



Establishing Nanopore Long-Read Sequencing as a Cost-Effective Approach for Axolotl Regeneration Research

Citation

Singer, Hani Dani. 2025. Establishing Nanopore Long-Read Sequencing as a Cost-Effective Approach for Axolotl Regeneration Research. Master's thesis, Harvard University Division of Continuing Education.

Link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37380137>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

Establishing Nanopore Long-Read Sequencing as a Cost-Effective Approach for Axolotl
Regeneration Research

Hani D. Singer

A Thesis in the Field of Biotechnology
for the Degree of Master of Liberal Arts in Extension Studies

Harvard University

February 2025

Abstract

RNA sequencing (RNA-seq) has become a cornerstone methodology to decipher the regulation of intricate molecular pathways. The development of RNA-seq technology in the last decade has exponentially increased the scope in which researchers collect data on gene expression, yet the high cost of executing modern sequencing methodologies can be a prohibitive barrier. Oxford Nanopore Technology (ONT) has rapidly emerged as a transformative force in the field of RNA-seq by enabling low-cost long-read sequencing.

The goal of this study was to investigate whether nanopore long-read sequencing can serve as a cost-effective and viable alternative to conventional RNA-seq technologies for investigating transcriptomic expression patterns in axolotl salamanders (*Ambystoma mexicanum*). The first aim of this study was to optimize the MinION long-read RNA-seq protocol for the axolotl model from RNA isolation to differential gene expression (DGE) analysis. The second aim was to utilize the optimized protocol within various experimental research contexts in axolotl regeneration research. The third aim was to benchmark the more robust PromethION sequencer utilizing the optimized workflow from aim one, to assess its utility for experimental applications relative to the MinION.

Our study presents a novel and optimized workflow in which nanopore sequencing accurately captures transcriptomic expression patterns in the axolotl, demonstrating its potential as a viable, cost-effective alternative to traditional RNA-seq platforms in regeneration research. This study marks the first successful published execution of nanopore RNA-seq in the axolotl salamander.

Author's Biographical Sketch

Hani D. Singer is a scientific researcher with over five years of experience in the field of regenerative biology. Starting as a research assistant, he is currently the research laboratory manager in the Whited Laboratory at Harvard University in the Department of Stem Cell and Regenerative Biology. He holds a B.A. in Biochemistry from the University of Massachusetts Boston.

Hani believes in the power of collaboration and has been deeply involved in the project management infrastructure of academic research labs. His passion for both micro- and macro- level project design and execution led him to explore cost-effective alternative approaches to RNA sequencing. To better execute computational work in the Whited Lab, he pursued a graduate certificate in Bioinformatics from the Harvard Extension School (HES), ultimately focusing on establishing a workflow utilizing nanopore sequencing technology in axolotl regeneration research for an ALM in Biotechnology at HES.

In his free time, Hani enjoys playing piano and producing music, caring for his ever-growing plant collection, and playing tennis. His future goals include continuing to engage in effective project design and problem-solving to help organizations address complex challenges, while continuing to learn and grow his technical skillsets within molecular biology research.

Dedication

This work is dedicated to everybody who has supported me through all the ups and downs as I have navigated my path in life.

First and foremost, I dedicate this work to my parents, whose unwavering encouragement and unconditional support have been the foundation of all my successes, and whose love and belief in me have helped me rise from every failure. I owe my work ethic and pride in creating a body of work to the values my parents have instilled in me.

I also dedicate this work to my friends, past and present. I am forever grateful for those who have selflessly given me their time, energy, and care. Your companionship has been invaluable, and I would not be where I am today without all your support. To all my friends who now know more about RNA sequencing than they ever wanted to – thank you for enduring this journey with me.

Lastly to my cat Tsugi, who ran to the door to greet me after every late night spent in the laboratory and kept my lap warm while writing this work. Even during the toughest times, you reminded me to find joy in the little things. This work would not have been the same without you by my side.

Acknowledgments

This work would not have been possible without the support and guidance of my mentor and supervisor Dr. Jessica Whited, Assistant Professor at the Harvard Department of Stem Cell and Regenerative Biology. Her mentorship in my scientific research cannot be overstated. Her unwavering encouragement and trust in my ability to independently design and execute my ideas allowed me to grow immensely as a researcher and manager, for which I am forever grateful.

I would also like to thank my colleague and close friend, Adnan Abouelela, now Ph.D. student at ETH Zürich, for being an instrumental collaborator on this project. His expertise as a computational biologist and mentorship have been invaluable in developing my bioinformatics skills. Simply put, this work would not have been possible without him, and I am grateful we had the opportunity to work on this project together.

Similarly, my gratitude also goes to Connor Powell, currently a Ph.D. student at Northeastern University. Starting as an intern on this project, his dedication and insightful contribution far exceeded any of our expectations. Witnessing his growth as a scientist was one of the most rewarding aspects of developing this work.

Finally, this project included contributions from current and former Whited lab members Dr. Aaron Savage, Noah Lopez, Dr. Duygu Payzin, Dr. Benjamin Tajer, Dr. Sangwon Min, as well as current and former undergraduate students Shifa Hossain, Ryan Kim, Miranda Thai, Ashley Juarez, Anna Luong, and Rachna Chilambi. Thank you all for your invaluable contributions and for being a part of my journey.

Table of Contents

Abstract	iii
Author's Biographical Sketch	iv
Dedication	v
Acknowledgments	vi
List of Figures	x
Chapter I. Introduction	1
Moore's Law & Next-Generation Sequencing	2
Nanopore Long-Read Sequencing	4
ONT Library Preparation and Sequencing Platforms	5
Bioinformatic Analysis of Bulk RNA-seq Sequencing Data	8
Axolotl Regeneration Research	10
Chapter II. Materials and Methods	12
Axolotl Amputation Surgery	12
Axolotl Pancreas Resection Surgery	13
Axolotl AV951 Treatment Surgery	14
Total RNA Isolation	14
In Situ Hybridization Chain Reaction (HCR-FISH)	15
ONT Library Preparation and Sequencing	16
Illumina Sequencing and Analysis	17
<i>De Novo</i> ONT Sequencing Analysis	18

EPI2ME Labs ONT Sequencing Analysis.....	19
PCA and Z-Score Correlation Analysis.....	20
Chapter III. Results	21
Total RNA Isolation.....	21
Run 001 ONT Library Preparation	23
Run 001 Sequencing Performance Analysis.....	25
Run 010 PCR Optimizations.....	26
Run 010 Sequencing Length Distributions	27
Run 011 Sequencing Performance Analysis.....	28
Run 011 Differential Gene Expression Analysis	29
Effect of Optimizations on DEG Discovery	31
Run 019 Pancreas Resection Experiment	32
Run 019 Sequencing Results	32
Run 019 Validation Results	33
Run 018 VEGF Pathway Inhibition Experiment	35
Run 018 Results and Validations.....	36
Run 021 PromethION Platform	38
Sequencing Run 011 vs. 021.....	39
Run 024 PromethION Pancreas Resection	40
Sequencing Run 019 vs. 024.....	41
Chapter IV. Discussion	43
Optimizations of Nanopore Workflow	43
Nanopore Bioinformatics and Axolotl Research	44

Testing MinION Workflow in Axolotl Research	45
PromethION Sequencing in Axolotl	48
Chapter V. Conclusion.....	51
Appendix 1. Overview of Oxford Nanopore Technology Hardware	53
Appendix 2. Axolotl Experimental Surgery Model Illustrations.....	54
Appendix 3. Tapestation RNA & DNA Distribution Plots.....	56
Appendix 4. Z-Score Correlation Analysis Plots.....	60
Appendix 5. Run 011 Sequencing Performance Analysis	62
Appendix 6. MinION vs. PromethION Gene Discovery Plots.....	65
Works Cited	67

List of Figures

Figure 1.	Cost per Raw Megabase of DNA Sequence, 2001-2022	3
Figure 2.	Visualization of Nanopore Sequencing and Basecalling	5
Figure 3.	ONT PCR-cDNA Kit Workflow	8
Figure 4.	<i>De Novo</i> Nanopore Sequencing Analysis Pipeline	19
Figure 5.	Total RNA Isolation Length Distribution, Post-Optimization.....	23
Figure 6.	Run 001 Library Prepared cDNA and Sequencing Length Distributions.	24
Figure 7.	Run 001 Read Classification and Alignment Efficiency	26
Figure 8.	Run 010 Sequencing Read Length Distribution	28
Figure 9.	Run 011 Sequencing Distribution and Read Classification.....	29
Figure 10.	Run 011 Principal Component Analysis.....	30
Figure 11.	Run 001 vs. 011, Unique Genes Discovered	31
Figure 12.	Run 019 DGE Heatmap and PCA Plot	33
Figure 13.	HCR-FISH Staining of Sham and Resected Pancreas	34
Figure 14.	Run 018 Read Classification and Alignment Efficiency	36
Figure 15.	Run 018 PCA plot and HCR-FISH Validations	37
Figure 16.	Run 021 Read Classification and Alignment Efficiency	38
Figure 17.	Run 021 Unique Genes Discovered and Correlation of Z-Scores	39
Figure 18.	Run 024 Read Classification and Alignment Efficiency	41
Figure 19.	Run 024 Correlation of Z-Scores.....	42
Figure 20.	ONT Flow Cells and Sequencing Platforms.....	53
Figure 21.	Axolotl Limb Amputation Experiment Illustrations.....	54

Figure 22.	Axolotl Pancreas Resection Experiment Illustrations.....	55
Figure 23.	Axolotl VEGF inhibitor (AV951) Experiment Illustrations	55
Figure 24.	Total RNA Isolation Length Distribution, Pre-Optimization	56
Figure 25.	Run 001 Library Prepared cDNA Length Distribution.....	57
Figure 26.	Tapestation cDNA Length Distributions for 15-30 PCR Cycles.....	58
Figure 27.	Run 019 Tapestation cDNA Length Distribution	59
Figure 28.	Run 018 Z-Score Correlation Analysis, ONT-seq vs. Illumina-seq	60
Figure 29.	Run 021 Z-Score Correlation Analysis, EPI2ME vs. <i>De Novo</i> Pipelines	61
Figure 30.	Run 011 Sequencing Length Distribution.....	62
Figure 31.	Run 011 Heatmap of Differentially Expressed Genes	63
Figure 32.	Run 011 Volcano Plot of Differentially Expressed Genes	64
Figure 33.	Run 011 vs. 021 Comparison of Gene Discovery.....	65
Figure 34.	Run 019 vs. 024 Comprison of Gene Discovery	66

Chapter I.

Introduction

Watson and Crick's discovery of the three-dimensional structure of DNA in 1953 initiated a race to elucidate the nucleic acid residue sequences in biological samples (Heather & Chain, 2016; Watson & Crick, 1953). The fundamental logic being, if one can uncover the specific sequence of the nucleic acid residues, deduction and analysis of the macro regulation of genetic, transcriptomic and proteomic signaling pathways could be performed.

Innovation in the budding field of nucleic acid sequencing was limited at first, yet exponential development occurred in the turn of the new millennium. This progress is best reflected by the cost of sequencing the whole human genome - from \$3 billion during the Human Genome Project in 2003 to \$25 million in 2006, to below \$1,000 by 2016 (Schloss, 2008; Wetterstrand, 2024). Significant milestones in these technological advancements have been categorized into discrete generations of sequencing technologies, starting with Frederick Sanger's chain-termination technique in 1977, recognized as the first major breakthrough and thus referred to as first-generation sequencing (FGS) (Adewale, 2020). This technique involves the use of radiolabeled dideoxynucleotides (ddNTPs) that result in all possible lengths of a given template DNA sequence being run via polyacrylamide gel electrophoresis utilizing autoradiography to identify the nucleic residue sequence (Sanger et al., 1977). Over four decades later, Sanger sequencing is still considered a gold standard in accuracy and read quality,

although its inherent low-throughput design makes it inefficient for large scale experiments (Totomoch-Serra et al., 2017).

Moore's Law & Next-Generation Sequencing

Moore's Law is a predictive model of the evolution of computational development that suggests that compute power, quantified by number of microprocessors, can be expected to double approximately every two years (Lundstrom & Alam, 2022; Moore, 1965). As high-throughput genomic sequencing requires significant computational effort, Moore's Law has been an accurate predictor of sequencing cost through the completion of the first drafts of the human genome in the turn of the millennium. However, since 2007, significant improvements in sequencing software technology design and likewise the capacity for sequencing technology platforms to generate large amounts of data, have enabled the cost of sequencing to far outpace the expected trend line defined by Moore's Law (Fig. 1; (McCombie & McPherson, 2019; Wetterstrand, 2024).

Named Next-Generation Sequencing (NGS), or Second-Generation Sequencing (SGS), short-read sequencing approaches utilizing sequencing by synthesis (SBS) technology became commercially available in the late 1990s and continue to represent the majority of the current market share for sequencing technology platforms. SBS-NGS is dominated by Illumina, Inc. based out of San Diego, California (Pervez et al., 2022). Illumina SBS-NGS involves the library preparation technique of fragmenting the DNA into small pieces (referred to as "short reads"), ligating adapters, and amplifying them to form clusters. Fluorescently labeled nucleotides are then added one by one, at which point the fluorescent signals are detected as each nucleotide is incorporated into the

template DNA strand (Goodwin et al., 2016; Pareek et al., 2011). Illumina-SBS is cited as one of the main technological breakthroughs that, in tandem with the well-funded Human Genome Project and efforts by the National Human Genome Research Institute (NHGRI) to further improve sequencing efficiency, catalyzed the significant reduction in cost for commercial offerings of nucleotide sequencing (Goodwin et al., 2016; Schloss, 2008).

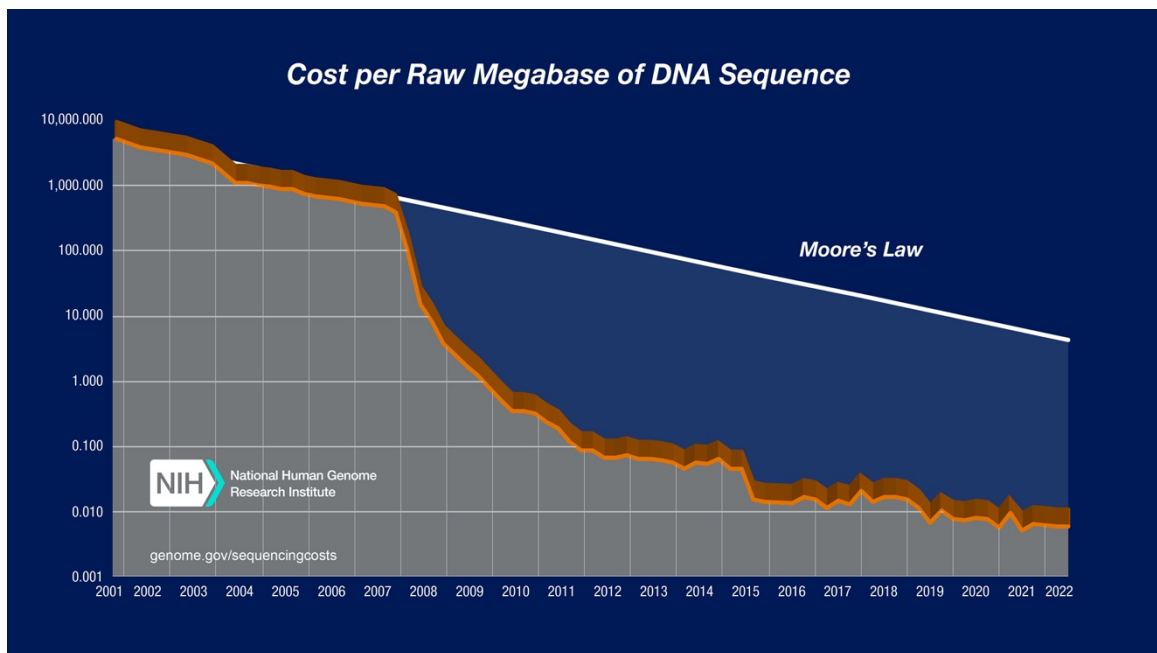


Figure 1. Cost per Raw Megabase of DNA Sequence, 2001-2022

Semi-log plot illustrating the dramatic reduction in sequencing cost since 2001. The y-axis is presented in log scale to emphasize the sharp decline after 2007, spurred by advancements in NGS platforms like Illumina. The large drop in cost per megabase far outpaced the expected reduction as defined by Moore's Law, which largely has been a reliable predictor of computational development. Plot sourced from the National Human Genome Research Institute (Wetterstrand, 2024).

Nanopore Long-Read Sequencing

The core principles of sequencing full-length DNA molecules were independently proposed by two groups in the late 1980's - David Deamer at University of California-Santa Cruz and George Church at Harvard University (Branton et al., 2008; Kasianowicz et al., 1996). By embedding nanopores – nanometer-scale channel proteins – in a membrane, a steady ionic current can be established to flow across, enabling the identification of certain molecules passing through the pore and disrupting of the current. Therefore, if the single-stranded-DNA (ssDNA) or RNA can be ensured to translocate through the pore in strict sequential order, it was proposed that the full sequence of nucleotides could be deduced through the modulation of current flowing across the membrane (Fig. 2). Significant advances have been made in the nanopore core workflow since its inception, including optimization of the stochastic sensing required to measure the nucleotide identities from the change in current (Gu et al., 1999), bioengineering improvements of the nanopore proteins themselves to improve basecalling speed and accuracy (Wang et al., 2021), and software development of the basecaller (Wick et al., 2019).

Oxford Nanopore Technologies (ONT) based out of Oxford, United Kingdom, was established in 2014 and introduced the first commercial nanopore flow cells to market, kickstarting the era of Third-Generation Sequencing (TGS) or long-read sequencing (Jain et al., 2016; Van Dijk et al., 2018). As the name suggests, long-read sequencing enables users to maintain the full length of the input nucleic acid material without having the fragmentation step present in NGS methods.

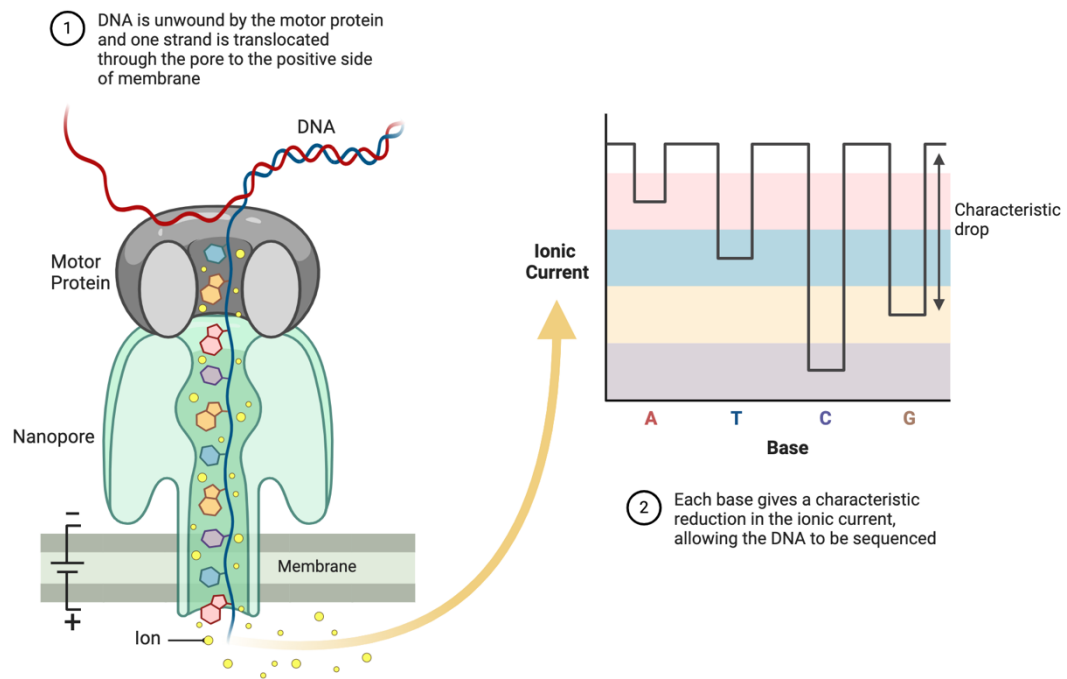


Figure 2. Visualization of Nanopore Sequencing and Basecalling

Illustrations depicting a cDNA strand being translocated via motor protein through a nanometer scale channel protein embedded in a membrane (left). Ionic current data is collected and modulations in current can be basecalled to identify the nucleic acid sequence (right). Figure sourced and edited via BioRender, created originally by user Flo Glencross.

ONT Library Preparation and Sequencing Platforms

After the first nanopore flow cell and sequencer came to market in 2014, ONT has expanded their catalogue to cover a multitude of sequencing needs (Fig. 20). ONT offers single flow-cell sequencers like the Flongle device, (126 active nanopores), MinION (512 active nanopores), and P2 Solo (3,000 active nanopores). ONT also offers high-throughput multi flow-cell systems such as the GridION (5 flow cells x 512 active nanopores) and PromethION (48 flow cells x 3,000 active nanopores), designed for large-

scale projects. For library preparation, ONT offers kits that allow flexibility of input material, including genomic DNA, reverse transcribed cDNA with RNA as input, or directly sequencing RNA. The development of the nanopores and library preparation kit chemistry has increased the sequencing depth capacity significantly in recent years, increasing from a few gigabases (Gb) of data output to up to 7 terabases (Tb) per run as of 2023, with accuracy improved from 60% in when first introduced to now 98% and above (Cuber et al., 2023; Rang et al., 2018).

In this study, we focus on the barcoded variants of the PCR-cDNA library preparation kits utilizing V11 and V14 chemistry (SQK-PCB111.24 and SQK-PCB114.24, respectively). Using total RNA as input, a reverse transcription (RT) step utilizing a poly-T overhang primer is performed to selectively transcribe the polyadenylated mRNA (mRNA) into double-stranded cDNA (Fig. 3). The cDNA is then PCR amplified along with barcoded primers per sample, then pooled together equimolarly prior to ligation of the motor protein to facilitate sequencing. The pooled cDNA is loaded onto a flow cell at the femtomole (fmol) scale, adjusted relative to total pore capacity of the flow cell.

The Nanopore platform allows for sequencing to be observed in real time for length distribution and quality. If any issues are observed, the sequencing run can be paused, and a nuclease treatment wash step can be performed to remove the initial library and re-loaded with a new one. The raw ionic current data is logged for each read onto a hard drive in pod5 format for basecalling and further downstream analysis. Optionally, reads can be basecalled, filtered, demultiplexed and aligned to a reference in real time

during sequencing, although a powerful GPU is necessary to ensure sequencing is not disrupted.

The cost of the MinION sequencer is \$2000 for a starter pack which includes the sequencer itself, a library preparation kit capable of more than 10 multiplexed sequencing runs, and two flow cells. Every subsequent flow cell is between \$600 - \$900, dependent on whether they are purchased in bulk. After the initial purchase (or access to a sequencer) an average MinION sequencing run with in-house library preparation comes out to average roughly \$700 per run.

Assuming access to the more robust PromethION platform, the initial cost is \$3600 for a pack of four flow cells. Thus, a sequencing run on the PromethION platform averages to \$1000 per run, assuming in-house library prep. A serious consideration of the PromethION four-unit system is that flow cells have a suggested shelf life of 12 weeks due to degradation of the nanopores in storage. Through this study we have observed up to a 40% decrease in pore capacity even after just 10 weeks when stored at the recommended 4°C.

In comparison, a standard multiplexed Illumina run is quoted to be around \$3000 – 5000, including library preparation through a third party. Unlike nanopore sequencers, Illumina does not offer real time analysis tools, nor the ability to wash and re-load libraries if issues are encountered during sequencing.

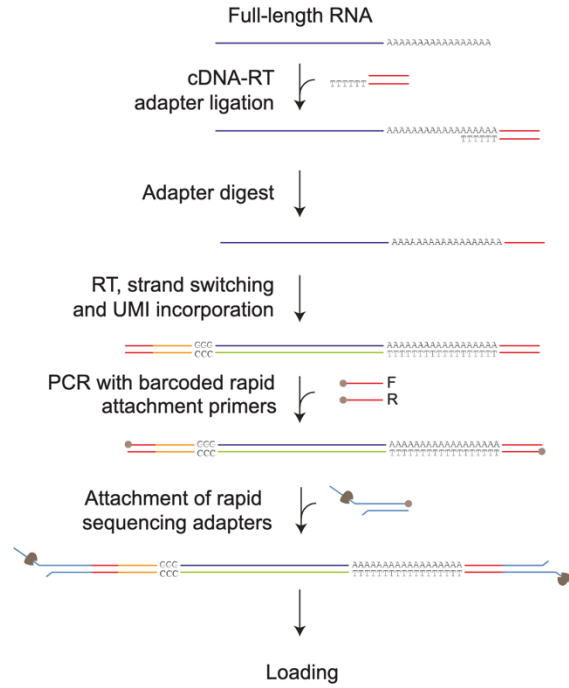


Figure 3. ONT PCR-cDNA Kit Workflow

Workflow schematic of cDNA library preparation using the ONT PCR-cDNA kits (PCB-111.24, PCB-114.24) utilized in this study. Input RNA is reverse transcribed (RT) into cDNA, then unique molecular identifiers (UMIs) are ligated prior to a polymerase chain reaction (PCR) step to ligate adapters onto the double stranded cDNA (dscDNA). Schematic sourced from Oxford Nanopore Technologies.

Bioinformatic Analysis of Bulk RNA-seq Sequencing Data

Regardless of sequencing technology, raw sequencing data (e.g., fluorescent intensity in Illumina-seq, current modulation data in Nanopore-seq) must be converted into nucleotide residue sequences in a process referred to as basecalling. Basecalled reads are processed by filtering for low quality reads, identifying barcode sequences and trimming all adapters and primers.

After some additional preprocessing and quality control, reads are then aligned to a reference genome or transcriptome using alignment tools like Minimap2, which uses an algorithm to map sequences based on seed-and-extend methods (Li, 2018). Minimap2 first identifies candidate regions in the reference using k-mer seeds to create an index file, performing chaining and base-level alignment to refine the mapping to output aligned reads in BAM file format. This approach allows the software to handle both short and long reads, making it a versatile tool for genomic and transcriptomic analyses (Sadasivan et al., 2023).

To calculate for transcript abundances from the BAM files produced by Minimap2, quantification tools like Salmon utilize a lightweight mapping algorithm to process the alignments. (Patro et al., 2017). These quantifications are not raw counts of how many reads were attributed to each transcript, but rather normalized estimates that account for factors like fragment length distribution, GC-content biases, and mapping uncertainty to generate more accurate count matrices for downstream sequencing analysis.

Differential expression analysis tools like edgeR, Limma-voom, and DESeq2 utilize count matrices as input to calculate statistical significance of transcript and gene expression differences across two experimental conditions (Law et al., 2014; Love et al., 2014; Robinson et al., 2010). DESeq2, utilized in this study, uses a negative binomial generalized linear model to provide an estimate of log fold changes and adjusted p-values to identify differentially expressed genes across conditions. Visualization tools like volcano plots and heatmaps help assess key genes of interest for the adjusted p-values

and log fold changes in each dataset and are typical methodologies for RNA-seq data representation (Rutter et al., 2019).

Additionally, principal component analysis (PCA) utilizes agnostic dimensional reduction techniques to confirm whether differences in transcriptomic expression patterns reflect the experimental conditions. By investigating the groups of transcripts of the principal component dimensions separating the samples, researchers can gain further insight into the factors driving change in transcript expression patterns across experimental conditions.

Axolotl Regeneration Research

The axolotl (*Ambystoma mexicanum*) is a species of aquatic salamander that possess a unique ability to completely regenerate their limbs and internal organs like its heart, brain and lungs (Cano-Martínez et al., 2010; Jensen et al., 2021; Vieira et al., 2020; Wei et al., 2022). The axolotl regenerates its limb through the development of a blastema, a mass of undifferentiated cells that forms at the site of amputation (Haas & Whited, 2017). The blastema cells then proliferate, differentiating into the various tissues required to reconstruct the lost limb including muscle, cartilage, nerves, and blood vessels (Tanaka, 2016). Understanding the molecular mechanisms underlying the axolotl's regenerative process has become a major focus in therapeutic research due to its potential applications in regenerative medicine (Fior, 2014; Roy & Gatien, 2008).

Genome and transcriptome resources have only recently been published, opening the door for further large-scale genetic analyses in axolotl (Bryant et al., 2017; Nowoshilow et al., 2018). The available genome poses a significant challenge for any analysis strategy due to the axolotl reference's giant 32Gb size (Schloissnig et al., 2021).

In addition, the current annotation of the references leaves much to be desired, for example with many genes identified to be enriched in the regenerating blastema to remain unannotated (Bryant et al., 2017).

The size and complexity of the axolotl genome feed into the potential of long-read sequencing in identifying isoforms to help improve the assembly of complex and repetitive regions. Nanopore sequencing offers distinct advantages in this context, as it allows for the generation of ultra-long reads that can span repetitive elements and capture full-length transcripts (Jain et al., 2016). This capability is crucial for annotating previously uncharacterized genes and identifying novel isoforms, particularly in the context of the axolotl's large and complex genome (Liem et al., 2017; Lopatriello et al., 2023; Sagniez et al., 2024). With over 65% of the axolotl genome consisting of repetitive sequences - significantly higher than the 50% found in the human genome – this further highlights the need for sequencing methodologies capable of addressing these complexities (Lander et al., 2001; Nowoshilow et al., 2018).

As of the time of writing, no peer-reviewed studies have been published utilizing nanopore sequencing on the axolotl salamander. Therefore, nanopore sequencing is a promising tool for improving our understanding of the molecular mechanisms involved in axolotl tissue repair and regeneration. We propose a novel workflow that allows researchers to unlock the nanopore sequencing platform for use in the axolotl salamander. This framework will allow future research to leverage nanopore-based genome and transcriptome assemblies, single-cell RNA sequencing, and potentially single-cell proteomics to gain deeper insights into axolotl regeneration (Fang et al., 2021; Motone & Nivala, 2023; Philpott et al., 2021).

Chapter II.

Materials and Methods

The following materials and methods were used in this project, and enough detail is provided to recreate the experiments and replicate the results. Additional data and instruction are available upon request.

Axolotl Amputation Surgery

All animal experimentation was approved and conducted in accordance with Harvard University's Institutional Animal Care and Use Committee (Protocol # 19-02-346). Age and size-matched adult leucistic axolotls were housed in axolotl housing water (40% holtfreter's solution with a salinity of 3000 μ S and 7.6 pH) and sroom temperature of 19°C. Animals were anesthetized in 0.1% Tricaine (MS-222) solution in axolotl housing water and confirmed after approximately 30 minutes to be fully anesthetized via a foot-pinch test. After transferring the animals to a damp paper towel soaked in tricaine solution, sterilized scissors were used to amputate the stylopod of the right forelimb. Any bone extending past the plane of amputation was trimmed back to ensure the wound epidermis would form uninhibited. Each collected limb tissue was immediately placed in 500 μ L of Trizol (Invitrogen, 15596018) in a microcentrifuge tube. After 5-10 minutes of waiting to ensure full absorption of Trizol into the tissue, the samples were frozen via dry ice before being stored at -80°C.

After surgery, animals were immediately placed in 0.5% sulfamerazine solution in axolotl housing water as a 24-hour antibiotic treatment to prevent infection of the wound site. After the treatment, all animals were returned to standard housing water, and observed to regenerate for 13 days post amputation. After 13 days, animals were once again treated with 0.1% Tricaine until fully anesthetized, and all regenerated tissue past the amputation plane was collected and placed into 500 μ L of Trizol, frozen on dry ice, and stored at -80°C (Fig. 21).

Axolotl Pancreas Resection Surgery

Age and size-matched adult leucistic axolotl salamanders were anesthetized in 0.1% tricaine solution and subjected to either a pancreatic resection surgery or a sham control surgery. The resection surgery consisted of a 2-inch lateral incision of the abdomen followed by an approximate 20% resection of the pancreas starting from the base of the duodenum. Cotton tipped applicators were used to clot any bleeding that was observed. The abdominal incision was closed with two horizontal mattress stitches using 6-0 nylon sutures. The sham surgery consisted of a 2-inch lateral incision of the abdomen followed by gentle perturbation of the pancreas after which the abdominal incision was closed with two horizontal mattress stitches using 6-0 nylon sutures. All animals were allowed to respond to this injury for 14 days after which the whole pancreas was collected. For sequencing, the pancreas was injected with RNAlater (Invitrogen, AM7020) upon collection and immediately minced using a sterile blade. Samples were then placed in RNAlater, flash frozen on dry ice and stored at -80°C. For histology, the pancreas was immediately fixed with 4% paraformaldehyde (PFA) (Fig. 22).

Axolotl AV951 Treatment Surgery

Age and size-matched leucistic axolotls were anesthetized in 0.1% tricaine solution and subjected to bilateral forelimb amputations at the mid stylopod. Animals were allowed to recover overnight in 5% sulfamerazine in standard housing water before being returned to Holtfreter's solution for the remainder of the experiment. For blastema AV951 treatments, animals were treated at 80nM AV951 to inhibit the function of VEGFR1, VEGFR2, and VEGFR3 without acute toxicity. AV951 was prepared as a 10mM stock in DMSO, and a volume was diluted into Holtfreter's solution to a concentration of 80nM into which animals were immersed for treatment. An equivalent volume of DMSO was added to Holtfreter's solution for control treatments. Animals were treated with AV951 or control DMSO at 3 days post amputation (dpa) until 11 dpa, having their treatment water changed daily, after which 11 dpa blastemas were collected for further processing (Fig. 23).

Total RNA Isolation

Samples in Trizol were defrosted on ice and homogenized at first with a disposable pestle style handheld tissue homogenizer (VWR, 76529-604), later changed to be a handheld drill homogenizer (Bio-Gen, PRO200). The pestle style homogenizer was used to disrupt samples in Trizol for about 3 minutes or until fully homogenous, while the handheld drill was used for 20 seconds or until fully homogenous. 100 μ L of chloroform was added to each sample and mixed thoroughly by hand shaking the tube. After five minutes, tubes were spun at 12,000 rpm at 4°C in a microcentrifuge to promote phase separation. The aqueous phase was collected carefully and pipetted into a new microcentrifuge tube, then equal volume of 100% Ethanol was added. The mixture was

then added to columns from the RNA Clean & Concentrator Kit (Zymo, R1013), and the subsequent cleanup steps from the protocol were followed. Samples were eluted in 15µL of RNase-free water and measured on a Nanodrop One to confirm A260/A280 above 2.0 and A260/A230 values were above 1.8. Extracted total RNA was measured for quality on an Agilent Tapestation 4200, using RNA ScreenTape (Agilent, 5067-5576) to confirm RIN scores were above 9.0. Qubit RNA BR Assay Kit (Invitrogen, Q10210) was used to measure RNA concentration on a Qubit 4 Fluorometer (Thermo Fisher, Q33238). Isolated total RNA was stored at -80 °C until thawed on ice for further use.

In Situ Hybridization Chain Reaction (HCR-FISH)

Tissues were fixed in 4% PFA in PBS, dehydrated into methanol in series, and stored at -20 °C. Upon use, samples were rehydrated and were left overnight in 30% sucrose in PBS. Samples were embedded in optimal cutting temperature compound (Sakura, 4583) and stored at -80 °C. Samples were cryosectioned at a thickness of 16 µm. *In situ* hybridization chain reaction (HCR) was performed as previously described (Lovely et al., 2022). Briefly, sectioned blastemas were washed in 2X SSC, treated with tissue clearing solution, washed in 2X SSC, incubated at 37°C in probe hybridization buffer (Molecular Instruments) before incubating overnight at 37°C in probe solution (1:200 dilution of probe in probe hybridization buffer). The next day, slides were washed in probe wash solution (Molecular Instruments) before washing in 5X SSCT and incubating at room temperature in amplification buffer (Molecular Instruments) and incubating overnight at room temperature in hairpin solution (2µl of each hairpin in amplification buffer). Slides were then washed in 5X SSCT, stained with DAPI (1:1000

in PBS), and mounted for imaging. Slides were imaged using the Nikon Spinning Disk Microscope, Zeiss Axioscan Microscope, and Zeiss LSM 900 Microscope.

ONT Library Preparation and Sequencing

Qubit RNA concentrations were used to calculate 200ng of input total RNA into the ONT PCR-cDNA v11 chemistry barcoding library preparation kit (Oxford Nanopore Technologies, SQK-PCB111.24), or the v14 chemistry kit (Oxford Nanopore Technologies, SQK-PCB114.24) for sequencing run 024. Library preparation was conducted following the v11 ONT protocol, with the only modification being the PCR conditions after some optimization experiments (see Chapter III).

Final PCR conditions used were: Initial denaturation for 3 min at 95°C. 15 cycles of denaturation (15 sec, 95°C), annealing (1 min, 62°C) and extension (4.5 min, 65°C). Final extension for 6 min, 65°C. Hold at 4°C.

After bead cleanup of the reverse transcribed PCR products, the cDNA was measured via Qubit 1X dsDNA HS Assay Kit (Invitrogen, Q33230) for concentration and Tapestation DNA 5000 ScreenTape (Agilent, 5067-5588) to quantify cDNA length distribution. The average length of cDNA was estimated from the Tapestation plots and molarity was calculated via the formula $[\text{moles dsDNA (mol)} = \text{mass of dsDNA (g)} / ((\text{length of dsDNA (bp)} \times 615.96 \text{ g/mol/bp}) + 36.04 \text{ g/mol})]$.

For MinION sequencing, samples were pooled together to sum to 25 fmol, and loaded onto the MinION R9.4.1 flow cell (Oxford Nanopore Technologies, FLO-MIN106D) on the MinION Mk1C sequencer following the standard protocol.

For PromethION sequencing run 021, samples were pooled together to sum to 25 fmol, and loaded onto the PromethION R9.4.1 flow cell on the PromethION 48 sequencer

following the standard protocol. For PromethION sequencing run 024, samples were pooled together to sum to 50 fmol, and loaded onto the PromethION R10.4.1 flow cell on the on the PromethION 48 sequencer following the standard protocol. Sequencing was run until the respective flow cell was at 0% flow cell health and no further reads could be collected – normally around 2-4 days of total sequencing time when library preparation was optimized.

Sequencing run 001 required pausing sequencing and washing the flow cell with nuclease supplied by ONT (ONT, EXP-WSH004) to reload the library multiple times to ensure maximizing the read depth. Subsequent sequencing runs did not require these washing steps as we determined quality of library preparation to be sufficient to maximize flow cell lifespan without re-loading. Raw reads were saved in pod5 format and transferred onto the FASRC cluster for downstream analysis.

Illumina Sequencing and Analysis

Total RNA samples were normalized to 5ng/ μ L in 60 μ L and submitted to the Bauer Sequencing Core Facility at Harvard University. Library preparation was executed using the Kapa mRNA HyperPrep kit and sequenced on an Illumina NextSeq 1000 at 50 bp read length through the Bauer Sequencing Core Facility at Harvard University. Fastq files were received from the Bauer Sequencing Core at Harvard University and transferred to the Harvard research computing cluster (FASRC) for further processing. Paired-end sequencing reads were quality-trimmed using Trimmomatic (v0.39). Kallisto (v22.04) was used to pseudoalign reads to the AmexT_v47 transcriptome assembly (Nowoshilow et al., 2018). Differential gene expression analysis was performed in R, using the DESeq2 package in RStudio (v2023.03.0) (Love et al., 2014).

De Novo ONT Sequencing Analysis

All data processing was conducted in the Harvard FASRC cluster, using a Nextflow/Singularity designed workflow aiming to be simple to use for the axolotl community (Fig. 4). This workflow was designed and developed in collaboration with Adnan Abouelela at ETH Zürich, who contributed most of the initial code development.

Raw reads saved in pod5 format were basecalled, demultiplexed, and filtered for minimum quality score of 8 using guppy (v6.5.7). Guppy sequencing summary files were used to generate sequencing statistic plots (e.g., percent reads passing quality threshold). Sequencing run 024, which utilized the newer PCB-114.24 library preparation kit with V14 chemistry and R10 flow cells, required the more updated dorado basecaller (v0.8.0) for basecalling, filtering (quality score 8) and demultiplexing. For run 024, dorado sequencing summary files were used to generate plots representing sequencing statistics.

Fastq files were oriented and trimmed for adapter and poly-A tails through Pychopper (v2.7.10) and Cutadapt (Martin, 2011). Reads were aligned with Minimap2, an aligner developed to work with long-read sequencing datasets (Li, 2018), to the AmexT_v47 transcriptome assembly (Nowoshilow et al., 2018). After sorting and indexing with Samtools, the resulting SAM file was quantified into a count matrix using Salmon (Patro et al., 2017). Differential gene expression analysis was performed in R, using the DESeq2 package in RStudio (v2023.03.0) (Love et al., 2014).

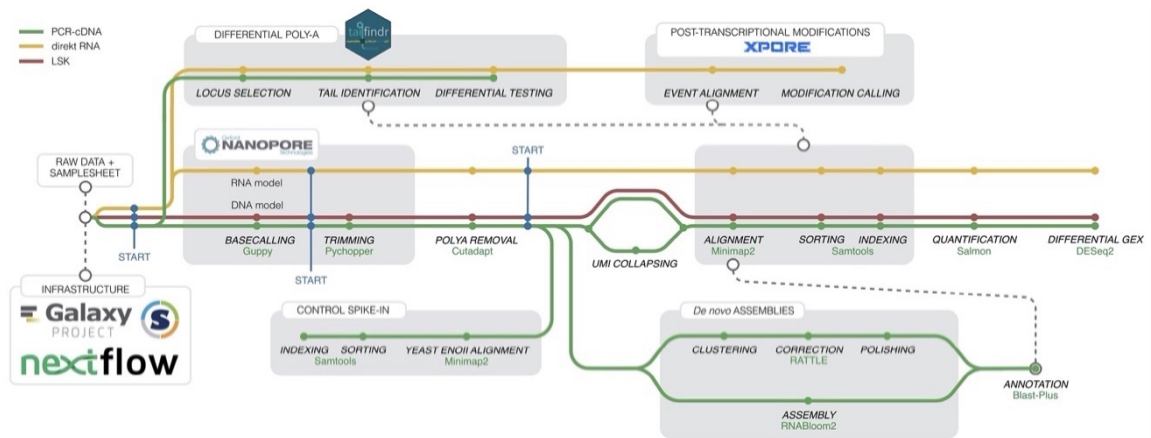


Figure 4. *De Novo* Nanopore Sequencing Analysis Pipeline

Workflow schematic of a de novo pipeline for analyzing nanopore long-read sequencing data. Included are analysis workflows for the PCR-cDNA kit (green), direct RNA kit (yellow), and ligation sequencing kit (red). The workflow also includes branches for differential poly-A analysis (top left) and de novo transcriptome assembly (bottom right). The linear center path depicts the PCR-cDNA analysis workflow used in this study. Figure generated by and adapted from Adnan Abouelela.

EPI2ME Labs ONT Sequencing Analysis

To utilize the ONT EPI2ME Labs workflow, the GitHub repository for wf-transcriptomes (<https://github.com/epi2me-labs/wf-transcriptomes>) was cloned onto the FASRC cluster. The workflow was modified to be fully functional with the cluster environment and required increasing SLURM memory requests for many of the workflow steps. Reads were aligned to the AmexT_v47 transcriptome assembly (Nowoshilow et al., 2018). After generating a count matrix from the workflow, differential gene expression analysis was performed in R, using the DESeq2 package in RStudio (v2023.03.0) (Love et al., 2014).

PCA and Z-Score Correlation Analysis

Principal Component Analysis (PCA) was conducted on output count matrices from either Salmon quantification of long reads, or Kallisto quantification of short reads. PCA was conducted using Rog or VST stabilized expression levels, which were also used for visualization in heatmaps and volcano plots. All plots were created via ggplot2 (v3.4.2) (Wickham, 2016) and pheatmap (v1.0.12), packages in Rstudio (v2023.03.0).

Correlation analysis comparing z-scores of t-statistics were created using a modified version of codebase sourced from Dong *et al.*, using DESeq2 objects generated from each relevant dataset (Dong et al., 2023).

Chapter III.

Results

The results obtained through this project were in pursuit of creating a novel reproducible Nanopore RNA-seq workflow in axolotl salamander research. To this end, we optimized the tissue isolation protocol for RNA quality and yield to ensure high quality cDNA library preparation and sequencing data. Analysis tools were developed and benchmarked pre- and post-optimization. The full workflow was then utilized for various research projects in the Whited lab at the Harvard University Department of Stem Cell and Regenerative Biology (HSCRB). Data was generated utilizing the MinION and PromethION sequencers to assay for viability of future experimental use in regeneration research.

Total RNA Isolation

Initial tests of RNA isolation were conducted with a disposable pestle style handheld tissue homogenizer, with tissues in Trizol and the lysate placed directly into a column-based RNA cleanup kit. Tissues tested included axolotl intact limb, 3-, 13-, and 28-days post amputation (dpa) blastema, heart, brain, and pancreas. Using this protocol, we observed low TapeStation RNA Integrity Number (RIN) scores, within 6.5 – 8.5 RIN range (Fig. 24). ONT recommends input RNA with RIN scores above 8.0 for Nanopore sequencing, while Illumina recommends 7.0, as low RIN scores suggest significant RNA degradation. After measuring the total RNA via Nanodrop for purity, we observed

A260/A230 values between 1.6 – 2.0. Pure nucleic acid mixtures are expected to yield an A260/A230 ratio of 2.0 – 2.2, suggesting that our initial RNA isolation protocol had a significant presence of guanidine thiocyanate, carried over from the Trizol.

To improve our RNA isolation protocol, we utilized a drill style handheld tissue homogenizer to ensure that any tissue would be fully lysed in Trizol. Adding a phenol-chloroform phase separation step to the protocol prior to the lysate being transferred to the column cleanup kit yielded significantly better Tapestation RIN scores (RIN 9 – 10) (Fig. 5). Via Nanodrop we observed A260/A230 values between 1.9 - 2.2. In addition, average Qubit concentration measurements showed between 20-80% improved total RNA yield compared to previous RNA isolation methods. This data suggested that our improved RNA isolation protocol yielded more higher quality total RNA, and the optimized protocol was used for all subsequent sequencing after sequencing run 001.

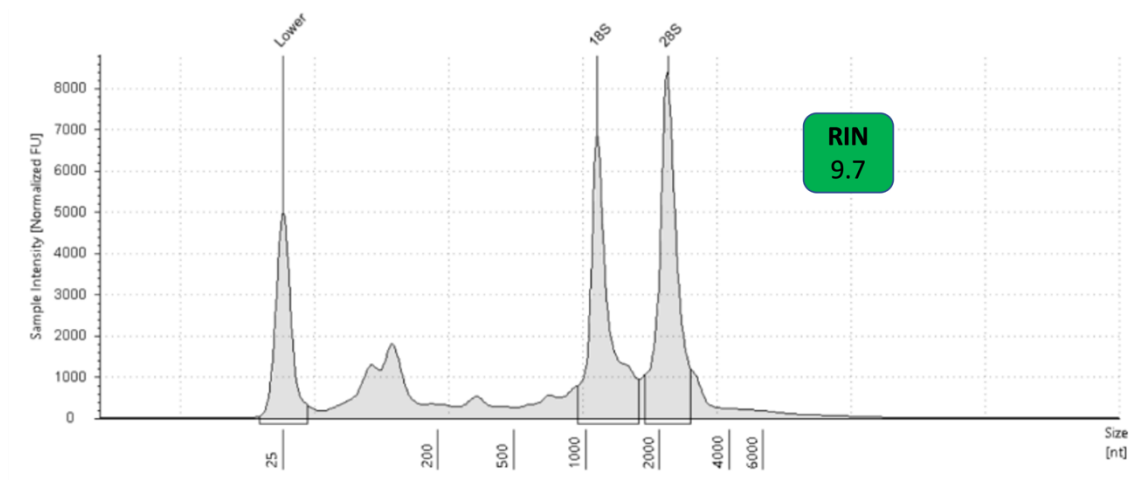


Figure 5. Total RNA Isolation Length Distribution, Post-Optimization

Representative Tapestation electropherogram depicting RNA integrity, represented by RNA integrity number (RIN) score. The x-axis represents RNA nucleotide length (nt), while the y-axis represents fluorescence (FU) to quantify abundance of RNA. The two peaks labeled 18S and 28S represent respective ribosomal RNA, while the lower peak is the internal standard (25bp) incorporated into the Tapestation buffer. The electropherogram post-optimization exhibited significantly improved RIN scores (RIN 9.7), along with increased RNA concentration relative to Figure 24.

Run 001 ONT Library Preparation

To test the efficacy of the ONT PCB-111.24 library preparation kit, we measured the output cDNA via Tapestation for length distribution and Qubit for accurate yield. Initial testing using the recommended 18 PCR amplifications cycles yielded insufficient cDNA for sequencing (<10 ng per sample). ONT support staff suggested increasing to 30 PCR amplification cycles. This yielded higher amounts of cDNA, at about 50ng per sample, and became the basis of our first run on the MinION sequencer, named sequencing run 001. Run 001 RNA was sourced from size-matched adult leucistic axolotls, comparing their intact limb and 13dpa blastemas (n=4).

We utilized the barcoded variant of the kit (“111.24” denotes V11 kit chemistry, with 24 available barcodes), allowing for multiplexing within one flow cell through the ligation of DNA barcodes during the PCR step of the library preparation. To ensure even distribution of sequencing reads, we equimolarly pooled samples, which relied on estimating the average length of cDNA in each sample via the Tapestation distribution plots (Fig. 25). After samples were loaded and sequenced, we observed lower than anticipated read count (10.5M) in our initial run. The sequencing length distribution plot showed significant deviation from the cDNA distribution plot prior to loading, notably through a spike in reads right above the recorded cutoff of 200bp - suggesting that the sequencer could have also had a significant proportion of total reads less than the 200bp threshold (Fig. 6). In addition, longer cDNA visualized in the Tapestation at 1000bp – 2000bp were absent from the read distribution.

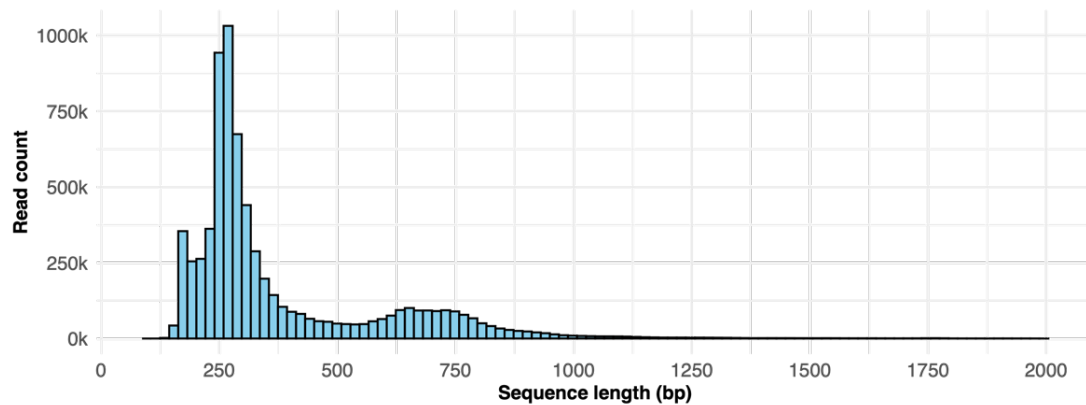


Figure 6. Run 001 Library Prepared cDNA and Sequencing Length Distributions

Read length distribution histogram generated from raw sequencing run 001 data. The sequencing distribution consists of a substantial peak at 250-300 bp, but no corresponding peak at 450-500bp as expected from data depicted in Figure 25.

Run 001 Sequencing Performance Analysis

In collaboration with ETH Zürich graduate student Adnan Abouelela, we developed a *de novo* axolotl ONT bioinformatics workflow using Nextflow. Utilizing this workflow on the Harvard FASRC cluster, we were able to basecall, demultiplex, and align run 001 to the Nowoshilow *et al.* reference transcriptome. Out of the 10.5M reads sequenced, 66.6% (7M) were classified as passing a quality threshold of quality score 8 and identified with their corresponding DNA barcode (Fig. 7). Out of these reads, only 26.5% (1.9M) were found to align to a reference transcript. This led to 1715 unique genes being identified out of the aligned transcripts. DESeq2 analysis was executed on this dataset, which uncovered 247 genes being significantly differentially expressed across the conditions ($P_{adj} < 0.05$). Differentially expressed genes (DEGs) recovered were significantly fewer relative to existing Illumina RNA-seq analysis of similar comparisons. As Illumina intact limb vs. blastema datasets uncovered ~4000 DEGs, this suggested that further optimizations were necessary. These results prompted revisiting our PCR settings (see Materials and Methods), to maximize our cDNA length distribution which was hypothesized to have a cascading beneficial effect to our read depth, percent alignment, and gene discovery.

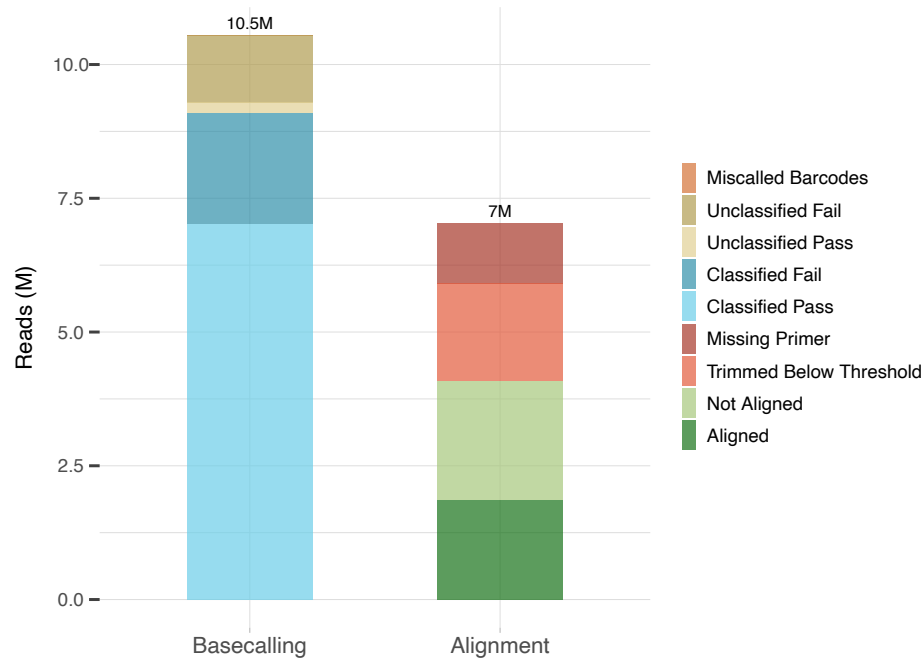


Figure 7. Run 001 Read Classification and Alignment Efficiency

Bar plots depicting read classification distribution of sequencing run 001, with y-axis representing read count in millions. Left bar represents read classification during basecalling and demultiplexing, in which passed reads are defined by mean quality score > 8 . Reads where there was no match for available barcode sequences were identified as unclassified. Right bar represents alignment and trimming statistics, in which failure to identify primers were labeled as missing primer. Instances where the trimming resulted in a read length below 200bp, reads were classified as trimmed below threshold. Out of 10.5M total reads, 1.87M (18%) aligned successfully (dark green) to the reference transcriptome.

Run 010 PCR Optimizations

High PCR cycles are not only known to increase the introduction of amplification artifacts and errors, but also overrepresentation of shorter fragments due to faster amplification of short sequences (Shagin et al., 1999). After adjustments to the denaturation, annealing and extension temperature settings to account for the high GC content of the axolotl transcripts, we conducted a PCR cycle gradient utilizing axolotl

adult leucistic 13dpa limb blastema cDNA samples to test if reducing PCR cycles would now prove effective for cDNA yield and higher average fragment length. We observed a clear reduction in the proportion of cDNA at 100-300bp, and a significant increase of cDNA of 1000bp and above in the Tapestation plots testing various PCR cycles (Fig. 26). In addition, the yield of RNA was increased to ~100ng per sample, double what we recovered with the initial PCR settings.

Run 010 Sequencing Length Distributions

To further assess the effect of various PCR cycles, 13dpa blastema cDNA generated from library preparation utilizing 15, 18, 21 and 30 cycles were loaded on a used flow cell and sequenced. The resulting sequencing run 010 generated distributions that were generally representative of their corresponding Tapestation distributions (Fig. 8). Notably, like run 001 we observed a higher proportion of shorter (<400bp) reads than the Tapestation depicted. Between the Tapestation and run 010 distribution output, we identified 15 PCR cycles as optimal for the ONT PCR-cDNA library preparation protocol.

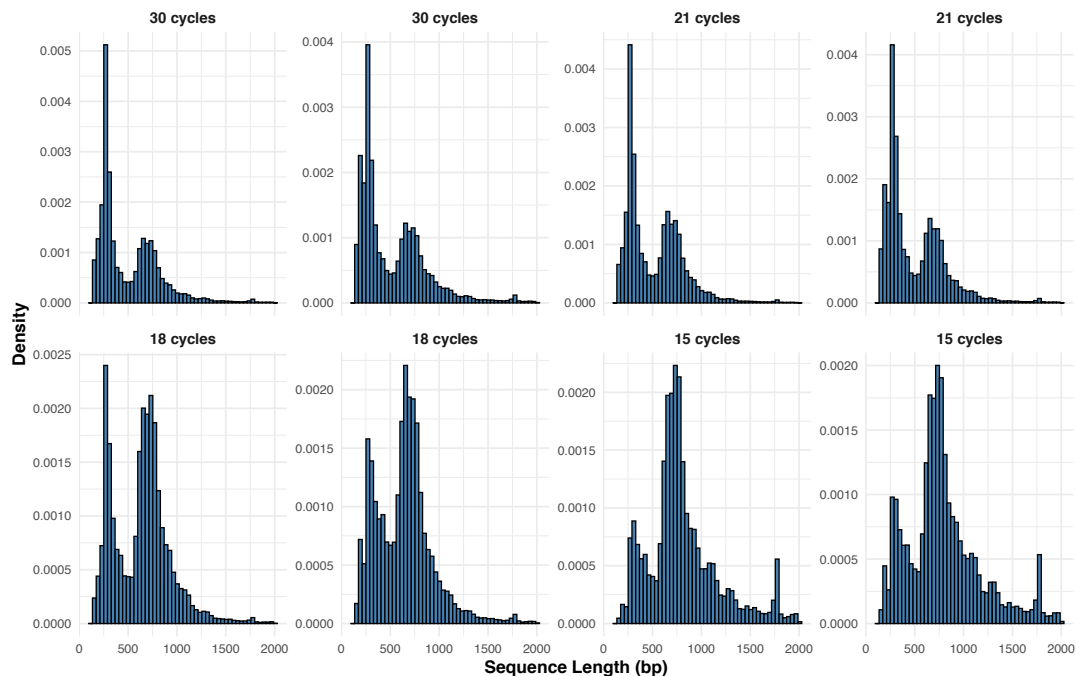


Figure 8. Run 010 Sequencing Read Length Distribution

Sequencing run 010 length distribution across 8 barcode pairs utilizing duplicates of RNA sourced from 13dpa blastemas. The x-axis represents read length (bp), and y-axis represents read count, normalized for each barcode. The distribution patterns after sequencing across 15 – 30 PCR cycles mirror patterns observed from the Tapestation electropherograms generated during library preparation (Fig. 26).

Run 011 Sequencing Performance Analysis

Using our optimized RNA isolation, PCR-cDNA library prep protocol, and our tested *de novo* ONT analysis pipeline, we collected intact limb and 13dpa blastema tissues from adult leucistic size-matched axolotls to characterize the full scope of improvements made through our workflow. Thanks to these updates, run 011 utilized higher quality RNA as input – the average RIN score was 9.9, resulting in the library prepared cDNA Tapestation distributions having an average transcript length of ~ 900bp,

and the sequencing length distribution having significantly fewer shorter reads (Fig. 30).

Run 011 acquired 32M reads, a three-fold increase from Run 001 (Fig. 9).

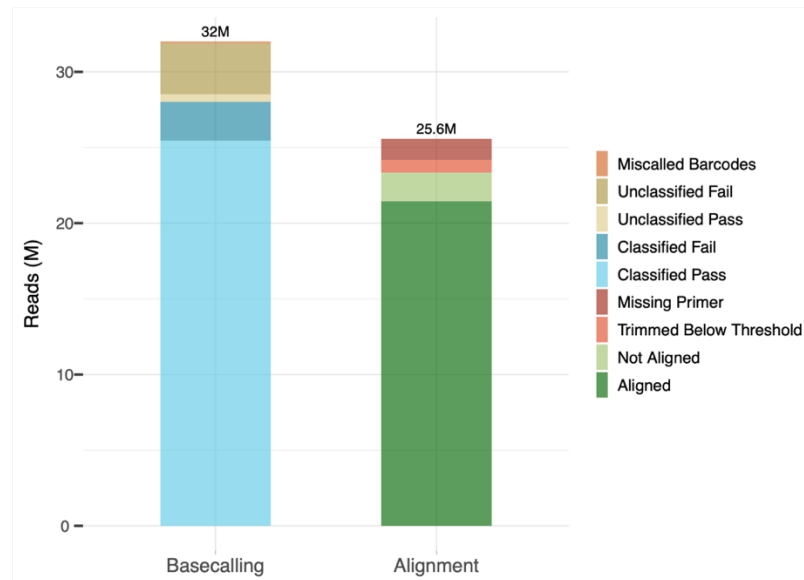


Figure 9. Run 011 Sequencing Distribution and Read Classification

BB Bar plots depicting read classification distribution of sequencing run 011, with y-axis representing read count in millions. Out of 32M total reads, 21.4M (67%) aligned successfully (dark green) to the reference transcriptome. Left bar represents read classification during basecalling and demultiplexing, in which passed reads are defined by mean quality score > 8. Right bar represents alignment and trimming statistics, in which failure to identify primers were labeled as missing primer.

Run 011 Differential Gene Expression Analysis

Out of 32M reads, 79.6% (25.5M) were classified as passing the quality threshold of quality score 8 and identified with their corresponding DNA barcode, a 3.6-fold improvement in quality-passed read depth from Run 001. Out of passed reads, 83.8% (21.4M) were found to align to a reference transcript, an 11.5-fold improvement from

Effect of Optimizations on DEG Discovery

Our optimizations to the ONT PCR-cDNA workflow led to the discovery of 7250 unique genes being identified out of the aligned transcripts, a nearly four-fold increase relative to Run 001 (Fig. 11). It is well-known that increased sequencing depth can result in a significant increase in coverage, gene discovery, and DEG identification (Li et al., 2021; Patterson et al., 2019; Wang et al., 2009). Consequently, we observed a 15-fold increase in DEGs Run 011 DEG numbers fall within 10% of the expected range (~4000 DEGs) of similar Illumina datasets, suggesting that the 32M reads was within the satisfactory sequencing depth required for this experiment using Nanopore-seq.

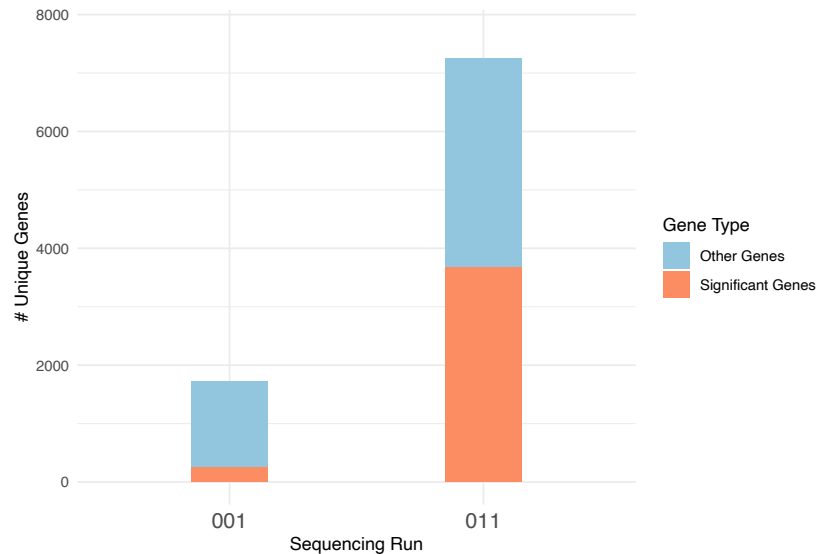


Figure 11. Run 001 vs. 011, Unique Genes Discovered

Bar plots depicting the increase in unique genes identified between sequencing run 001 (left) and 011 (right). The y-axis represents number of unique aligned genes. Each bar is divided into genes that are significantly differentially expressed (orange, $P_{adj} < 0.05$), and genes identified as not significant (blue). Run 011 shows a substantial increase in both the total number of genes (323%) and DEGs discovered (1385%) compared to run 001.

Run 019 Pancreas Resection Experiment

To test our MinION workflow in an experimental application, we isolated RNA from resected 14 days post injury (dpi) axolotl pancreases and a sham-operated control (Fig. 22). Standard homogenization techniques with test pancreas samples generated low RIN scores below 7.0 on average, and required modifications based on Jun *et al.*, in which treating samples with RNAlater led to increased RNA stability (Jun et al., 2018). Using this modified protocol, we identified significantly improved RIN scores, between 8.0 - 9.0. Interestingly, the library preparation produced a length distribution that was significantly different than run 011, suggesting that the changes in transcript expression profile of various tissues can be observed through this assay (Fig. 27). Unfortunately, due to the timing of the sequencing run being delayed by the pancreas RNA isolation optimization tests, a partially degraded flow cell was used for sequencing run 019.

Run 019 Sequencing Results

Sequencing run 019 generated a total of 12.5M reads, with 693 DEGs comprising the 2473 unique genes discovered from the dataset (n=4) (Fig. 12A). PCA identified 61% variance across PC1, with the plot showing clear separation across the sham and resection conditions (Fig. 12B). We identified key pancreas marker gene *Insulin* as differentially expressed, as well as known markers of regeneration like *Cirbp* (Bryant et al., 2017; Sørensen et al., 2022).

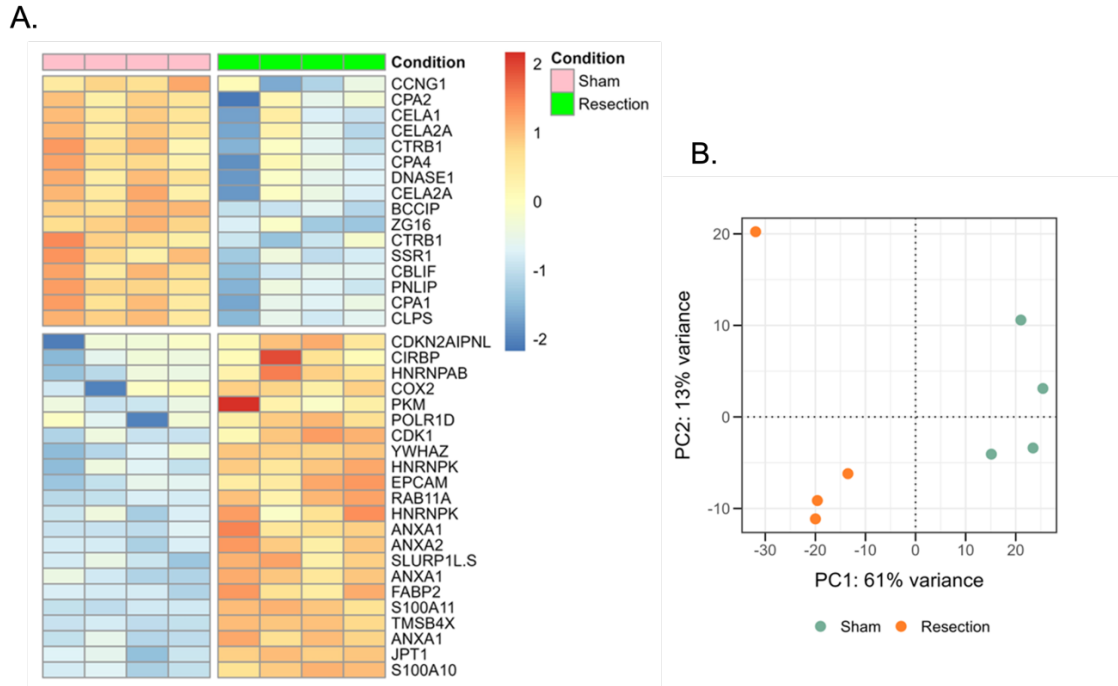


Figure 12. Run 019 DGE Heatmap and PCA Plot

(A) Heatmap depicting 38 key DEGs from sequencing run 019 (z-score normalized). Blastema enriched genes like *Cirbp* were identified as significantly upregulated in 14dpi pancreas resection relative to sham controls. (B) Principal component analysis (PCA) plot generated from run 019 data shows 61% variance of the first principal component (PC1), indicating significant differences in the transcript expression pattern captured between the sham and resected pancreas conditions. Plots generated by Connor Powell.

Run 019 Validation Results

To validate the presence of transcripts identified in the sequencing data, HCR-FISH was performed on 14dpi, 28dpi and their respective sham samples. As expected, *Insulin* was found to have positive signal in all conditions, while *Cirbp* was found to have signal in only the resected conditions (Fig. 13A). Interestingly, the endocrine marker *Pdx1* was found to have positive signal at 14dpi yet was not one of the genes identified in the run 019 dataset (Fig. 13B) (Ebrahim et al., 2022; Li et al., 2010).

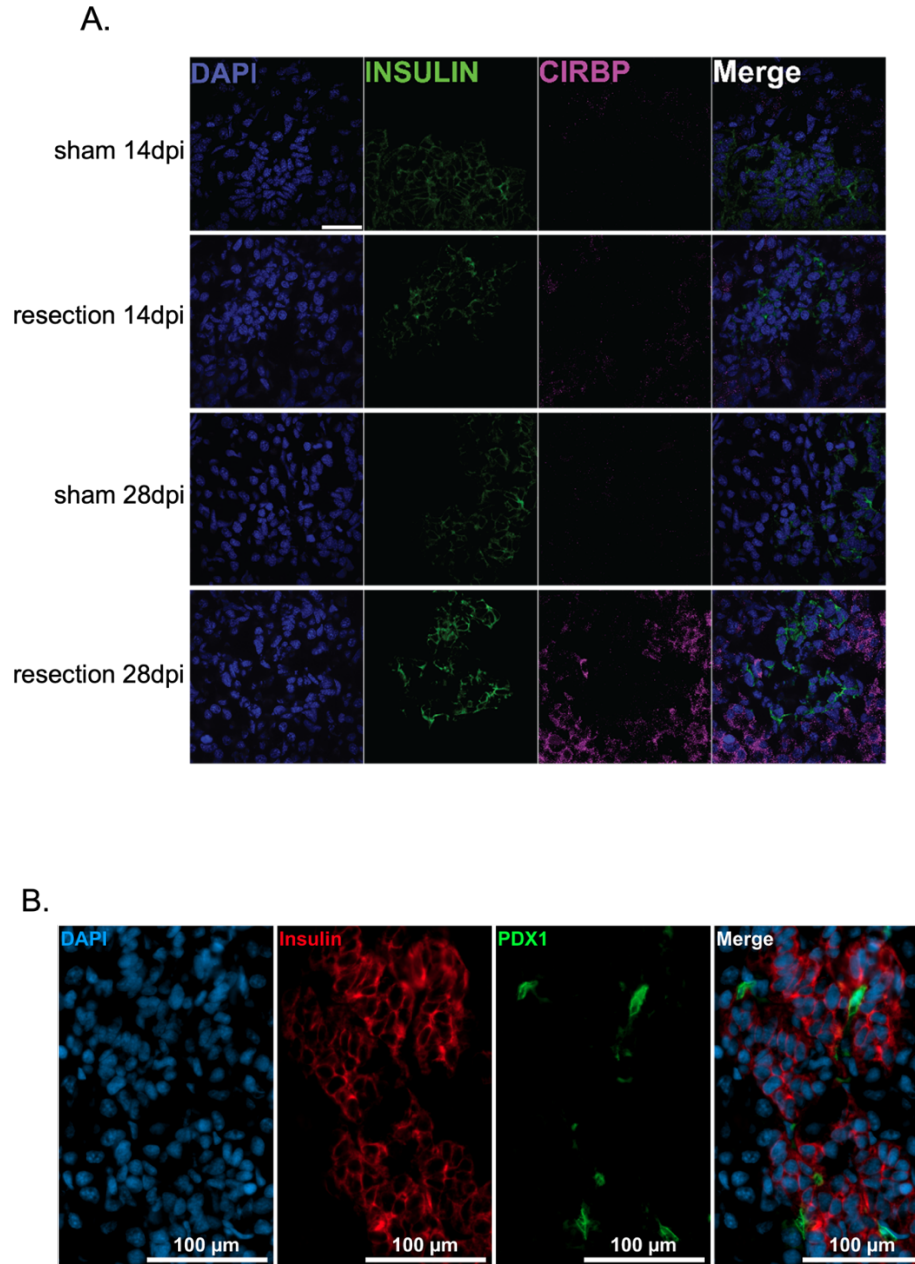


Figure 13. HCR-FISH Staining of Sham and Resected Pancreas

(A) HCR-FISH highlighting differential expression of *Insulin* and *Cirbp* transcripts in sham vs. resected pancreas samples at 14- and 28-days post injury (dpi). Interestingly, *Cirbp* signal was observed to be significantly increased at 28 dpi as well (scale bar = 50 μ m). (B) HCR-FISH of *Insulin* and *Pdx1* in resected 14dpi pancreas. *Pdx1* is a pancreas marker gene that was not identified by sequencing run 019, suggesting insufficient sequencing depth using the MinION sequencer. Images generated by Connor Powell and Ashley Juarez.

Run 018 VEGF Pathway Inhibition Experiment

To further test the MinION workflow in experimental axolotl applications, we isolated RNA from a VEGF pathway inhibition study in which animals were amputated and treated with control DMSO or AV951, a potent VEGF inhibitor, with tissues being collected at 11 days post amputation (Fig. 23). By inhibiting the VEGF pathway, we sought to show that VEGF signaling is essential for limb regeneration and the effects of a loss of VEGF signaling were on the blastema at the transcriptional level. Using the standard optimized protocol, we observed high RIN scores and optimal PCR yields across samples that were library prepared, suggesting that our workflow is well suited for chemically treated tissues.

Sequencing run 018 resulted in 17.8M reads on the MinION sequencer (n=3), with 72.6% reads passing quality score threshold 8 and out of the passed reads 87% being aligned to the reference transcriptome (Fig. 14).

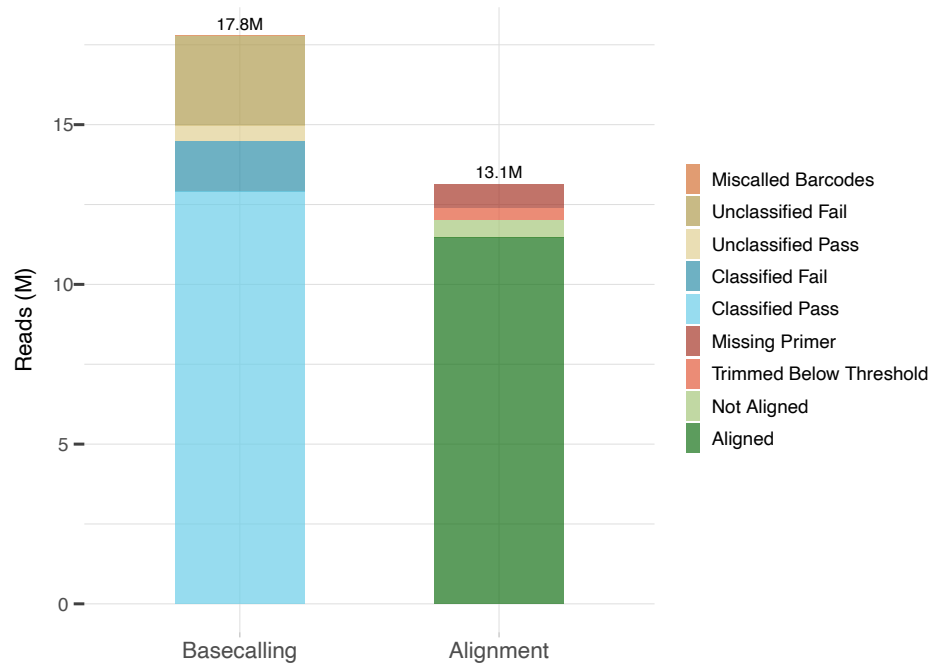


Figure 14. Run 018 Read Classification and Alignment Efficiency

Bar plots depicting read classification distribution of sequencing run 018, with y-axis representing read count in millions. Left bar represents read classification during basecalling and demultiplexing, in which passed reads are defined by mean quality score > 8 . Reads where there was no match for available barcode sequences were identified as unclassified. Right bar represents alignment and trimming statistics, in which failure to identify primers were labeled as missing primer. Instances where the trimming resulted in a read length below 200bp, reads were classified as trimmed below threshold. Out of 17.8M total reads, 11.5M (65%) aligned successfully (dark green) to the reference transcriptome.

Run 018 Results and Validations

From a total of 17.8M reads, sequencing run 018 resulted in the discovery of 2970 unique genes, including of 140 DEGs. The PCA plot showed significant (61%) variance of PC1, showing separation across the control and AV951 treated conditions (Fig. 15A). One DEG significantly downregulated in the AV951 treatment group was *Sparc*, a gene that has been implicated as critical for the wound healing process

(Barrenas et al., 2019). HCR-FISH validated the decreased presence of *Sparc* in the treatment group, suggesting its importance in the VEGF pathway (Fig. 15B-C).

After sequencing run 018, Illumina short-read sequencing was conducted on the same total RNA samples, in which 3286 DEGs were identified, comprising of all 140 DEGs discovered in the run 018 dataset. Additionally, the correlation of the t-test of the shared DEGs resulted in the expected range of correlation (Pearson's $r = 0.688$) between Illumina and ONT datasets (Fig. 28).

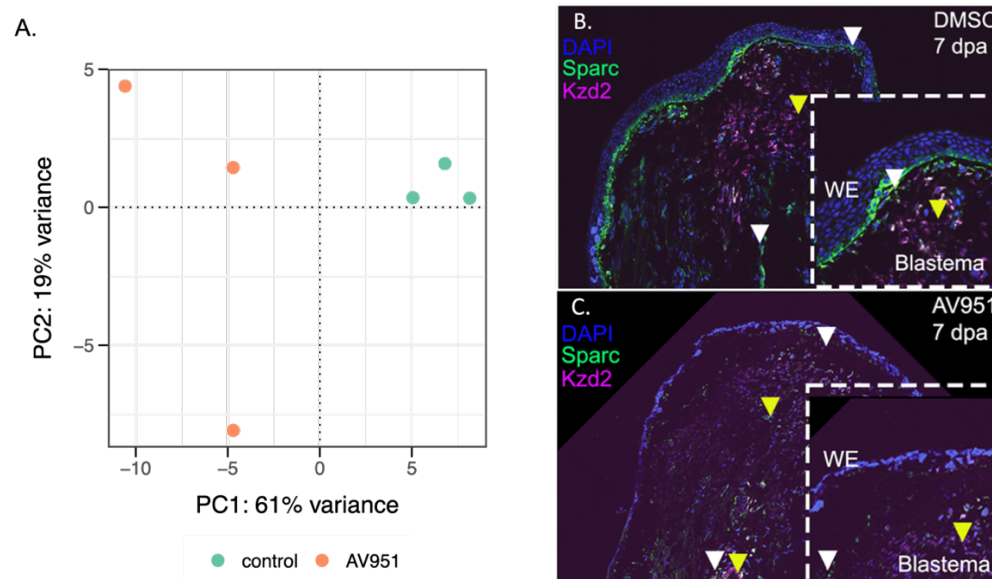


Figure 15. Run 018 PCA plot and HCR-FISH Validations

(A) Principal component analysis (PCA) plot generated from run 018 data shows 61% variance of the first principal component (PC1), indicating significant differences in the transcript expression pattern captured between the control and AV951 treated blastema conditions. (B-C) HCR-FISH validated the downregulation of *Sparc* in the control and AV951 treatment (credit Aaron M. Savage).

Run 021 PromethION Platform Initial Testing

Sequencing run 021 utilized the more robust PromethION platform (v11 chemistry), loading identical library prepared cDNA as run 011 sourced from intact limb and 13dpa blastema (n=4). Sequencing on a degraded flow cell with only 3,890 pores available out of 12,000, we recovered 56.7M reads, at 73.7% passing a quality threshold score of 8 and out of the passed reads 78% aligning to the reference transcriptome (Fig. 16).

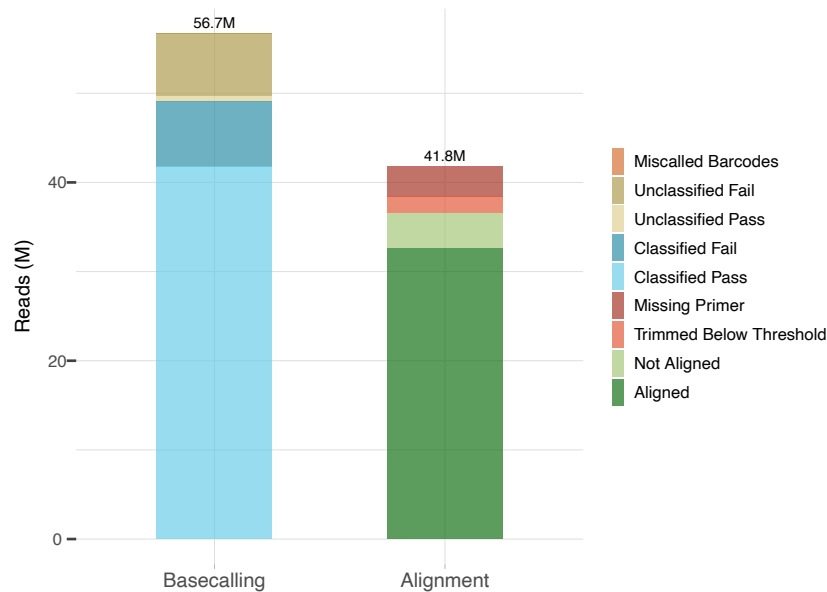


Figure 16. Run 021 Read Classification and Alignment Efficiency

Bar plots depicting read classification distribution of sequencing run 021, with y-axis representing read count in millions. Left bar represents read classification during basecalling and demultiplexing, in which passed reads are defined by mean quality score > 8 . Reads where there was no match for available barcode sequences were identified as unclassified. Right bar represents alignment and trimming statistics, in which failure to identify primers were labeled as missing primer. Instances where the trimming resulted in a read length below 200bp, reads were classified as trimmed below threshold. Out of 56.7M total reads, 32.6M (57%) aligned successfully (dark green) to the reference transcriptome.

Sequencing Run 011 vs. 021

Sequencing run 021 identified 8785 unique genes (21% increase from run 011), with 4222 DEGs discovered (15% increase) (Fig. 33). Out of the 3053 DEGs shared between the datasets, we found a highly significant correlation (Pearson's $r = 0.942$), suggesting strong agreement between the data (Fig. 17). The relatively mild improvement between the datasets despite the improvement of read depth (250% increase) indicates that run 011 had relatively saturated sequencing, with further increases in read depth unlikely leading to significant improvements in the dataset.

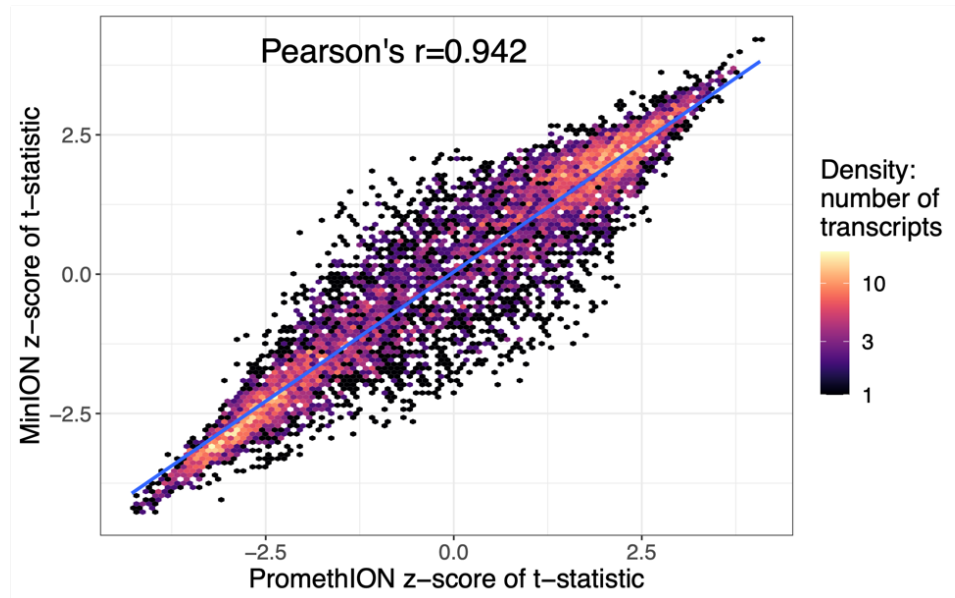


Figure 17. Run 021 Unique Genes Discovered and Correlation of Z-Scores

Correlation analysis plot between run 021 t-statistics of identified genes (x-axis) and run 011 t-statistics of identified genes (y-axis) using the same cDNA. High correlation (Pearson's $r=0.942$) was observed between the two sequencing runs.

Run 024 PromethION Platform Experimental Testing

PromethION sequencing run 024 used identical total RNA as MinION sequencing run 019, sourced from sham and resected axolotl pancreases (Fig. 22). Samples were sequenced on the PromethION platform (v14 chemistry) with a modified library preparation and analysis protocol detailed in the Materials and Methods section. Sequencing run 024 generated 107.6M reads in total, of which 80.7% passed quality score filter 8, and out of these 90% were aligned to the transcriptome (Fig. 18).

Interestingly, we observed a significant portion of reads that were classified to barcodes that were unused (7.9%), while in other runs most low-quality reads were classified as “unclassified”, meaning no match to the barcodes in the database. As the percent passing the quality score threshold is still consistent relative to the prior runs, the miscalled barcodes are a result of the change in basecalling software from Guppy to Dorado. Guppy marks reads as below quality threshold and does not assign them barcodes, while Dorado still assigns low quality reads a barcode when possible. Consequently, for Dorado, errors in the barcode region of the read lead to reads being identified as nonexistent barcodes rather than being thrown out as unclassified and failing to meet quality threshold.

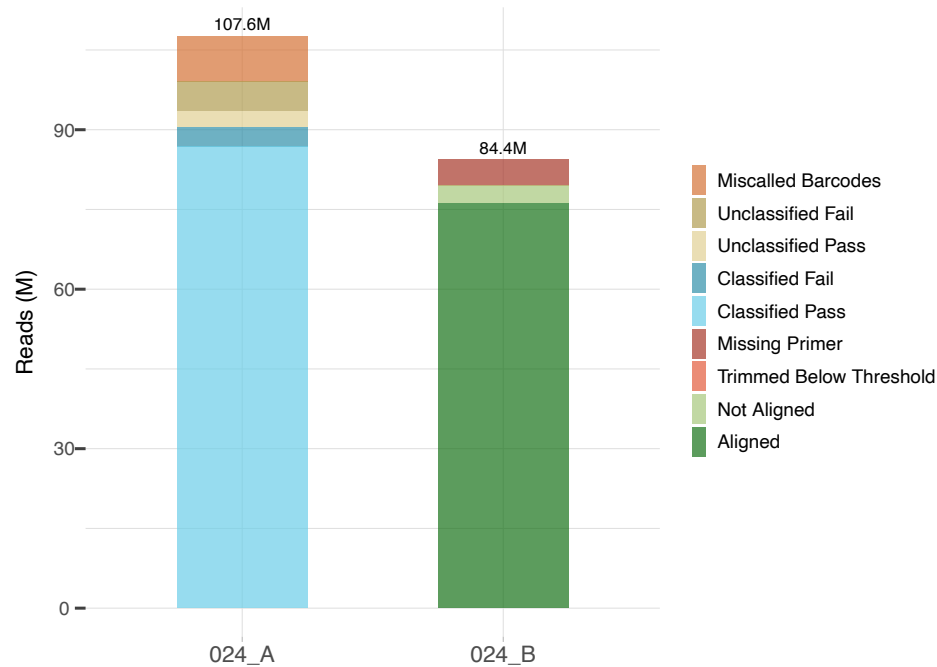


Figure 18. Run 024 Read Classification and Alignment Efficiency

Bar plots depicting read classification distribution of sequencing run 024, with y-axis representing read count in millions. Left bar represents read classification during basecalling and demultiplexing, in which passed reads are defined by mean quality score > 8. Reads where there was no match for available barcode sequences were identified as unclassified. Right bar represents alignment and trimming statistics, in which failure to identify primers were labeled as missing primer. Instances where the trimming resulted in a read length below 200bp, reads were classified as trimmed below threshold.

Sequencing Run 019 vs. 024

From the 107.6M reads in sequencing run 024, we identified 8539 unique genes (3.5-fold increase from run 019), and 1317 DEGs (a 1.9-fold increase) (Fig. 34).

Correlation between the 521 overlapping DEGs was extremely high (Pearson's $r=0.951$) (Fig. 19). The nearly 9-fold increase in sequencing depth between runs 019 and 024 led to the identification of key pancreatic marker gene *Pdx1* in the dataset, being identified as a DEG and found to be downregulated in the resected pancreas. Additionally, *Kazald2*, a

known regeneration marker was uniquely discovered in the run 024 dataset and exhibited significant upregulation with a log2FC of 5.03.

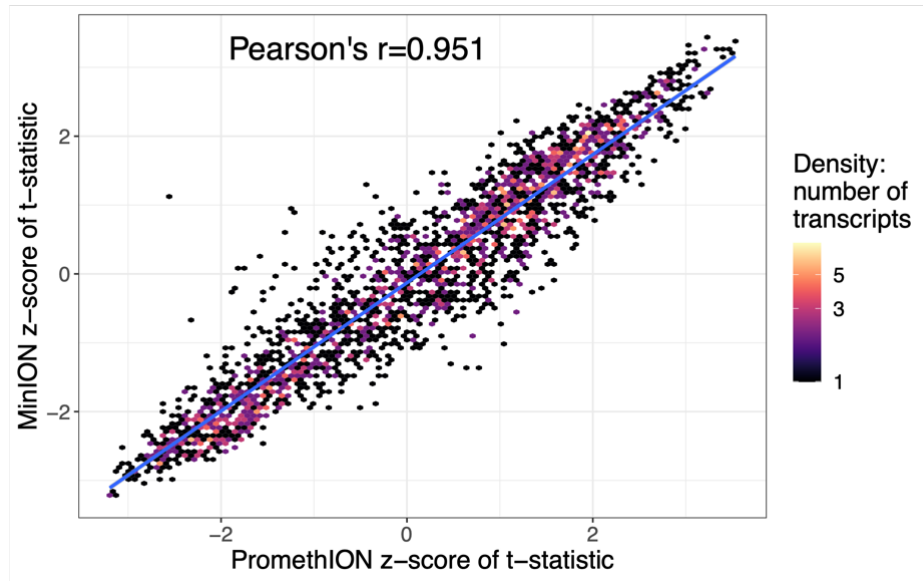


Figure 19. Run 024 Correlation of Z-Scores

Correlation analysis plot between run 024 t-statistics of identified genes (x-axis) and run 019 t-statistics of identified genes (y-axis) using the same cDNA. High correlation (Pearson's $r=0.951$) was observed between the two sequencing runs.

Chapter IV.

Discussion

This project and its experiments aimed to establish a protocol for long-read nanopore sequencing in the axolotl salamander model with the ONT PCR-cDNA kit. Our goal was to optimize and develop a reproducible nanopore bulk RNA-seq pipeline for the axolotl salamander model and provide researchers in the Whited Laboratory at Harvard University an opportunity to utilize long-read sequencing for their experimental applications. Using this novel workflow for nanopore RNA-seq in the axolotl, we executed multiple optimization and experimental sequencing runs utilizing the ONT MinION Mk1C sequencer and the PromethION 48 sequencer.

Optimizations of Nanopore Workflow

First, we focused on improving the tissues collection and RNA isolation protocol. For all axolotl tissues tested, increasing tissue disruption intensity and using standard phenol-chloroform phase separation paired with commercially available column-based RNA cleanup kits produced reproducible higher RNA yield and integrity than longer disruption times with less intensity (Fig. 5 and 24). Our findings show that with Trizol, a commercially available mixture of phenol and guanidinium isothiocyanate, tissues can be aggressively disrupted without reduction in RNA integrity.

In addition, optimization of PCR conditions in the cDNA library preparation protocol produced a marked improvement in the final distribution of read length prior to

loading, leading to increased sequencing read length distribution (Fig. 8). As expected, we found that attaining more full-length reads led to a higher percentage of reads passing quality thresholds as well as a higher percentage of reads aligning to the reference transcriptome (Fig. 9 and 30). More surprisingly, improvements in the cDNA library preparation led to more than three times the sequencing depth in the optimized run. As standard nanopore sequencing settings immediately remove all reads below 200bp, the sharp spike in reads observed near this threshold is attributed to a significant presence of fragmented reads in the pre-optimized sequencing run 001. Subsequent runs exhibit significant reduction in the fragmented read spike, leading to improved read quality and yield.

Across all sequencing runs, we observed a disparity between the cDNA size distribution represented by TapeStation electropherogram and the read-length sequencing histogram generated by the MinION and PromethION (Fig. 6 and 25). We propose that there is preferential sequencing of shorter cDNA due to longer strands of cDNA spending more time passing through the nanopores. This further highlights the importance of the initial RNA isolation and library preparation optimization, as the effect of any presence of initial RNA degradation – and therefore shorter fragmented cDNA - will be preferentially amplified during the PCR as well as sequencing and reduce downstream alignment performance.

Nanopore Bioinformatics and Axolotl Research

While ONT offers EPI2ME, a cloud-based data analysis platform designed for analyzing and visualizing nanopore sequencing data, the large size of the axolotl reference genome and transcriptome caused significant issues during the alignment step

of the pipeline, resulting in crashes of all workflows attempted. The high computational requirements of the alignment step in the RNA-seq analysis workflow led to the commercially available pipeline encountering catastrophic problems running axolotl data.

To ensure more nuanced control over troubleshooting each step of the workflow, we developed a *de novo* ONT PCR-cDNA analysis workflow using Nextflow, which was built to run Guppy to basecall, Pycloppe and Cutadapt to orient and trim, MiniMap2 to align, Samtools to sort and index, and Salmon for quantification (Fig. 4). Note that for run 024, which utilized the newer V14 chemistry and R10 flow cells, we used Dorado for basecalling as Guppy is no longer compatible with the newer kits.

The pipeline is run on the Harvard Faculty of Arts and Sciences Research Computing (FASRC) cluster, which enables access to the powerful computational resources necessary to handle the axolotl alignment. Through the course of this project, ONT released updates to EPI2ME Labs, a local version of EPI2ME which we modified to work on the FASRC cluster. Thus, we were able to validate results of our *de novo* pipeline by comparing it to the EPI2ME Labs workflow and found the output to be highly correlated (Fig. 29).

Testing MinION Workflow in Axolotl Research

Optimizations for the PCR-cDNA kit on the ONT MinION sequencer utilized axolotl intact limb and 13 days post amputation (dpa) limb blastema tissues, as this early- to mid-bud blastema stage has been previously characterized and relevant to future experimental tissue types in axolotl (Darnet et al., 2019; Leigh et al., 2018; Sousounis et al., 2020). However, typical experimental applications of RNA-seq in axolotl focus on more granular sample condition comparisons, like blastema transcript expression profile

across transgenic animals, or chemical inhibition of signaling pathways to characterize the effects on limb blastema formation (Fei et al., 2017; Leigh et al., 2018; Leigh et al., 2020). As the goal of bulk RNA-seq analysis is to identify and calculate significant changes in transcript abundance across conditions, experimental designs in which the conditions express highly similar transcript expression profiles require improved sequencing depth to ensure the full capture of the subtle changes in transcript abundance. Therefore, while it was promising that our benchmarking tests comparing intact limb to 13dpa blastemas yielded ~3700 DEG's, we were aware that the ~30M read sequencing depth achieved could prove to be a limiting factor in reproducing these results in more granular experimental comparisons.

To test our workflow in more granular experimental contexts, we began to supply various axolotl research projects with long-read sequencing data. In an axolotl pancreas study, in which a novel resection surgery of the pancreas (20% of the duodenum) was utilized to characterize pancreas injury response over time, we used our optimized nanopore RNA-seq workflow using the MinION sequencer (Fig. 12A-B). Whole pancreas from 14 days post injury (dpi) and a sham control were collected, isolated for RNA, sequenced and analyzed (n = 4). We discovered nearly 700 DEGs and observed clear separation of the resected and sham conditions in the PCA plot, suggesting sufficient sequencing depth was achieved. In addition, expected marker genes of the pancreas like *Insulin* was identified in the sequencing data, while known regeneration marker *Cirbp* showed up as significantly upregulated ($P_{adj} < 0.05$) in the resected condition. *Insulin* and *Cirbp* were validated via HCR-FISH to confirm the findings from

the sequencing data (Fig. 13A). However, *Pdx1*, a known marker of endocrine cells, was not represented in the final count matrix (Fig. 13B).

We observed a range of sequencing successes and failure for other experiments as well. One project, focusing on the vascular endothelial growth factor (VEGF) pathway, tested 11dpa limb blastemas injected with AV951, a VEGF receptor inhibitor, compared to dimethyl sulfoxide (DMSO)-treated control limbs (n=3). AV951 is a known potent inhibitor of the VEGF signaling pathway, which triggers several key signaling pathways such as the Ras/MAPK pathway. Thus, it was hypothesized that there would be many transcriptional changes resulting from the inhibitor treatment, yet MinION sequencing led to only 140 DEGs being discovered (Fig. 14). It is likely that the similarity of input tissue across the conditions (both involved blastemas receiving injections) required a higher sequencing depth than the MinION's capability. Still, one of the genes of interest identified from the DEG list, *Sparc*, was validated in a follow up HCR-FISH staining (Fig. 15B-C). In addition, this preliminary pilot sequencing showed enough promising data that the lab group chose to invest in a full Illumina short-read sequencing run, in which nearly all DEGs discovered with the MinION appeared in the Illumina dataset and was correlated within the expected range as shown in Figure 28 (Dong et al., 2023). The Illumina dataset uncovered 3286 DEGs, further showing the limitations of the MinION sequencing depth in more sensitive experimental contexts.

Success of a bulk RNA-seq experiment is not strictly defined by raw DEG yield, but include other factors like observing expected marker genes, unique new findings, and identifying separation of experimental conditions through PCA analysis. Illumina bulk RNA-seq runs in axolotl typically yield between 2000-4000 DEGs, while no

experimental MinION nanopore sequencing run in this study yielded more than 800 DEGs. However, apart from DEG yield, we also observed key marker genes missing from MinION datasets, and limited novel DEG identifications due to the lower sequencing depth. Our conclusion is therefore that the MinION does not have the necessary sequencing depth to function as the primary sequencing methodology for experiments that typically encompasses axolotl regeneration research.

PromethION Sequencing in Axolotl

We sought to overcome the lack of sequencing depth identified from the MinION experimental runs by attempting to sequence on the PromethION platform. This more robust platform uses different flow cells, which are designed with 5x the number of nanopores, allowing for up to 180-200M reads, vs. 20-30M for the MinION (Fig. 20). As the optimization experiments showed that increased sequencing depth can result in a significant increase in gene discovery and number of DEGs, our hypothesis was that a PromethION run would rescue the experiments in which we identified insufficient sequencing depth from the MinION.

Since the PromethION flow cells are sold in a set of four, it was beyond the scope of this project to purchase a full set. However, the Harvard Bauer Sequencing Core offered us a partially degraded flow cell (33% capacity) at a discount and allowed us to use their PromethION 48 sequencer at no additional cost. Note that even a 1/3 capacity PromethION flow cell has the potential to sequence at 3x the read depth of a MinION flow cell. We chose to repeat sequencing run 011 (now PromethION run 021), to identify whether the library preparation improvements we observed would be applicable to the PromethION platform. Since the library preparation protocol for a MinION and

PromethION sequencer are fundamentally identical, we utilized cDNA prepped from run 011 (stored in -80°C). The PromethION was able to sequence 2.3-fold more reads (57M) than the MinION (25M) and identified 1300 more overall genes, contributing to an increase of 10% more DEGs (Fig. 16 and 33). In addition, the correlation between the MinION and PromethION datasets were extremely high (Pearson's $r=0.942$) (Fig. 17).

The combination of these improvements on a dataset that was already approaching sequencing depth saturation, as well as the large overlap between MinION and PromethION data, were both promising results for the future of nanopore sequencing in axolotl research. We hypothesized that the 57M reads we observed from the PromethION will be further increased by using a full capacity flow cell.

In collaboration with the Jain lab at Northeastern University, we were given an opportunity to sequence on a full capacity PromethION flow cell for the pancreas resection experiment. We utilized leftover total RNA from MinION sequencing run 019 to library prepare and sequence using the newer generation R10 flow cell. We were able to sequence 108M reads, nearly 9x more than the MinION run, and identified 6,066 more genes. 1,317 DEGs were discovered, a nearly two-fold increase relative to run 019 with high correlation (Pearson's $r = 0.951$) (Fig. 19 and 34). Key genes from the MinION run (*Cirbp*, *Insulin*) were reproduced in the PromethION data, while additional pancreas marker genes like *Pdx1*, and regeneration markers like *Kazald2* are discovered to be DEGs when they were previously filtered out of the run 019 dataset. *Pdx1* was previously validated via HCR-FISH to be expressed in the regenerating pancreas, further highlighting the suboptimal sequencing depth of the MinION in this specific experimental context (Fig. 13B).

The significant increases in sequencing depth led to increases in overall gene discovery and DEG identification of both MinION to PromethION comparisons using identical input RNA (Fig. 17 and 33). These improvements in the data are evidence that our optimizations in RNA isolation and library preparation were applicable to both ONT sequencing platforms. While the MinION sequencer may not achieve the coverage necessary for most experiments in axolotl research applications, our findings show that MinION data can accurately predict the outcome of a larger sequencing experiment – the high correlation between the datasets show that DEGs identified from the MinION are still represented accurately not only in the PromethION datasets, but also in Illumina data.

Chapter V.

Conclusion

Through this project we were able to design a novel and reproducible nanopore bulk RNA-seq pipeline from tissue collection to an output of differential gene expression analysis. We present the first study employing nanopore RNA-seq in the axolotl salamander. Future studies will benefit from our optimization of RNA isolation of various axolotl tissue types and cells due to higher quality RNA as input material improves the performance of both short- and long-read sequencing. As axolotl research utilizing long-read sequencing will inevitably encounter computational load issues due to the size of the reference genome and transcriptome, we propose that our *de novo* pipeline offers a more efficient alternative workflow for axolotl nanopore sequencing data analysis.

While our initial aim was to optimize a workflow specifically for use of our in-house MinION sequencer, we found that, in combination with the PCR-cDNA library preparation kit, the MinION has suboptimal sequencing depth for tested axolotl research applications. We therefore validated our workflow and analysis pipelines to ensure full compatibility with the more robust PromethION platform. We found significant improvements in read depth and differentially expressed gene (DEG) discovery with the PromethION platform through a study on axolotl pancreas, validating key genes through HCR-FISH. Interestingly, most MinION genes were still present in the PromethION dataset with extremely high correlation, highlighting the validity of the smaller scale

sequencer. The pancreas resection study utilizing this novel nanopore bulk-RNA sequencing in the axolotl is currently being prepared for submission to a research journal.

Additionally, even though a MinION sequencing run in an axolotl VEGF pathway inhibition study encountered insufficient read depth and identified only 140 DEGs, we were able to validate a key DEG (*Sparc*) through HCR-FISH. As a pilot experiment, these initial results generated enough interest to justify investment for an Illumina sequencing run. Subsequent discoveries and follow-up experiments based on the Illumina sequencing run data helped enable the submission of the study to a research journal, demonstrating the value of the MinION platform as a pilot sequencing device to acquire preliminary data. Therefore, we conclude that nanopore long-read bulk RNA-seq can serve in axolotl research as a cost-effective alternative to Illumina short-read sequencing, whether for generating pilot data (MinION run 018) or as the main sequencing platform of a project (PromethION run 024).

Bulk RNA-seq is only scratching the surface of the capabilities of the nanopore sequencing technology. Nanopore-seq can enable epigenomic analysis, genome and transcriptome assembly, and protein sequencing is currently being developed on the platform. Axolotl research will greatly benefit from the ability to sequence molecules in real time at a fraction of the cost of competing technologies, but future studies should be mindful of the limitations in working with the axolotl model.

Appendix 1.

Overview of Oxford Nanopore Technology Hardware

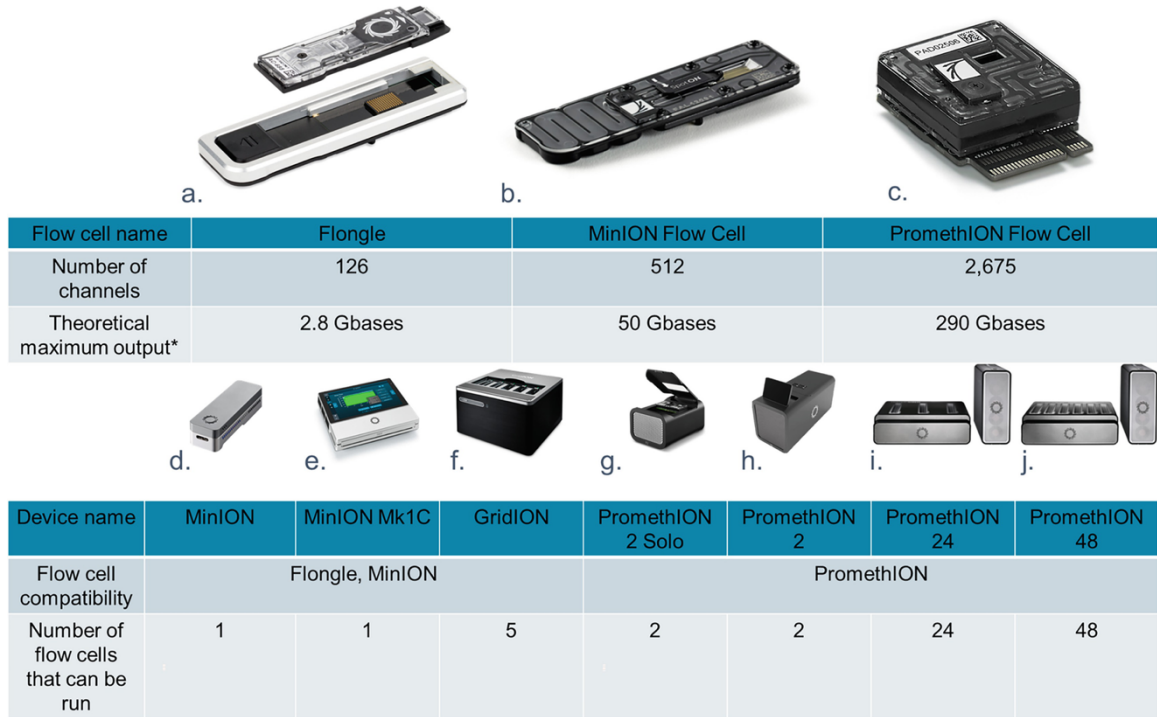


Figure 20. ONT Flow Cells and Sequencing Platforms

List of flow cells (a-c) and sequencing platforms (d-j) provided by ONT. Note that in this study we benchmarked the MinION flow cell (b) in combination with the MinION Mk1C sequencer (e), and the PromethION flow cell (c) in combination with the PromethION 48 sequencer (j). Figure sourced from Oxford Nanopore Technologies.

Appendix 2.

Axolotl Experimental Surgery Model Illustrations

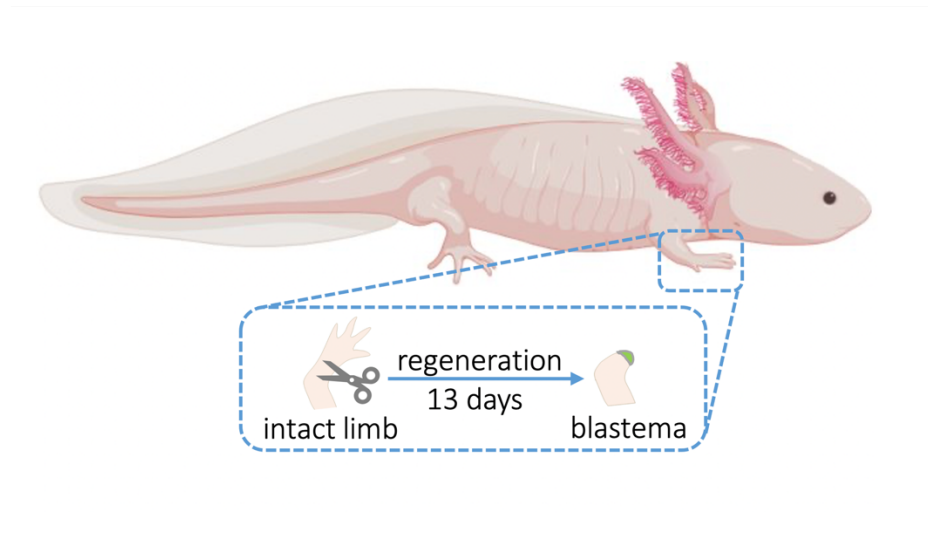


Figure 21. Axolotl Limb Amputation Experiment Illustrations

Illustration of the axolotl 13 days post amputation (dpa) limb blastema surgery used for sequencing runs 001, 010, 011 and 021. Each pair of intact limbs and 13dpa blastema were utilized for sequencing (n=4). Image generated via Biorender.

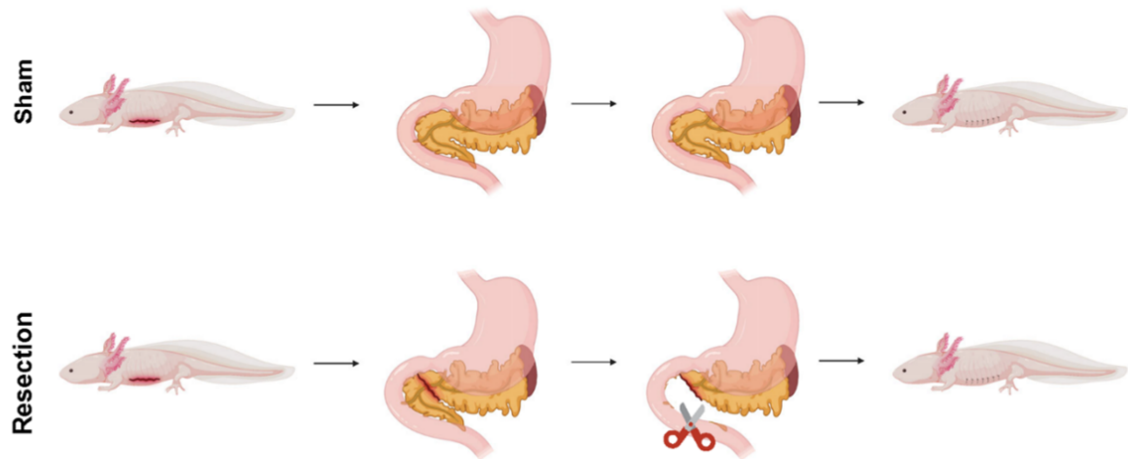


Figure 22. Axolotl Pancreas Resection Experiment Illustrations

Illustration of the pancreas sham surgery (top) and resection surgery (bottom), used for sequencing runs 019 and 024. Sham and resected pancreas RNA was collected at 14 days post injury (dpi) (credit to Connor Powell).

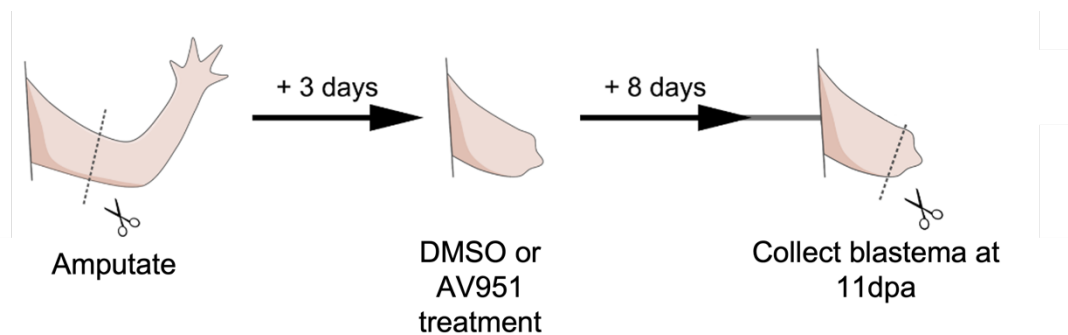


Figure 23. Axolotl VEGF inhibitor (AV951) Experiment Illustrations

Illustration of the VEGF inhibitor experiment, used for sequencing run 018. DMSO or AV951 treated blastemas were collected at 11 days post amputation (dpa) (credit to Aaron M. Savage).

Appendix 3.

Tapestation RNA & DNA Distribution Plots

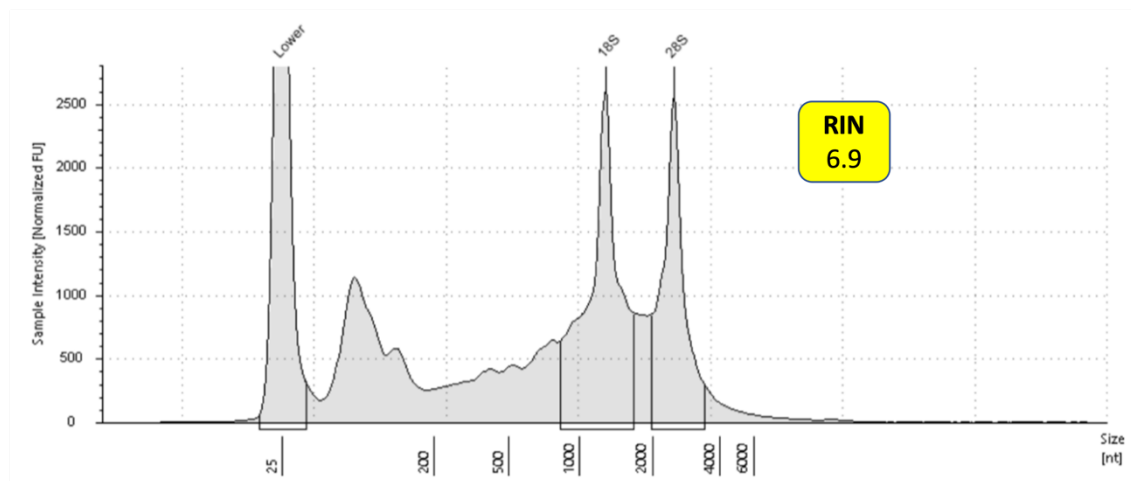


Figure 24. Total RNA Isolation Length Distribution, Pre-Optimization

Representative TapeStation electropherogram depicting poor RNA integrity Number (RIN) score showing RNA degradation pre-optimization of the RNA isolation extraction protocol. The x-axis represents RNA fragment length (nt), while the y-axis represents fluorescence (FU) to quantify abundance of RNA. The two peaks labeled 18S and 28S represent respective ribosomal RNA, while the lower peak is the internal standard (25bp) incorporated into the TapeStation buffer.

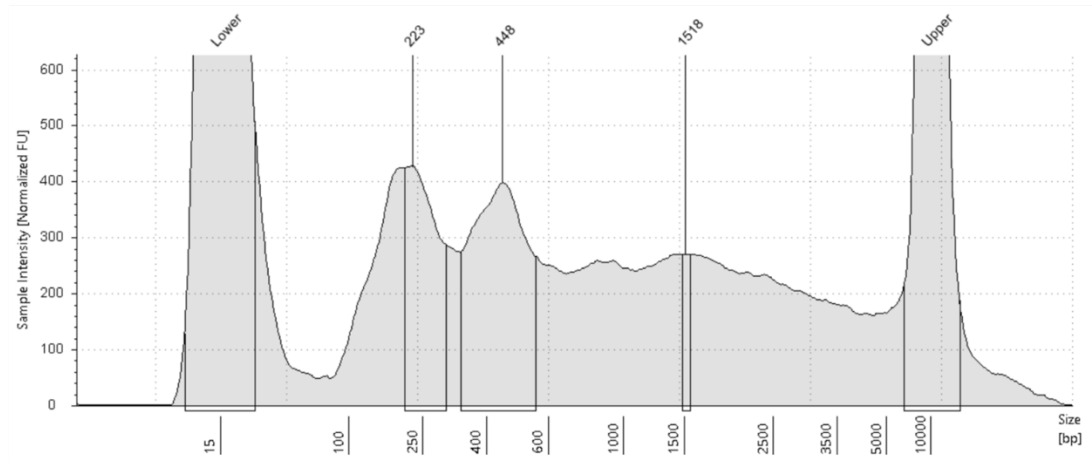


Figure 25. Run 001 Library Prepared cDNA Length Distribution

Representative Tapestation electropherogram depicting sequencing run 001 cDNA library preparation showing two peaks at 223 bp and 448 bp.

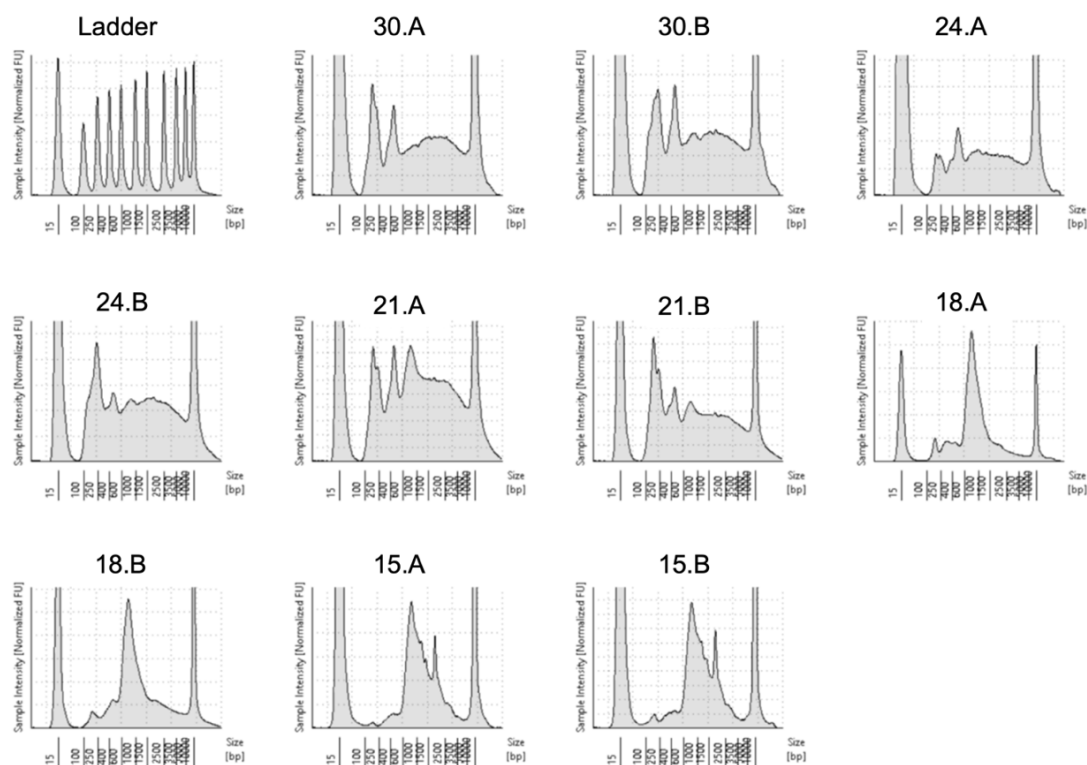


Figure 26. Tapestation cDNA Length Distributions for 15-30 PCR Cycles

Tapestation electropherograms depicting library prepared cDNA at various PCR cycle settings - PCR cycle 30 (30.A-B), 24 (24.A-B), 21 (21.A-B), 18 (18.A-B) and 15 (15.A-B). Each pair utilized duplicate RNA samples as input. The electropherograms showed differences in cDNA length distribution at each cycle setting. PCR cycle 15 was identified to be optimal for long-read sequencing, as it maximizes read length while limiting amplification of less desirable shorter fragments. PCR cycle 13 and 14 (not depicted) resulted in insufficient yield for sequencing

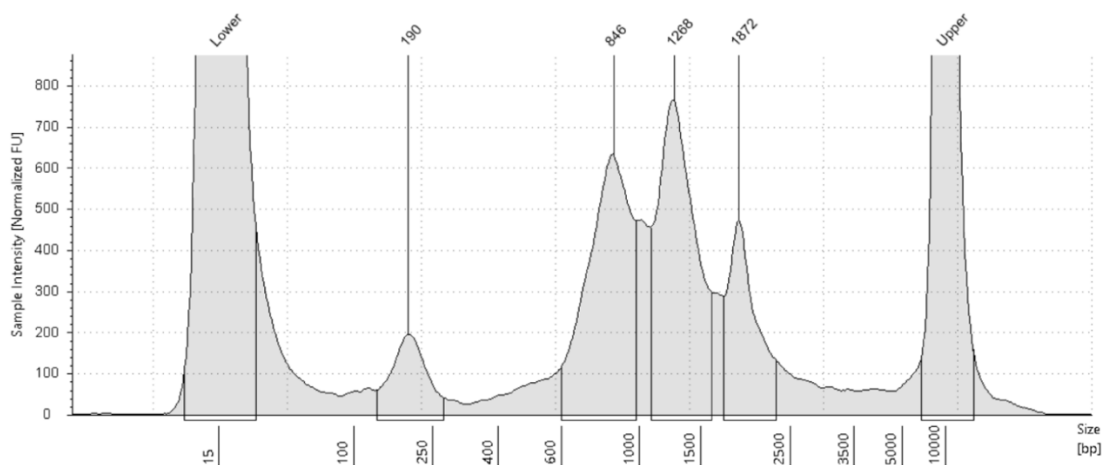


Figure 27. Run 019 Tapestation cDNA Length Distribution

Representative Tapestation electropherogram depicting library prepared cDNA for sequencing run 019. The x-axis represents cDNA base pair length (bp) and y-axis represents fluorescence intensity (FU) to quantify concentration. The distribution of cDNA is significantly different from other runs, as tissue types can have distinct transcript expression profiles.

Appendix 4.

Z-Score Correlation Analysis Plots

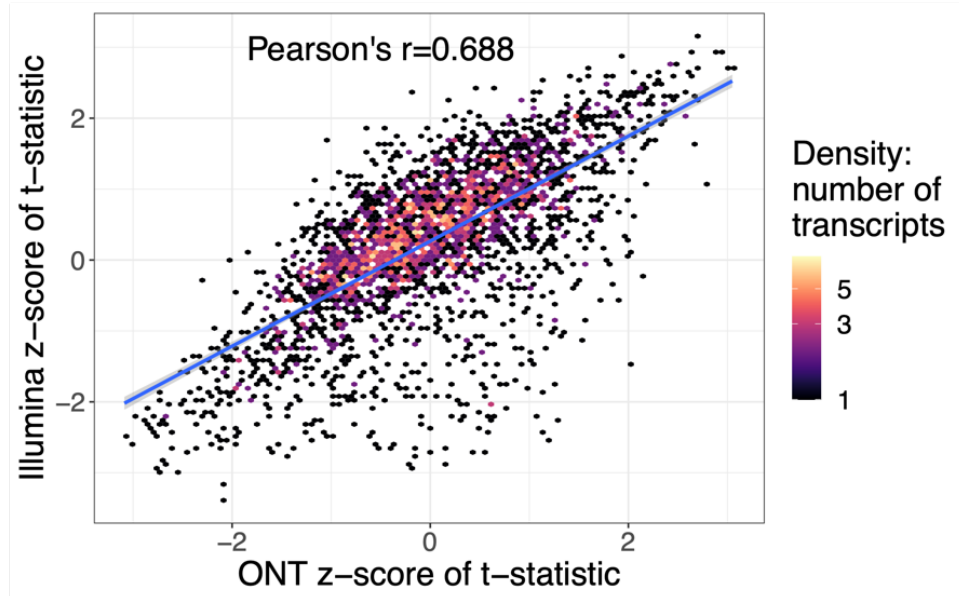


Figure 28. Run 018 Z-Score Correlation Analysis, ONT-seq vs. Illumina-seq

Correlation analysis was performed between run 018 t-statistics of identified genes (x-axis) and a follow-up Illumina short-read sequencing t-statistics of identified genes (y-axis) using the same input cDNA. High correlation (Pearson's $r=0.688$) was observed between the two sequencing runs.

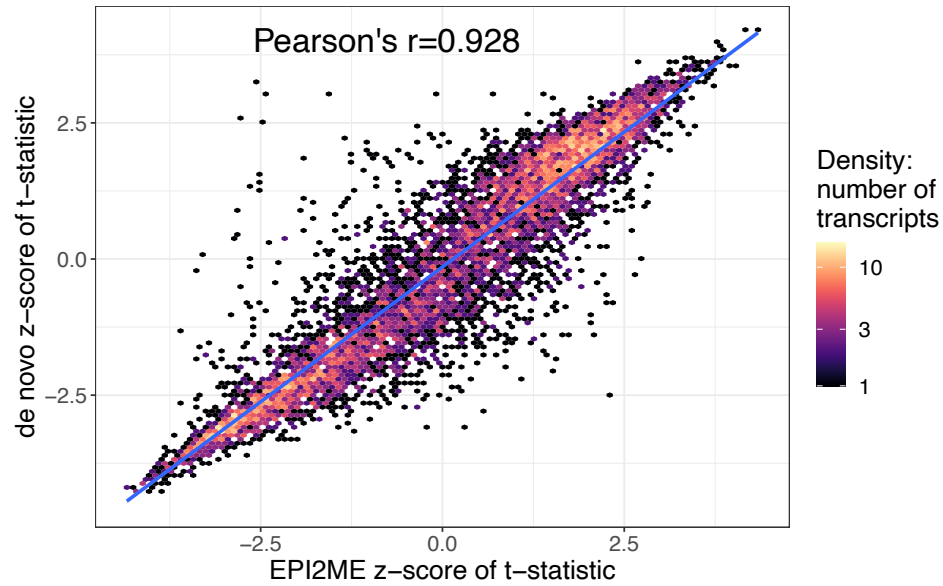


Figure 29. Run 021 Z-Score Correlation Analysis, EPI2ME vs. *De Novo* Pipelines

Correlation analysis plot between t-statistics of identified genes analyzed by EPI2ME Labs wf-transcriptomes pipeline (x-axis) and t-statistics of identified genes analyzed by the de novo workflow (y-axis) using the same raw data from sequencing run 021. High correlation (Pearson's $r=0.928$) was observed between the two sequencing runs.

Appendix 5.

Run 011 Sequencing Performance Analysis

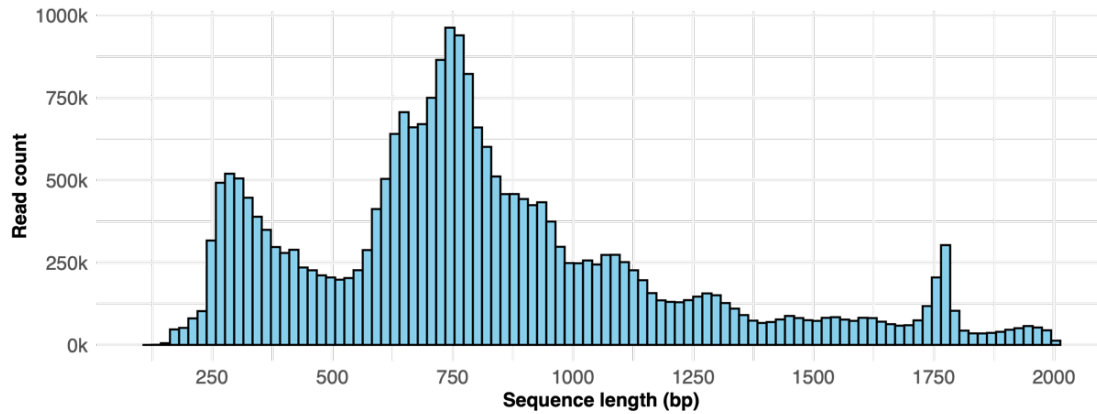


Figure 30. Run 011 Sequencing Length Distribution

Read length distribution histogram generated from raw sequencing run 011 data. The sequencing distribution consists of a less substantial peak at 250 bp than run 001 (Fig. 6).

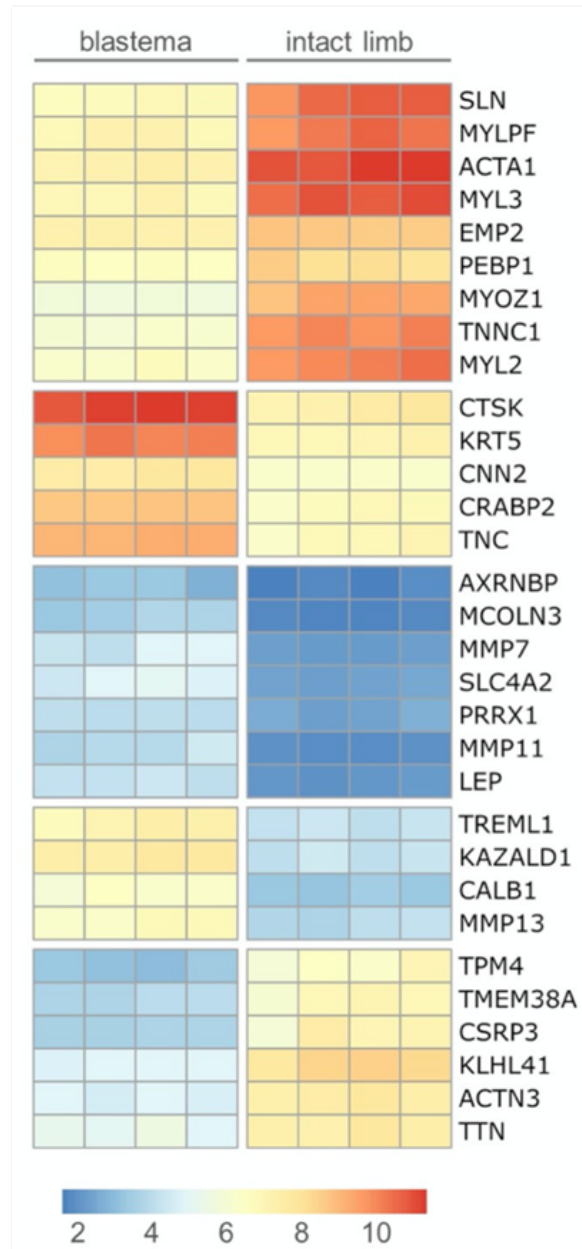


Figure 31. Run 011 Heatmap of Differentially Expressed Genes

Heatmap depicting 31 key DEGs from sequencing run 011. Expected blastema enriched genes like Kazald2 (misabeled in the reference as Kazald1), Axrnbp, and Prrx1 were identified. Transcript expression per replicate (red = high, blue = low) is normalized across the full sequencing run.

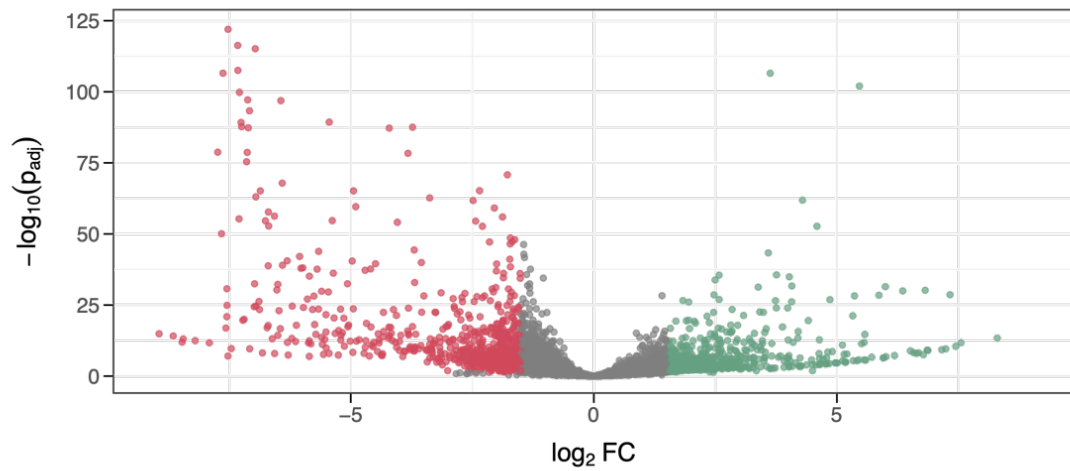


Figure 32. Run 011 Volcano Plot of Differentially Expressed Genes

Volcano plot representing DEGs with genes upregulated (green) and downregulated (red) in the blastema visualized for clarity. Nonsignificant genes ($P_{adj} > 0.05$ and $|\log_2 FC| < 2$) are depicted in grey.

Appendix 6.

MinION vs. PromethION Gene Discovery Plots

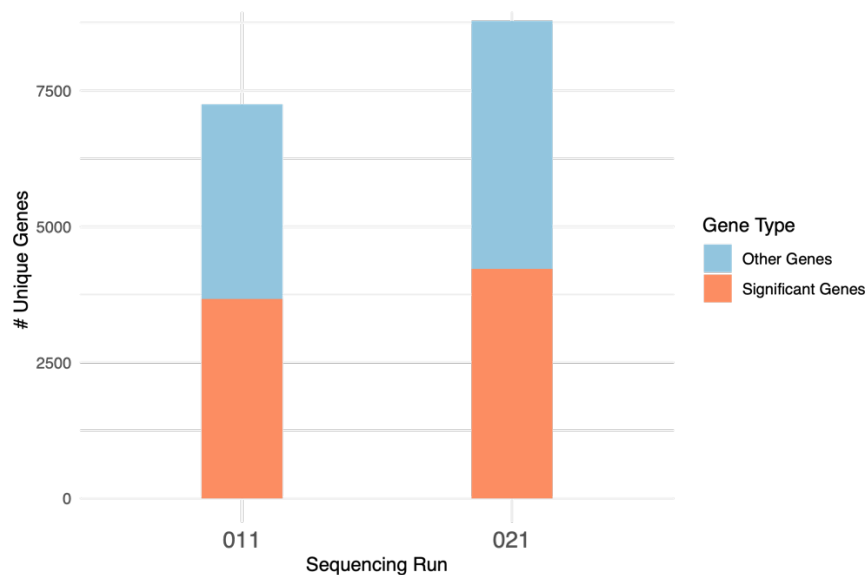


Figure 33. Run 011 vs. 021 Comparison of Gene Discovery

Bar plots depicting the increase in unique genes identified between MinION sequencing run 011 (left) and PromethION sequencing run 021 (right). The y-axis represents number of unique aligned genes. Each bar is divided into genes that are significantly differentially expressed (orange, $P_{adj} < 0.05$), and genes identified as not significant (blue).

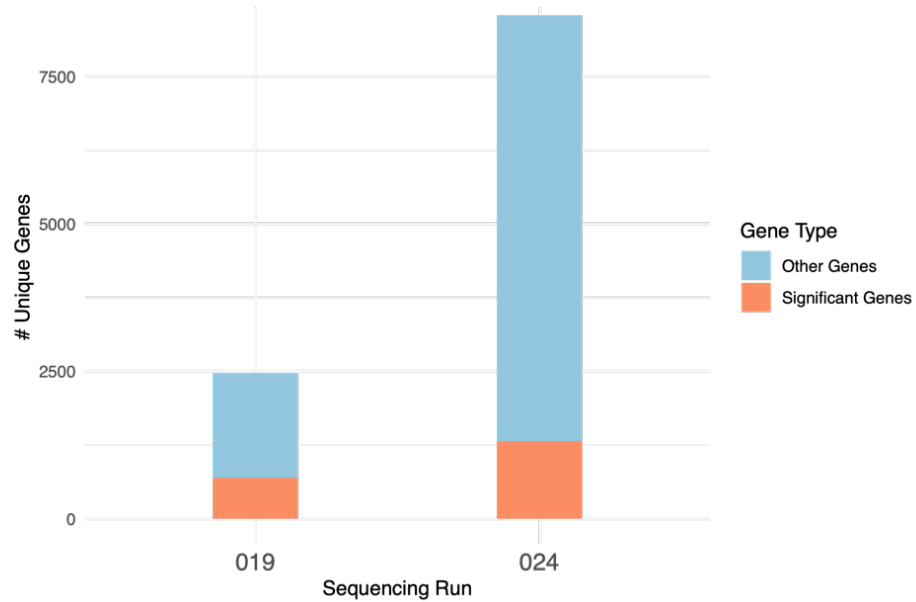


Figure 34. Run 019 vs. 024 Comprison of Gene Discovery

Bar plots depicting the increase in unique genes identified between MinION sequencing run 019 (left) and PromethION sequencing run 024 (right). The y-axis represents number of unique aligned genes. Each bar is divided into genes that are significantly differentially expressed (orange, $P_{adj} < 0.05$), and genes identified as not significant (blue).

Works Cited

- Adewale, B. A. (2020). Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? *Afr J Lab Med*, 9(1), 1340.
<https://doi.org/10.4102/ajlm.v9i1.1340>
- Barrenas, F., Raehtz, K., Xu, C., Law, L., Green, R. R., Silvestri, G., Bosinger, S. E., Nishida, A., Li, Q., Lu, W., Zhang, J., Thomas, M. J., Chang, J., Smith, E., Weiss, J. M., Dawoud, R. A., Richter, G. H., Trichel, A., Ma, D., . . . Gale, M., Jr. (2019). Macrophage-associated wound healing contributes to African green monkey SIV pathogenesis control. *Nat Commun*, 10(1), 5101.
<https://doi.org/10.1038/s41467-019-12987-9>
- Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V., Ramsey, J. M., . . . Schloss, J. A. (2008). The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 26(10), 1146-1153.
<https://doi.org/10.1038/nbt.1495>
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T. H., Davis, F. G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Jr., Peshkin, L., Tabin, C. J., . . . Whited, J. L. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep*, 18(3), 762-776. <https://doi.org/10.1016/j.celrep.2016.12.063>
- Cano-Martínez, A., Vargas-González, A., Guarner-Lans, V., Prado-Zayago, E., León-Oleda, M., & Nieto-Lima, B. (2010). Functional and structural regeneration in the axolotl heart (*Ambystoma mexicanum*) after partial ventricular amputation. *Arch Cardiol Mex*, 80(2), 79-86.
- Cuber, P., Chooneea, D., Geeves, C., Salatino, S., Creedy, T. J., Griffin, C., Sivess, L., Barnes, I., Price, B., & Misra, R. (2023). Comparing the accuracy and efficiency of third generation sequencing technologies, Oxford Nanopore Technologies, and Pacific Biosciences, for DNA barcode sequencing applications. *Ecological Genetics and Genomics*, 28, 100181.
<https://doi.org/https://doi.org/10.1016/j.egg.2023.100181>
- Darnet, S., Dragalzew, A. C., Amaral, D. B., Sousa, J. F., Thompson, A. W., Cass, A. N., Lorena, J., Pires, E. S., Costa, C. M., Sousa, M. P., Fröbisch, N. B., Oliveira, G., Schneider, P. N., Davis, M. C., Braasch, I., & Schneider, I. (2019). Deep evolutionary origin of limb and fin regeneration. *Proceedings of the national*

- academy of sciences*, 116(30), 15106-15115.
<https://doi.org/doi:10.1073/pnas.1900475116>
- Dong, X., Du, M. R. M., Gouil, Q., Tian, L., Jabbari, J. S., Bowden, R., Baldoni, P. L., Chen, Y., Smyth, G. K., Amarasinghe, S. L., Law, C. W., & Ritchie, M. E. (2023). Benchmarking long-read RNA-sequencing analysis tools using in silico mixtures. *Nature Methods*, 20(11), 1810-1821. <https://doi.org/10.1038/s41592-023-02026-3>
- Ebrahim, N., Shakirova, K., & Dashinimaev, E. (2022). PDX1 is the cornerstone of pancreatic β -cell functions and identity. *Front Mol Biosci*, 9, 1091757. <https://doi.org/10.3389/fmolb.2022.1091757>
- Fang, Y., Chen, G., Chen, F., Hu, E., Dong, X., Li, Z., He, L., Sun, Y., Qiu, L., Xu, H., Cai, Z., & Liu, X. (2021). Accurate transcriptome assembly by Nanopore RNA sequencing reveals novel functional transcripts in hepatocellular carcinoma. *Cancer Science*, 112(9), 3555-3568. <https://doi.org/https://doi.org/10.1111/cas.15058>
- Fei, J. F., Schuez, M., Knapp, D., Taniguchi, Y., Drechsel, D. N., & Tanaka, E. M. (2017). Efficient gene knockin in axolotl and its use to test the role of satellite cells in limb regeneration. *Proc Natl Acad Sci U S A*, 114(47), 12501-12506. <https://doi.org/10.1073/pnas.1706855114>
- Fior, J. (2014). Salamander regeneration as a model for developing novel regenerative and anticancer therapies. *J Cancer*, 5(8), 715-719. <https://doi.org/10.7150/jca.9971>
- Gerber, T., Murawala, P., Knapp, D., Masselink, W., Schuez, M., Hermann, S., Gac-Santel, M., Nowoshilow, S., Kageyama, J., Khattak, S., Currie, J. D., Camp, J. G., Tanaka, E. M., & Treutlein, B. (2018). Single-cell analysis uncovers convergence of cell identities during axolotl limb regeneration. *Science*, 362(6413). <https://doi.org/10.1126/science.aag0681>
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333-351. <https://doi.org/10.1038/nrg.2016.49>
- Gu, L. Q., Braha, O., Conlan, S., Cheley, S., & Bayley, H. (1999). Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature*, 398(6729), 686-690. <https://doi.org/10.1038/19491>
- Haas, B. J., & Whited, J. L. (2017). Advances in Decoding Axolotl Limb Regeneration. *Trends Genet*, 33(8), 553-565. <https://doi.org/10.1016/j.tig.2017.05.006>
- Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, 107(1), 1-8. <https://doi.org/https://doi.org/10.1016/j.ygeno.2015.11.003>

- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17, 1-11.
- Jensen, T. B., Giunta, P., Schultz, N. G., Griffiths, J. M., Duerr, T. J., Kyeremateng, Y., Wong, H., Adesina, A., & Monaghan, J. R. (2021). Lung injury in axolotl salamanders induces an organ-wide proliferation response. *Dev Dyn*, 250(6), 866-879. <https://doi.org/10.1002/dvdy.315>
- Jun, E., Oh, J., Lee, S., Jun, H. R., Seo, E. H., Jang, J. Y., & Kim, S. C. (2018). Method Optimization for Extracting High-Quality RNA From the Human Pancreas Tissue. *Transl Oncol*, 11(3), 800-807. <https://doi.org/10.1016/j.tranon.2018.04.004>
- Kasianowicz, J. J., Brandin, E., Branton, D., & Deamer, D. W. (1996). Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the national academy of sciences*, 93(24), 13770-13773. <https://doi.org/doi:10.1073/pnas.93.24.13770>
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., . . . The Wellcome, T. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921. <https://doi.org/10.1038/35057062>
- Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, 15(2), R29. <https://doi.org/10.1186/gb-2014-15-2-r29>
- Leigh, N. D., Dunlap, G. S., Johnson, K., Mariano, R., Oshiro, R., Wong, A. Y., Bryant, D. M., Miller, B. M., Ratner, A., Chen, A., Ye, W. W., Haas, B. J., & Whited, J. L. (2018). Transcriptomic landscape of the blastema niche in regenerating adult axolotl limbs at single-cell resolution. *Nat Commun*, 9(1), 5153. <https://doi.org/10.1038/s41467-018-07604-0>
- Leigh, N. D., Sessa, S., Dragalzew, A. C., Payzin-Dogru, D., Sousa, J. F., Aggouras, A. N., Johnson, K., Dunlap, G. S., Haas, B. J., Levin, M., Schneider, I., & Whited, J. L. (2020). von Willebrand factor D and EGF domains is an evolutionarily conserved and required feature of blastemas capable of multitissue appendage regeneration. *Evol Dev*, 22(4), 297-311. <https://doi.org/10.1111/ede.12332>
- Li, D., Gong, B., Xu, J., Ning, B., & Tong, W. (2021). Impact of Sequencing Depth and Library Preparation on Toxicological Interpretation of RNA-Seq Data in a "Three-Sample" Scenario. *Chem Res Toxicol*, 34(2), 529-540. <https://doi.org/10.1021/acs.chemrestox.0c00368>
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094-3100. <https://doi.org/10.1093/bioinformatics/bty191>

- Li, W. C., Rukstalis, J. M., Nishimura, W., Tchipashvili, V., Habener, J. F., Sharma, A., & Bonner-Weir, S. (2010). Activation of pancreatic-duct-derived progenitor cells during pancreas regeneration in adult rats. *J Cell Sci*, 123(Pt 16), 2792-2802. <https://doi.org/10.1242/jcs.065268>
- Liem, M., Jansen, H. J., Dirks, R. P., Henkel, C. V., van Heusden, G. P. H., Lemmers, R., Omer, T., Shao, S., Punt, P. J., & Spaink, H. P. (2017). De novo whole-genome assembly of a wild type yeast isolate using nanopore sequencing. *F1000Res*, 6, 618. <https://doi.org/10.12688/f1000research.11146.2>
- Lopatriello, G., Maestri, S., Alfano, M., Papa, R., Di Vittori, V., De Antoni, L., Bellucci, E., Pieri, A., Bitocchi, E., Delledonne, M., & Rossato, M. (2023). CRISPR/Cas9-Mediated Enrichment Coupled to Nanopore Sequencing Provides a Valuable Tool for the Precise Reconstruction of Large Genomic Target Regions. *International Journal of Molecular Sciences*, 24(2), 1076. <https://www.mdpi.com/1422-0067/24/2/1076>
- 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. <https://doi.org/10.1186/s13059-014-0550-8>
- Lovely, A. M., Duerr, T. J., Qiu, Q., Galvan, S., Voss, S. R., & Monaghan, J. R. (2022). Wnt Signaling Coordinates the Expression of Limb Patterning Genes During Axolotl Forelimb Development and Regeneration. *Front Cell Dev Biol*, 10, 814250. <https://doi.org/10.3389/fcell.2022.814250>
- Lundstrom, M. S., & Alam, M. A. (2022). Moore's law: The journey ahead. *Science*, 378(6621), 722-723. <https://doi.org/doi:10.1126/science.ade2191>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads [next generation sequencing; small RNA; microRNA; adapter removal]. *2011*, 17(1), 3. <https://doi.org/10.14806/ej.17.1.200>
- McCombie, W. R., & McPherson, J. D. (2019). Future Promises and Concerns of Ubiquitous Next-Generation Sequencing. *Cold Spring Harb Perspect Med*, 9(9). <https://doi.org/10.1101/cshperspect.a025783>
- Moore, G. E. (1965). Gramming more components onto integrated circuits. *Electronics*, 38, 8.
- Motone, K., & Nivala, J. (2023). Not if but when nanopore protein sequencing meets single-cell proteomics. *Nature Methods*, 20(3), 336-338. <https://doi.org/10.1038/s41592-023-01800-7>
- Nowoshilow, S., Schloissnig, S., Fei, J. F., Dahl, A., Pang, A. W. C., Pippel, M., Winkler, S., Hastie, A. R., Young, G., Roscito, J. G., Falcon, F., Knapp, D., Powell, S., Cruz, A., Cao, H., Habermann, B., Hiller, M., Tanaka, E. M., & Myers, E. W. (2018). The axolotl genome and the evolution of key tissue

- formation regulators. *Nature*, 554(7690), 50-55.
<https://doi.org/10.1038/nature25458>
- Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. *J Appl Genet*, 52(4), 413-435.
<https://doi.org/10.1007/s13353-011-0057-x>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417-419. <https://doi.org/10.1038/nmeth.4197>
- Patterson, J., Carpenter, E. J., Zhu, Z., An, D., Liang, X., Geng, C., Drmanac, R., & Wong, G. K.-S. (2019). Impact of sequencing depth and technology on de novo RNA-Seq assembly. *BMC Genomics*, 20(1), 604. <https://doi.org/10.1186/s12864-019-5965-x>
- Pervez, M. T., Hasnain, M. J. U., Abbas, S. H., Moustafa, M. F., Aslam, N., & Shah, S. S. M. (2022). A Comprehensive Review of Performance of Next-Generation Sequencing Platforms. *Biomed Res Int*, 2022, 3457806.
<https://doi.org/10.1155/2022/3457806>
- Philpott, M., Watson, J., Thakurta, A., Brown, T., Brown, T., Oppermann, U., & Cribbs, A. P. (2021). Nanopore sequencing of single-cell transcriptomes with scCOLOR-seq. *Nature Biotechnology*, 39(12), 1517-1520. <https://doi.org/10.1038/s41587-021-00965-w>
- Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, 19(1), 90. <https://doi.org/10.1186/s13059-018-1462-9>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. <https://doi.org/10.1093/bioinformatics/btp616>
- Roy, S., & Gatien, S. (2008). Regeneration in axolotls: a model to aim for! *Experimental Gerontology*, 43(11), 968-973.
<https://doi.org/https://doi.org/10.1016/j.exger.2008.09.003>
- Rutter, L., Moran Lauter, A. N., Graham, M. A., & Cook, D. (2019). Visualization methods for differential expression analysis. *BMC Bioinformatics*, 20(1), 458.
<https://doi.org/10.1186/s12859-019-2968-1>
- Sadasivan, H., Maric, M., Dawson, E., Iyer, V., Israeli, J., & Narayanasamy, S. (2023). Accelerating Minimap2 for Accurate Long Read Alignment on GPUs. *J Biotechnol Biomed*, 6(1), 13-23. <https://doi.org/10.26502/jbb.2642-91280067>
- Sagniez, M., Budhraj, A., Paré, B., Simpson, S. M., Vinet-Ouellette, C., Rozendaal, M., & Smith, M. A. (2024). Assembly Arena: Benchmarking RNA isoform

- reconstruction algorithms for nanopore sequencing. *bioRxiv*, 2024.2003.2021.586080. <https://doi.org/10.1101/2024.03.21.586080>
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12), 5463-5467.
- Schloissnig, S., Kawaguchi, A., Nowoshilow, S., Falcon, F., Otsuki, L., Tardivo, P., Timoshevskaya, N., Keinath, M. C., Smith, J. J., Voss, S. R., & Tanaka, E. M. (2021). The giant axolotl genome uncovers the evolution, scaling, and transcriptional control of complex gene loci. *Proc Natl Acad Sci U S A*, 118(15). <https://doi.org/10.1073/pnas.2017176118>
- Schloss, J. A. (2008). How to get genomes at one ten-thousandth the cost. *Nature Biotechnology*, 26(10), 1113-1115. <https://doi.org/10.1038/nbt1008-1113>
- Shagin, D. A., Lukyanov, K. A., Vagner, L. L., & Matz, M. V. (1999). Regulation of average length of complex PCR product. *Nucleic Acids Research*, 27(18), e23-i-e23-iii. <https://doi.org/10.1093/nar/27.18.e23-i>
- Sørensen, P. L., Dittrich, A. L., & Lauridsen, H. L. (2022). A Novel Animal Model in Diabetes: Regeneration of β cells in the Axolotl Salamander? *The FASEB Journal*, 36(S1). <https://doi.org/10.1096/fasebj.2022.36.S1.R1914>
- Sousounis, K., Bryant, D. M., Martinez Fernandez, J., Eddy, S. S., Tsai, S. L., Gundberg, G. C., Han, J., Courtemanche, K., Levin, M., & Whited, J. L. (2020). Eya2 promotes cell cycle progression by regulating DNA damage response during vertebrate limb regeneration. *Elife*, 9. <https://doi.org/10.7554/eLife.51217>
- Tajer, B., Savage, A. M., & Whited, J. L. (2023). The salamander blastema within the broader context of metazoan regeneration. *Front Cell Dev Biol*, 11, 1206157. <https://doi.org/10.3389/fcell.2023.1206157>
- Tanaka, E. M. (2016). The Molecular and Cellular Choreography of Appendage Regeneration. *Cell*, 165(7), 1598-1608. <https://doi.org/10.1016/j.cell.2016.05.038>
- Totomoch-Serra, A., Marquez, M. F., & Cervantes-Barragán, D. E. (2017). Sanger sequencing as a first-line approach for molecular diagnosis of Andersen-Tawil syndrome. *F1000Res*, 6, 1016. <https://doi.org/10.12688/f1000research.11610.1>
- Van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The third revolution in sequencing technology. *Trends in Genetics*, 34(9), 666-681.
- Vieira, W. A., Wells, K. M., & McCusker, C. D. (2020). Advancements to the Axolotl Model for Regeneration and Aging. *Gerontology*, 66(3), 212-222. <https://doi.org/10.1159/000504294>

- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11), 1348-1365. <https://doi.org/10.1038/s41587-021-01108-x>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10(1), 57-63. <https://doi.org/10.1038/nrg2484>
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738.
- Wei, X., Fu, S., Li, H., Liu, Y., Wang, S., Feng, W., Yang, Y., Liu, X., Zeng, Y.-Y., Cheng, M., Lai, Y., Qiu, X., Wu, L., Zhang, N., Jiang, Y., Xu, J., Su, X., Peng, C., Han, L., . . . Gu, Y. (2022). Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*, 377(6610), eabp9444. <https://doi.org/doi:10.1126/science.abp9444>
- Wetterstrand, K. A. (2024). *DNA sequencing costs: Data from the NHGRI genome sequencing program (GSP)*. National Human Genome Research Institute. <https://www.genome.gov/sequencingcostsdata>
- Wick, R. R., Judd, L. M., & Holt, K. E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20(1), 129. <https://doi.org/10.1186/s13059-019-1727-y>
- Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. <https://ggplot2.tidyverse.org>