. The plot only depicts genes that have been annotated with a known transcript length based on the *O. mykiss* genome. Transcript length is based on the longest transcript for a gene.

compared to QuantSeq (Table S2 on Dryad). However, the overall magnitude in log baseMean gene expression differences across genes with known transcript lengths is similar between the two methods (Figure 5).

4 | DISCUSSION

The increasing use of RNA-Seq for ecological, physiological, and evolutionary studies on wild caught organisms has required evaluating the influence of different sampling techniques, storage methods, processing times, and tissue types on RNA quality and data production (Camacho-Sanchez et al., 2013; Cheviron et al., 2011; Nakatsuji et al., 2019). Among the most important applications of RNA-Seq currently is testing for rapid adaptation to environmental change (e.g., Connon et al., 2018; Narum & Campbell, 2015) and for transgenerational inheritance (e.g., Charlesworth et al., 2017; Christie et al., 2016; Navarro-Martin et al., 2020; Skvortsova et al., 2018), and for addressing questions in evolutionary developmental biology (e.g., Liu et al., 2020; Roux et al., 2015).

In our work, we tested if different sampling techniques influenced gene expression in different tissues from westslope cutthroat trout. Overall, we obtained high RNA quality for all tissues (mean RIN ≥ 9.0 for the different tissues). We found no difference in RNA quality among samples obtained through dip netting or electrofishing, even when tissue was not harvested until 5 min after euthanasia. While opinions differ about a cutoff threshold RIN value to obtain reliable gene expression data, it has been shown that partially degraded RNA may still detect the same uniquely mapped genes as nondegraded RNA, although the coverage of mapped reads is lower for partially degraded RNA and is gene specific (Romero et al., 2014; Wang et al., 2016). However, while RNA degradation may not strongly affect mapping, it may drastically influence estimates of differential gene expression (Chen et al., 2014; Romero et al., 2014). Furthermore, different RNA-Seq techniques may be differentially affected by RNA degradation (Adiconis et al., 2013), requiring selecting the most appropriate RNA-Seq library depending on RNA quality (Adiconis et al., 2013).

We found that gene expression among individuals belonging to the same treatment group and tissue type were highly similar for the majority of comparisons (correlation coefficients ≥0.9), independent of the sampling method or harvesting time. However, we observed among-sample variation in gene expression, reflecting the importance of having larger sample sizes in RNA-Seq studies to decrease the influence of stochastic effects on variation in gene expression that could otherwise be interpreted as biologically relevant (Ching et al., 2014). Furthermore, we also observed similar expression levels among samples obtained with the two sampling methods, dip netting or electrofishing, or subjected to different tissue harvest times (immediate or 5 min after euthanasia). Sampling

individuals of the same age, in the same environment, and on the same day, with many biological replicates per treatment and using only samples with highly similar RNA quality most probably reduced the effects of nonbiological variation and of nonrelevant biological variation in our experiments (Fang & Cui, 2011; Wong et al., 2012; Yu et al., 2014).

We detected a 31% higher number of mapped genes for samples processed with NEB compared to QuantSeq, which supports previous research showing that whole mRNA libraries detect more genes than 3′ RNA-Seq (e.g., Crow et al., 2022; Jarvis et al., 2020; Ma et al., 2019; Xiong et al., 2017). Unlike previous studies (Jarvis et al., 2020), we did not find that transcript length affected the likelihood of detecting a given gene with either library type. We found significantly different gene expression between NEB and QuantSeq, with 51% more genes having greater expression for NEB relative to QuantSeq. Furthermore, although overall we did not find differences in gene expression among the treatment groups (sampling methods and processing time after death) with either of the two library types, there were no differentially expressed genes between processing the samples immediately or 5 min after death using QuantSeq, while we found six loci (0.03% of total loci) that were differentially expressed using NEB, suggesting that the two library types may produce different results (e.g., Crow et al., 2022; Jarvis et al., 2020; Ma et al., 2019; Tandonnet & Torres, 2017). Future research could further investigate in-depth comparisons of individual gene expression variation among library types, especially when observing differences in gene expression among experimental treatments (e.g., Ma et al., 2019).

Differential detection and expression of genes between NEB and QuantSeq library types has been proposed to depend on the length of the transcript and on the accuracy and completeness of the genome annotation. Since QuantSeq library data rely on mapping reads to the 3′ UTR of a gene for gene detection, and since UTR regions are generally more variable than protein coding regions and may be especially challenging to properly annotate, better annotated and complete genomes facilitate mapping and detection of transcripts/genes (Lawson et al., 2020). In this study, mapping was carried out on a closely related salmonid species (*O. mykiss*), since the genome of the westslope cutthroat trout is currently not available. This can explain why many more expressed genes were detected with NEB than QuantSeq. The higher number of detected expressed genes suggests that researchers should use whole mRNA-Seq for work on species with limited genomic resources (Crow et al., 2022). Furthermore, as QuantSeq libraries only allow amplification of the 3′ end of the transcript, different transcripts resulting from alternative cleavage sites and splicing would only be detected if the 3′ UTR differs. Traditional whole mRNA-Seq is therefore preferred when identifying distinct spliced transcripts is of interest.

One of the goals of this study was to test if different sampling methods or processing times affect gene expression. Although stress levels associated with dip netting and electrofishing may differ, we found that sampling technique did not affect gene expression. This result was independent of the RNA-Seq library type (QuantSeq or NEB) and tissue used. Although whole mRNA-Seq has been reported to be more sensitive to differentially expressed genes than 3′ RNA-Seq methods (Ma et al., 2019), independent of the RNA-Seq library used, we found no difference in estimated gene expression between the two field collection methods. As field conditions often change among sampling locations, researchers could opt to use electrofishing, where more efficient, and compare with fish obtained by netting in other localities without introducing extraneous variation in gene expression.

We also found that harvesting tissue immediately versus 5 min after euthanasia did not produce differences in gene expression, suggesting that it is safe to euthanize fish in batches and then proceed to tissue harvesting. In our study, the maximum processing time of the last tissue harvested after euthanasia was approximately 10 min (for fish processed starting 5 min after euthanasia). Although sampling techniques and tissue processing time did not influence variation in gene expression, we observed a large proportion of differentially expressed genes among different tissues.

In summary, our study indicates that differential gene expression results are likely to be comparable for dip netting and electrofishing. Additionally, gill, blood, and muscle all produce good quality RNA with reliable results if sampled within 5–10 min after euthanasia. Finally, although the NEB library detected more expressed genes, this did not lead to different results in terms of distinct gene expression among the treatments tested here. If detecting alternative splicing is not of interest for the study question and if working with an organism (or closely related species) with good genomic resources, QuantSeq is a reliable option for processing larger numbers of samples. Researchers can confirm results on a subset of samples using NEB, maximizing the amount of samples that can be studied while also reducing the cost of NEB library sequencing. Conversely, when it is crucial to detect as many genes as possible, when alternative splicing is of interest, or when working with an organism lacking good genomic resources, whole RNA-Seq is recommended.

AUTHOR CONTRIBUTIONS

Ylenia Chiari, Matthew C. Boyer, Marty Kardos, and Gordon Luikart designed the study. Leif Howard, Scott Relyea, James Dunnigan and Gordon Luikart performed the sampling. Analyses were performed by Scott Glaberman, Ylenia Chiari, and commissioned to Admera Health. Ylenia Chiari, Leif Howard, Nickolas Moreno, and Scott Glaberman wrote the manuscript. Ylenia Chiari and Gordon Luikart provided funding for this project. All authors have edited and approved the manuscript.

ACKNOWLEDGEMENTS

Matthew C. Boyer, Scott Relyea, and James Dunnigan were supported by Bonneville Power Administration grant no. 199101903 to Montana Fish, Wildlife & Parks. Ingo Braasch provided helpful comments regarding genome variability and genome annotation in fish. We are very thankful to the three anonymous reviewers who provided valuable comments and improved the quality of this manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw reads produced for this study have been uploaded to NCBI (SRA PRJNA691889). All the Supporting Information tables mentioned in this study have been deposited on Dryad and can be found here: https://doi.org/10.5061/dryad.ns1rn8ptb (Chiari et al., 2022). The codes used for the analyses described in this paper can be found on Zenodo https://zenodo.org/record/7439509#.Y-rzwnbMK3A.

ORCID

Ylenia Chiari https://orcid.org/0000-0003-2338-8602
Leif Howard https://orcid.org/0000-0001-8697-0582
Marty Kardos https://orcid.org/0000-0002-7587-3004
Scott Glaberman https://orcid.org/0000-0003-0594-4732

REFERENCES

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D. S., Busby, M. A., Berlin, A. M., Sivachenko, A., Thompson, D. A., Wysoker, A., Fennell, T., & Gnirke, A. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nature Methods*, *10*(7), 623–629.

Allendorf, F. W., & Leary, R. F. (1988). Conservation and distribution of genetic variation in a polytypic species, the cutthroat trout. *Conservation Biology*, *2*(2), 170–184.

Alvarez, M., Schrey, A. W., & Richards, C. L. (2015). Ten years of transcriptomics in wild populations: What have we learned about their ecology and evolution? *Molecular Ecology*, *24*(4), 710–725.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169.

Andrews, S. (2014). *FastQC A quality control tool for high throughput sequence data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Auer, P. L., & Doerge, R. W. (2010). Statistical design and analysis of RNA sequencing data. *Genetics*, *185*(2), 405–416.

Barton, B. A. (2002). Stress in fishes: A diversity of responses with particular reference to changes in circulating corticosteroids. *Integrative and Comparative Biology*, *42*(3), 517–525.

Barton, B. A., & Dwyer, W. P. (1997). Physiological stress effects of continuous-and pulsed-DC electroshock on juvenile bull trout. *Journal of Fish Biology*, *51*(5), 998–1008.

Barton, B. A., & Grosh, R. S. (1996). Effect of AC electroshock on blood features in juvenile rainbow trout. *Journal of Fish Biology*, *49*(6), 1330–1333.

Bi, R., & Liu, P. (2016). Sample size calculation while controlling false discovery rate for differential expression analysis with RNA-sequencing experiments. *BMC Bioinformatics*, *17*(1), 146.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bracewell, P., Cowx, I. G., & Uglow, R. F. (2004). Effects of handling and electrofishing on plasma glucose and whole blood lactate of *Leuciscus cephalus*. *Journal of Fish Biology*, *64*(1), 65–71.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., & Galtier, N. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Molecular Ecology Resources*, *12*(5), 834–845.

Camacho-Sanchez, M., Burraco, P., Gomez-Mestre, I., & Leonard, J. A. (2013). Preservation of RNA and DNA from mammal samples under field conditions. *Molecular Ecology Resources*, *13*(4), 663–673.

Charlesworth, D., Barton, N. H., & Charlesworth, B. (2017). The sources of adaptive variation. *Proceedings of the Royal Society B: Biological Sciences*, https://doi.org/10.1098/rspb.2016.2864

Chen, E. A., Souaiaia, T., Herstein, J. S., Evgrafov, O. V., Spitsyna, V. N., Rebolini, D. F., & Knowles, J. A. (2014). Effect of RNA integrity on uniquely mapped reads in RNA-Seq. *BMC Research Notes*, *7*(1), 753.

Cheviron, Z. A., Carling, M. D., & Brumfield, R. T. (2011). Effects of post-mortem interval and preservation method on RNA isolated from field-preserved avian tissues. *The Condor*, *113*(3), 483–489.

Chiari, Y., & Galtier, N. (2011). RNA extraction from sauropsids blood: Evaluation and improvement of methods. *Amphibia-Reptilia*, *32*(1), 136–139.

Chiari, Y., Howard, L., Moreno, N., Relyea, S., Dunnigan, J., Boyer, M. C., Kardos, M., Glaberman, S., & Luikart, G. (2022). *RNA-Seq reads and analysis outputs. NCBI (SRA PRJNA691889) and Dryad*.

Ching, T., Huang, S., & Garmire, L. X. (2014). Power analysis and sample size estimation for RNA-Seq differential expression. *RNA*, *20*(11), 1684–1696.

Christie, M. R., Marine, M. L., Fox, S. E., French, R. A., & Blouin, M. S. (2016). A single generation of domestication heritably alters the expression of hundreds of genes. *Nature Communications*, *7*, 10676.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1), 13.

Connon, R. E., Jeffries, K. M., Komoroske, L. M., Todgham, A. E., & Fangue, N. A. (2018). The utility of transcriptomics in fish conservation. *Journal of Experimental Biology*, *221*, jeb148833.

Corlett, R. T. (2017). A bigger toolbox: Biotechnology in biodiversity conservation. *Trends in Biotechnology*, *35*, 55–65.

Crow, T. M., Gill, J. A., Whitehead, A., & Runcie, D. E. (2022). 3′ RNA sequencing does not increase power or reduce costs for gene expression analysis. *bioRxiv*. https://doi.org/10.1101/2022.04.13.488225

Debey, S., Schoenbeck, U., Hellmich, M., Gathof, B. S., Pillai, R., Zander, T., & Schultze, J. L. (2004). Comparison of different isolation techniques prior gene expression profiling of blood derived cells: Impact on physiological responses, on overall expression and the role of different cell types. *The Pharmacogenomics Journal*, *4*(3), 193–207.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21.

Dobin, A., & Gingeras, T. R. (2015). Mapping RNA-Seq reads with STAR. *Current Protocols in Bioinformatics*, *51*, 11.14.1–11.14.19.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*(1), 1–15.

Fang, Z., & Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Briefings in Bioinformatics*, *12*(3), 280–287.

Gayral, P., Weinert, L., Chiari, Y., Tsagkogeorga, G., Ballenghien, M., & Galtier, N. (2011). Next-generation sequencing of transcriptomes: A guide to RNA isolation in nonmodel animals. *Molecular Ecology Resources*, *11*(4), 650–661.

Hansen, K. D., Wu, Z., Irizarry, R. A., & Leek, J. T. (2011). Sequencing technology does not eliminate biological variability. *Nature Biotechnology*, *29*(7), 572–573.

Huitink, J. M., Heimerikxs, M., Nieuwland, M., Loer, S. A., Brugman, W., Velds, A., Sie, D., & Kerkhoven, R. M. (2010). Volatile anesthetics modulate gene expression in breast and brain tumor cells. *Anesthesia & Analgesia*, *111*(6), 1411–1415.

Jarvis, S., Birsa, N., Secrier, M., Fratta, P., & Plagnol, V. (2020). A comparison of low read depth QuantSeq 3′ sequencing to total RNA-Seq in FUS mutant mice. *Frontiers in Genetics*, *11*, 562445.

Jeffries, M. K. S., Kiss, A. J., Smith, A. W., & Oris, J. T. (2014). A comparison of commercially-available automated and manual extraction kits for the isolation of total RNA from small tissue samples. *BMC Biotechnology*, *14*(1), 94.

Jeffries, M. K. S., Teffer, A., Michaleski, S., Bernier, N. J., Heath, D. D., & Miller, K. M. (2021). The use of non-lethal sampling for transcriptomics to assess the physiological status of wild fishes. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *256*, 110629.

Kaern, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, *6*(6), 451–464.

Kim, J. K., Kolodziejczyk, A. A., Ilicic, T., Teichmann, S. A., & Marioni, J. C. (2015). Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nature Communications*, *6*(1), 8687.

Koch, C. M., Chiu, S. F., Akbarpour, M., Bharat, A., Ridge, K. M., Bartom, E. T., & Winter, D. R. (2018). A beginner's guide to analysis of RNA sequencing data. *American Journal of Respiratory Cell and Molecular Biology*, *59*(2), 145–157.

Krasnov, A., Koskinen, H., Pehkonen, P., Rexroad, C. E., Afanasyev, S., & Mölsä, H. (2005). Gene expression in the brain and kidney of rainbow trout in response to handling stress. *BMC Genomics*, *6*(1), 3.

Krishnan, J., Persons, J. L., Peuß, R., Hassan, H., Kenzior, A., Xiong, S., Olsen, L., Maldonado, E., Kowalko, J. E., & Rohner, N. (2020). Comparative transcriptome analysis of wild and lab populations of *Astyanax mexicanus* uncovers differential effects of environment and morphotype on gene expression. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *334*, 530–539.

Lawson, N. D., Li, R., Shin, M., Grosse, A., Yukselen, O., Stone, O. A., Kucukural, A., & Zhu, L. (2020). An improved zebrafish transcriptome annotation for sensitive and comprehensive detection of cell type-specific genes. *eLife*, *9*, e55792.

Li, C. I., Su, P. F., & Shyr, Y. (2013). Sample size calculation based on exact test for assessing differential expression analysis in RNA-seq data. *BMC Bioinformatics*, *14*(1), 357.

Liu, J., Frochaux, M., Gardeux, V., Deplancke, B., & Robinson-Rechavi, M. (2020). Inter-embryo gene expression variability recapitulates the hourglass pattern of evo-devo. *BMC Biology*, *18*, 129.

Liu, Y., Zhou, J., & White, K. P. (2014). RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics*, *30*(3), 301–304.

Lohman, B. K., Weber, J. N., & Bolnick, D. I. (2016). Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Molecular Ecology Resources*, *16*, 1315–1321.

Long, Y., Yan, J., Song, G., Li, X., Li, X., Li, Q., & Cui, Z. (2015). Transcriptional events co-regulated by hypoxia and cold stresses in zebrafish larvae. *BMC Genomics*, *16*(1), 385.

Lopez-Patino, M. A., Hernandez-Perez, J., Gesto, M., Libran-Perez, M., Miguez, J. M., & Soengas, J. L. (2014). Short-term time course of liver metabolic response to acute handling stress in rainbow trout, *Oncorhynchus mykiss*. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *168*, 40–49.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

Ma, F., Fuqua, B. K., Hasin, Y., Yukhtman, C., Vulpe, C. D., Lusis, A. J., & Pellegrini, M. (2019). A comparison between whole transcript and 3′ RNA sequencing methods using Kapa and Lexogen library preparation methods. *BMC Genomics*, *20*(1), 9.

Mandelboum, S., Manber, Z., Elroy-Stein, O., & Elkon, R. (2019). Recurrent functional misinterpretation of RNA-seq data caused by sample-specific gene length bias. *PLoS Biology*, *17*(11), e3000481.

Meyer, E., Aglyamova, G. V., & Matz, M. V. (2011). Profiling gene expression responses of coral larvae (*Acropora millepora*) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Molecular Ecology*, *20*(17), 3599–3616.

Milla, S., Mathieu, C., Wang, N., Lambert, S., Nadzialek, S., Massart, S., Henrotte, E., Douxfils, J., Mélard, C., Mandiki, S. N. M., & Kestemont, P. (2010). Spleen immune status is affected after acute handling stress but not regulated by cortisol in Eurasian perch, Perca Fluviatilis. *Fish & Shellfish Immunology*, *28*(5–6), 931–941.

Milligan, C. L. (1996). Metabolic recovery from exhaustive exercise in rainbow trout. *Comparative Biochemistry and Physiology Part A: Physiology*, *113*(1), 51–60.

Morrissy, A. S., Morin, R. D., Delaney, A., Zeng, T., McDonald, H., Jones, S., Zhao, Y., Hirst, M., & Marra, M. A. (2009). Next-generation tag sequencing for cancer gene expression profiling. *Genome Research*, *19*(10), 1825–1835.

Mutch, D. M., Tordjman, J., Pelloux, V., Hanczar, B., Henegar, C., Poitou, C., Veyrie, N., Zucker, J. D., & Clement, K. (2008). Needle and surgical biopsy techniques differentially affect adipose tissue gene expression profiles. *The American Journal of Clinical Nutrition*, *89*(1), 51–57.

Mutter, G. L., Zahrieh, D., Liu, C., Neuberg, D., Finkelstein, D., Baker, H. E., & Warrington, J. A. (2004). Comparison of frozen and RNALater solid tissue storage methods for use in RNA expression microarrays. *BMC Genomics*, *5*(1), 88.

Nakatsuji, N., Adachi, K., & Morioka, K. (2019). Long-term stability of RNA isolated from muscle of red seabream (*Parus major*) during ice storage. *Fish Physiology and Biochemistry*, *45*(2), 819–828.

Narum, S. R., & Campbell, N. R. (2015). Transcriptomic response to heat stress among ecologically divergent populations of redband trout. *BMC Genomics*, *16*, 103.

Navarro-Martín, L., Martyniuk, C. J., & Mennigen, J. A. (2020). Comparative epigenetics in animal physiology: An emerging frontier. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *36*, 100745. https://doi.org/10.1016/j.cbd.2020.100745

Olsvik, P. A., Lie, K. K., & Hevrøy, E. M. (2007). Do anesthetics and sampling strategies affect transcription analysis of fish tissues? *BMC Molecular Biology*, *8*(1), 48.

Passow, C. N., Kono, T. J., Stahl, B. A., Jaggard, J. B., Keene, A. C., & McGaugh, S. E. (2019). Nonrandom RNAseq gene expression associated with RNAlater and flash freezing storage methods. *Molecular Ecology Resources*, *19*(2), 456–464.

Pérez-Portela, R., & Riesgo, A. (2013). Optimizing preservation protocols to extract high-quality RNA from different tissues of echinoderms for next-generation sequencing. *Molecular Ecology Resources*, *13*(5), 884–889.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9), 1650–1667.

Qian, X., Ba, Y., Zhuang, Q., & Zhong, G. (2014). RNA-Seq technology and its application in fish transcriptomics. *Omics: A Journal of Integrative Biology*, *18*(2), 98–110.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Reese, S. E., Archer, K. J., Therneau, T. M., Atkinson, E. J., Vachon, C. M., De Andrade, M., Kocher, J. P. A., & Eckel-Passow, J. E. (2013). A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. *Bioinformatics*, *29*(22), 2877–2883.

Rey, O., Eizaguirre, C., Angers, B., Baltazar-Soares, M., Sagonas, K., Prunier, J. G., & Blanchet, S. (2020). Linking epigenetics and biological conservation: Towards a conservation epigenetics perspective. *Functional Ecology*, *34*, 414–427.

Romero, I. G., Pai, A. A., Tung, J., & Gilad, Y. (2014). RNA-seq: Impact of RNA degradation on transcript quantification. *BMC Biology*, *12*(1), 42.

Romero, L. M., & Reed, J. M. (2005). Collecting baseline corticosterone samples in the field: Is under 3 min good enough? *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, *140*(1), 73–79.

Ross, J. (1995). mRNA stability in mammalian cells. *Microbiology and Molecular Biology Reviews*, *59*(3), 423–450.

Roux, J., Rosikiewicz, M., & Robinson-Rechavi, M. (2015). What to compare and how: Comparative transcriptomics for Evo-Devo. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *324B*, 372–382.

Ruane, N. M., Huisman, E. A., & Komen, J. (2001). Plasma cortisol and metabolite level profiles in two isogenic strains of common carp during confinement. *Journal of Fish Biology*, *59*(1), 1–12.

Shepard, B. B., Urie, W., & May, B. E. (2003). *Status of westslope cutthroat trout (Oncorhynchus clarkii lewisi) in the United States: 2002*. USDA Forest Service.

Skvortsova, K., Iovino, N., & Bogdanović, O. (2018). Functions and mechanisms of epigenetic inheritance in animals. *Nature Reviews Molecular Cell Biology*, *19*, 774–790. https://doi.org/10.1038/s41580-018-0074-2

Smith, S., Bernatchez, L., & Beheregaray, L. B. (2013). RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics*, *14*(1), 375.

Sopinka, N. M., Donaldson, M. R., O'Connor, C. M., Suski, C. D., & Cooke, S. J. (2016). Stress indicators in fish. In C. B. Scherck, L. Tort, A. P. Farrell, & C. J. Brauner (Eds.), *Fish physiology* (Vol. *35*, pp. 405–462). Academic Press.

Staton, J. M., Thomson, A. M., & Leedman, P. J. (2000). Hormonal regulation of mRNA stability and RNA-protein interactions in the pituitary. *Journal of Molecular Endocrinology*, *25*(1), 17–34.

Tandonnet, S., & Torres, T. T. (2017). Traditional versus 3′ RNA-seq in a non-model species. *Genomics Data*, *11*, 9–16.

Todd, E. V., Black, M. A., & Gemmell, N. J. (2016). The power and promise of RNA-seq in ecology and evolution. *Molecular Ecology*, *25*(6), 1224–1241.

Wang, L., Nie, J., Sicotte, H., Li, Y., Eckel-Passow, J. E., Dasari, S., Vedell, P. T., Barman, P., Wang, L., Weinshiboum, R., & Jen, J. (2016). Measure transcript integrity using RNA-seq data. *BMC Bioinformatics*, *17*(1), 58.

Wang, Y., Yang, L., Wu, B., Song, Z., & He, S. (2015). Transcriptome analysis of the plateau fish (*Triplophysa dalaica*): Implications for adaptation to hypoxia in fishes. *Gene*, *565*(2), 211–220.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63.

Wiseman, S., Osachoff, H., Bassett, E., Malhotra, J., Bruno, J., VanAggelen, G., Mommsen, T. P., & Vijayan, M. M. (2007). Gene expression pattern in the liver during recovery from an acute stressor in rainbow trout. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *2*(3), 234–244.

Wong, P. B., Wiley, E. O., Johnson, W. E., Ryder, O. A., O'Brien, S. J., Haussler, D., Koepfli, K. P., Houck, M. L., Perelman, P., Mastromonaco, G., & Bentley, A. C. (2012). Tissue sampling methods and standards for vertebrate genomics. *GigaScience*, *1*(1), 2047–217X.

Wood, C. M., Turner, J. D., & Graham, M. S. (1983). Why do fish die after severe exercise? *Journal of Fish Biology*, *22*(2), 189–201.

Wu, Z. J., Meyer, C. A., Choudhury, S., Shipitsin, M., Maruyama, R., Bessarabova, M., Nikolskaya, T., Sukumar, S., Schwartzman, A., Liu, J. S., & Polyak, K. (2010). Gene expression profiling of human breast tissue samples using SAGE-Seq. *Genome Research*, *20*(12), 1730–1739.

Xiong, Y., Soumillon, M., Wu, J., Hansen, J., Hu, B., Van Hasselt, J. G., Jayaraman, G., Lim, R., Bouhaddou, M., Ornelas, L., Bochicchio, J., Lenaeus, L., Stocksdale, J., Shim, J., Gomez, E., Sareen, D., Svendsen, C., Thompson, L. M., Mahajan, M., … Birtwistle, M. R. (2017). A comparison of mRNA sequencing with random primed and 3-directed libraries. *Scientific Reports*, *7*, 14626.

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Marugán, J. C., Cummins, C., Davidson, C., Dodiya, K., Fatima, R., Gall, A., Giron, C. G., … Flicek, P. (2020). Ensembl 2020. *Nucleic Acids Research*, *48*(D1), D682–D688. https://doi.org/10.1093/nar/gkz966

Yu, Y., Fuscoe, J. C., Zhao, C., Guo, C., Jia, M., Qing, T., Bannon, D. I., Lancashire, L., Bao, W., Du, T., & Luo, H. (2014). A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nature Communications*, *5*(1), 1–11.

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: Chiari, Y., Howard, L., Moreno, N., Relyea, S., Dunnigan, J., Boyer, M C., Kardos, M., Glaberman, S., & Luikart, G. (2023). Influence of RNA-Seq library construction, sampling methods, and tissue harvesting time on gene expression estimation. *Molecular Ecology Resources*, *23*, 803–817. https://doi.org/10.1111/1755-0998.13757