



Article

Evaluation of Oxford Nanopore MinION RNA-Seq Performance for Human Primary Cells

Ilaria Massaiu ¹, Paola Songia ¹, Mattia Chiesa ¹ , Vincenza Valerio ^{1,2}, Donato Moschetta ^{1,3}, Valentina Alfieri ¹, Veronika A. Myasoedova ¹ , Michael Schmid ⁴ , Luca Cassetta ⁵, Gualtiero I. Colombo ¹ , Yuri D'Alessandra ^{1,†} and Paolo Poggio ^{1,*}

- ¹ Centro Cardiologico Monzino IRCCS, 20131 Milan, Italy; ilaria.massaiu@cardiologicomonzino.it (I.M.); paola.songia@cardiologicomonzino.it (P.S.); mattia.chiesa@cardiologicomonzino.it (M.C.); vincenza.valerio@cardiologicomonzino.it (V.V.); donato.moschetta@cardiologicomonzino.it (D.M.); valentina.alfieri@cardiologicomonzino.it (V.A.); veronika.myasoedova@cardiologicomonzino.it (V.A.M.); gualtiero.colombo@cardiologicomonzino.it (G.I.C.); yuri.dalessandra@cardiologicomonzino.it (Y.D.)
- ² Dipartimento di Medicina Clinica e Chirurgia, Università degli Studi di Napoli Federico II, 80131 Napoli, Italy
- ³ Dipartimento di Scienze Farmacologiche e Biomolecolari, Università degli Studi di Milano, 20133 Milano, Italy
- ⁴ Genexa AG, Dienerstrasse 7, CH-8004 Zürich, Switzerland; michael.schmid@genexa.ch
- ⁵ The Queen's Medical Research Council Centre for Reproductive Health, University of Edinburgh, Edinburgh EH16 4TJ, UK; luca.cassetta@ed.ac.uk
- * Correspondence: paolo.poggio@cardiologicomonzino.it
- † These authors contributed equally.



Citation: Massaiu, I.; Songia, P.; Chiesa, M.; Valerio, V.; Moschetta, D.; Alfieri, V.; Myasoedova, V.A.; Schmid, M.; Cassetta, L.; Colombo, G.I.; et al. Evaluation of Oxford Nanopore MinION RNA-Seq Performance for Human Primary Cells. *Int. J. Mol. Sci.* **2021**, *22*, 6317. <https://doi.org/10.3390/ijms22126317>

Academic Editors: Michiaki Hamada and James K. Bashkin

Received: 20 April 2021

Accepted: 9 June 2021

Published: 12 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Transcript sequencing is a crucial tool for gaining a deep understanding of biological processes in diagnostic and clinical medicine. Given their potential to study novel complex eukaryotic transcriptomes, long-read sequencing technologies are able to overcome some limitations of short-read RNA-Seq approaches. Oxford Nanopore Technologies (ONT) offers the ability to generate long-read sequencing data in real time via portable protein nanopore USB devices. This work aimed to provide the user with the number of reads that should be sequenced, through the ONT MinION platform, to reach the desired accuracy level for a human cell RNA study. We sequenced three cDNA libraries prepared from poly-adenosine RNA of human primary cardiac fibroblasts. Since the runs were comparable, they were combined in a total dataset of 48 million reads. Synthetic datasets with different sizes were generated starting from the total and analyzed in terms of the number of identified genes and their expression levels. As expected, an improved sensitivity was obtained, increasing the sequencing depth, particularly for the non-coding genes. The reliability of expression levels was assayed by (i) comparison with PCR quantifications of selected genes and (ii) by the implementation of a user-friendly multiplexing method in a single run.

Keywords: nanopore technology; RNA-seq; MinION; sequencing depth; multiplexed samples; barcoding; SQK-PCS109

1. Introduction

Knowledge about the human genome plays a crucial role in modern medicine [1,2]. In-depth analysis of tissue transcriptomes, leveraging the power of genome-wide gene expression investigation, is increasingly used for clinical decisions in the new era of precision medicine [3]. RNA sequencing (RNA-Seq), based on next-generation sequencing technologies, is the current method of choice for quantifying gene expression at the genomic level with higher depth and accuracy than probe-based microarray approaches [4].

Recently, third-generation sequencing (TGS) approaches have been designed to overcome some of the limitations of the second-generation sequencing (SGS) technologies, which rely on short-read length analysis [5]. Despite massive throughput, the use of SGS for de novo transcriptome assembly and analysis of large structural variations remains

challenging [6,7]. TGS technologies have the advantage of capturing many full-length single-molecule transcripts, avoiding the problematic errors of assembly processes, and reducing the time-to-results window [8].

The MinION sequencer, released by Oxford Nanopore Technologies (ONT), is a new, low-cost, handheld device that processes thousands of long reads in parallel, using 50 ng starting material [9,10]. The process is based on the passage of DNA/RNA strands through a biological nanopore, generating base-specific changes in electrical conductivity and leading to the identification of specific sequences using a neural network. Although Nanopore long-read sequencing suffers from some limitations compared to short-read methods, such as the high error rate during the base-calling step and the sensitivity to RNA degradation, significant advantages in identifying novel RNA molecules and complex isoforms have been widely demonstrated in previous works [6,11–15]. In particular, the long-read sequencing technology was more efficient in quantifying long non-coding RNAs [16]. Moreover, thanks to its ability to perform rapid long-read sequencing analysis requiring minimal supporting laboratory infrastructure or technical expertise, MinION has been widely used for the diagnosis of viral disease [17–19]. It played a crucial role in supporting the COVID-19 pandemic, particularly in isolated or resource-poor settings [20–22].

The turnaround time and cost for ONT RNA-Seq can be further reduced by sequencing multiple samples on a single run [19,23–25]. Wick et al. [23] sequenced 12 bacterial DNA isolates simultaneously on a single MinION flow cell using the ONT native barcoding kit. The long reads were then combined in a hybrid assembly with Illumina data to fully resolve the bacterial genome. Recently, King et al. [24] proposed a rapid workflow for multiplexed sequencing of influenza A viruses using the ONT technology for real-time analysis, in combination with a one-step RT-PCR and the Rapid Barcoding Kit.

As with any RNA-Seq study, the read depth is one of the most important factors to reach the desired level of accuracy and sensitivity, in addition to the number of biological replicates. Indeed, the read depth changes in accordance with the purpose of the study, requiring a significant increase when low-expression genes are to be evaluated. Numerous investigations regarding the performance of short-read RNA-Seq with varying reads depth have been published [26–30]. Conversely, a very limited number of studies focusing on the real capabilities of the MinION sequencing platform are available [6,15,31–35]. In particular, an RNA-Seq evaluation with the latest PCR-cDNA sequencing kit (i.e., SQK-PCS109) is missing. This kit is highly recommended for users who have a limited amount of input material, want to optimize their sequencing experiment for throughput, would like to identify and quantify full-length transcripts, and are interested in differential gene expression.

Here, we have qualitatively and quantitatively analyzed the performance of MinION through the sequencing of human primary cardiac fibroblasts, both in terms of the number of detected genes and corresponding quantifications, by sequentially changing the number of reads, employing the SQK-PCS109 kit. Our goal was to estimate, according to the target accuracy of a given study, a suitable sequencing depth to obtain a reliable number of detected genes as well as their expression levels from human primary cells.

2. Results

The workflow diagram of the study is reported in Figure 1.

2.1. Sequencing Performance Changing the Number of Reads

Three cDNA libraries, prepared from poly-adenosine (poly-A) RNA of three human primary cardiac fibroblast samples, were sequenced by an ONT MinION sequencer using R9.4 flow cells. The comparison of the gene expression levels, expressed as $\log_2(\text{CPM})$, showed a strong correlation among these three independent replicates (Pearson's correlation coefficient (r_p) equal to 0.98; Figure 2).

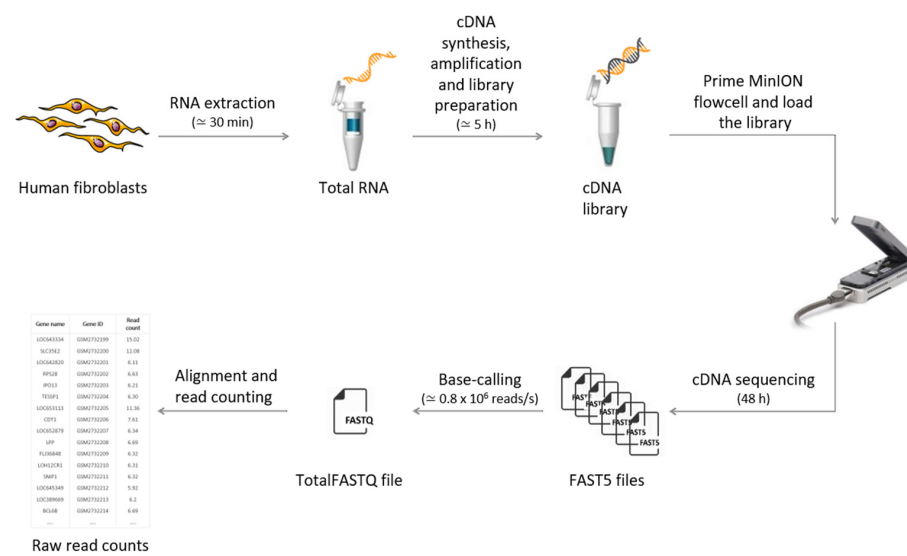


Figure 1. Study workflow. RNA was extracted from human cardiac fibroblast cells. Then, cDNA library preparation (including size, quantity, and quality analysis), base calling, read mapping, and counting were performed.

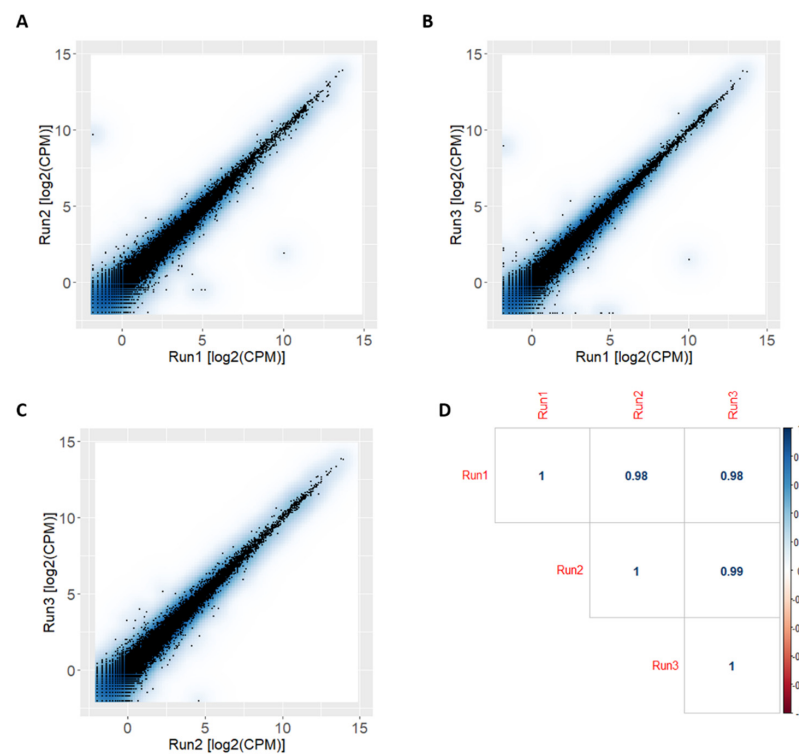


Figure 2. Correlation analysis comparing the three libraries. (A) Scatter plot of $\log_2(\text{CPM})$ values from run1 vs. run2. (B) Scatter plot of $\log_2(\text{CPM})$ values from run1 vs. run3. (C) Scatter plot of $\log_2(\text{CPM})$ values quantified from run2 vs. run3. (D) Correlation matrix with Pearson correlation coefficients.

Thus, data obtained from these runs were combined to obtain a large dataset (hereinafter DS100) for the evaluation of ONT performance as a function of read number. DS100 was composed of more than 48 million reads (about 23 gigabases), with an average length of 483.2 base pairs (bp) that passed the quality filter. We used this total dataset to generate progressively fractional synthetic subsets (90% to 5%, i.e., DS90 to DS5). The total and mapped reads are reported in Table 1 for each dataset.

Table 1. The numbers of total and mapped reads are reported for each subset (DS90–DS5) and the total dataset (DS100). The numbers of mapped reads for the subsets are the average among 10 randomly picked replicated subsets with the same size ($CV < 0.8\%$).

Dataset	Avg. Total Reads	Avg. Mapped Reads
DS100	48,495,343	28,239,507
DS90	43,645,809	25,415,207
DS80	38,796,274	22,592,768
DS70	33,946,740	19,768,060
DS60	29,097,206	16,943,277
DS50	24,247,672	14,098,354
DS40	19,398,137	11,302,374
DS30	14,548,603	8,461,139
DS20	9,699,069	5,368,370
DS10	4,849,534	2,676,107
DS5	2,424,767	1,340,245

We obtained similar length distributions of reads among the different subsets (Supplementary Figure S1 and Supplementary Table S1). Moreover, a strong agreement was found between the average gene expression levels obtained from the eleven datasets (DS100–DS5), even when comparing the two extreme datasets, DS5 and DS100 ($r_p = 0.97$ and $p < 10^{-5}$; Supplementary Figure S2). The reliability of the results was evaluated by comparing the quantifications obtained from DS5, DS30, and DS100 with the dCt values of ten selected genes (IL4, MALAT1, COL1A1, DCN, MMP2, H19, CAT, SOD3, BCL2, and BMP2) detected by qPCR. There was a significant correlation between each dataset quantification and the qPCR values ($r_p \geq 0.8$ and $p < 0.01$; Figure 3). Of note, the DS30 subset (about 14 million reads) is equivalent to a single flow-cell run. Indeed, the SQK-PCS109 kit can generate 10–15 million reads in 48 h per flow cell.

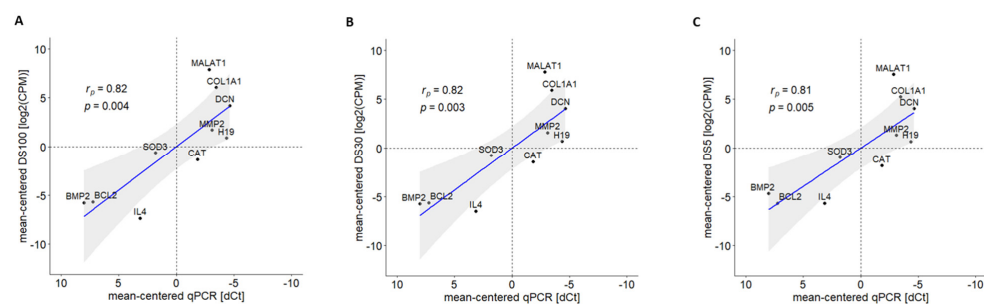


Figure 3. Correlation analysis of gene expression levels obtained from qRT-PCR and ONT sequencing using (A) DS100, (B) DS30, and (C) DS5. Interleukin 4 (IL4), metastasis associated lung adenocarcinoma transcript 1 (MALAT1), collagen type I alpha 1 chain (COL1A1), decorin (DCN), matrix metalloproteinase 2 (MMP2), h19 imprinted maternally expressed transcript (H19), catalase (CAT), superoxide dismutase 3 (SOD3), BCL2 apoptosis regulator (BCL2), and bone morpho-genetic protein 2 (BMP2) genes were selected, and qPCR experiments were performed. Pearson correlation coefficients (r_p) between $\log_2(\text{CPM})$ and dCt were computed and reported with the respective p -value (p).

The sequencing performance, when changing the number of reads, was assessed by evaluating the number of detected genes as well as the corresponding quantification levels based on different read depths. Moreover, we investigated coding and non-coding genes considering the Ensembl 97 database (GRCh38.p12). A total of 21,816 expressed genes were detected using DS100, and the detection sensitivity decreased after reducing the size of the subsets (from DS90 to DS5; Figure 4A, black line, and Supplementary Table S2). In particular, we identified 17,633 genes in DS30 and 12,114 in DS5. The detection trend revealed a significant improvement from DS5 up to DS30. Indeed, the number of expressed genes detected by DS30 was about 50% higher than that for DS5. Instead, compared with

DS30, an increase of about 25% was reached using the largest dataset, DS100. As expected, the genes with low expression levels were challenging to detect (Supplementary Figure S3).

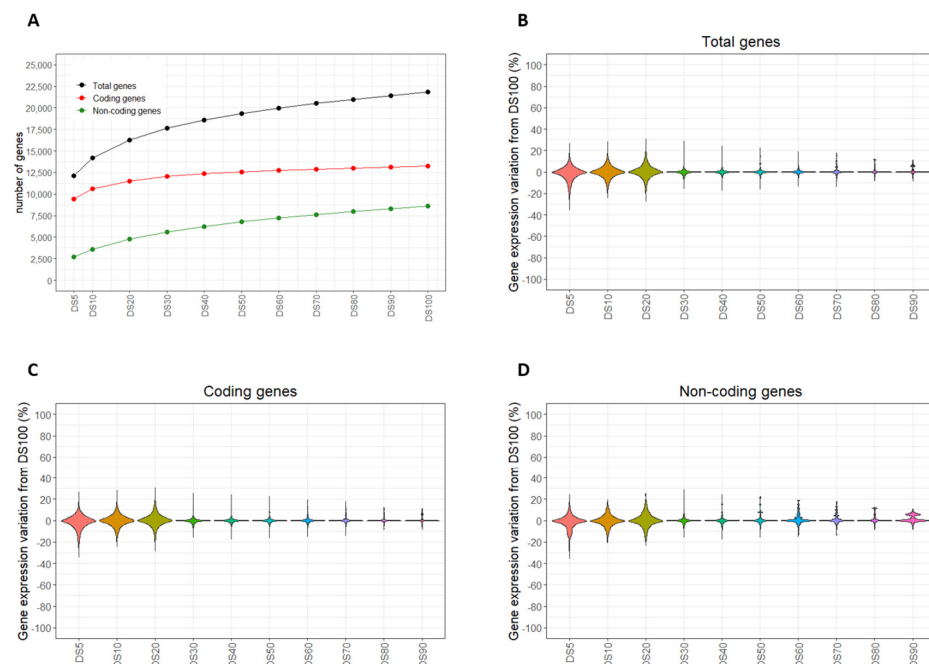


Figure 4. Sequencing results of total, coding, and non-coding genes as a function of the read depth. (A) Number of total, coding, and non-coding genes in each dataset (black, red, and green lines, respectively). Distributions of gene expression variation of (B) total, (C) coding, and (D) non-coding genes in each subset.

Focusing on the biotype classification, we observed that >60% of the genes were annotated as protein-coding (Figure 4A, red line, and Supplementary Table S2) and the remaining part was composed of non-coding genes (Figure 4A, green line, and Supplementary Table S2). The improvement in the detection of coding genes reached a plateau after DS20. In particular, from DS5 to DS20, we observed an increment of about 22% in the number of coding genes. As for the non-coding genes, an increment of >100% was achieved in DS30 compared with DS5, and 53% going from DS30 to DS100.

The gene quantifications obtained from each subset (from DS90 to DS5) were compared to the results of DS100 (as reference) to compute the gene expression variation (%GEV). A substantial reduction in the %GEV was shown considering the subsets larger than DS20 (Figure 4B and Supplementary Table S3). The %GEV values obtained for coding and non-coding genes had similar characteristics (Figure 4C,D; Supplementary Tables S4 and S5). However, in the latter, we observed a slightly higher %GEV.

Finally, the %GEV values at different ranges were investigated in the two biotypes. In particular, coding and non-coding genes were grouped into high, mid-high, mid-low, and low expression subsets and analyzed. An overall negative correlation between the %GEV and the expression class was observed for all the subsets (Supplementary Figures S4 and S5).

2.2. Multiplex Sequencing

To further validate the results obtained from the synthetic subsets, we sequenced one sample with six different barcodes in a single run and compared it to DS5 (about two million reads). More than 10 million reads passed the quality filter (99.9% of the overall passed reads). These reads were then assigned to each barcode and six different datasets were generated. An average of 1.7 million high-quality reads made up each dataset, of which about 60% were mapped to the reference genome, and more than 10,000 genes were

identified in each barcoded sample (5.26% CV; Table 2 and Figure 5), similar to the DS5 synthetic subset.

Table 2. Sequencing data of each barcoded sample.

Dataset	Total Reads	Mapped Reads	Number of Total Genes
Barcode1	1,293,001	878,970	10,869
Barcode2	2,029,051	1,382,710	12,006
Barcode3	1,401,956	909,050	10,917
Barcode4	2,640,182	1,331,922	11,690
Barcode5	1,871,851	784,207	11,442
Barcode6	1,196,100	821,091	10,418

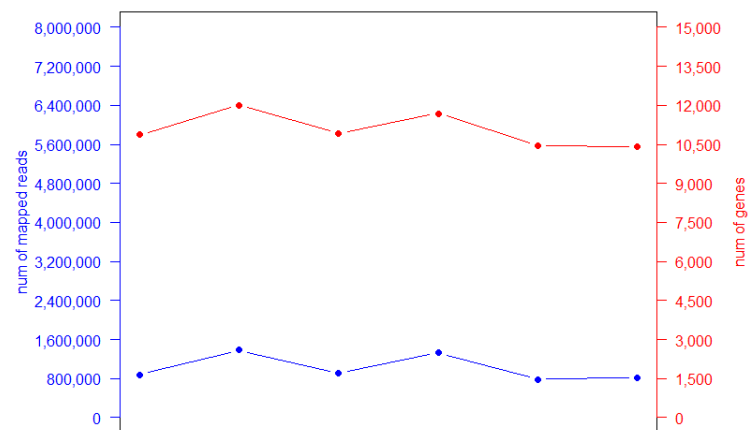


Figure 5. Sequencing results of each barcoded sample. Number of mapped reads (**red**) and identified genes (**blue**) obtained for each barcoded sample.

Furthermore, the expression levels of the barcoded samples were compared to the values quantified in the DS5 subset (Figure 6). The high correlation ($r_p \geq 0.9$ and $p < 10^{-5}$) between each barcoded sample and the DS5 subset revealed the reproducibility and high quality of the obtained results.

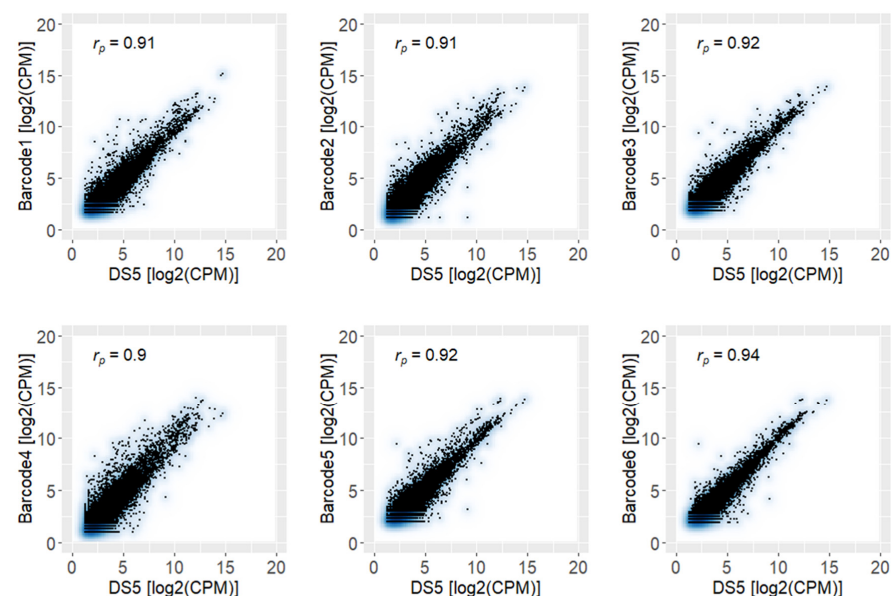


Figure 6. Correlation analysis of results obtained from the six barcoded samples with the smallest subset DS5.

3. Discussion

The ability to sequence long DNA fragments provided by the ONT MinION system offers easier assembly of complex genomes than short-read methods do, but its high error rate is still an obstacle [6,7]. Thanks to the decreased turnaround time at a low cost compared to short-read sequencing, and the continuous performance improvements, also including the option to sequence multiple samples in a single run [19,23,24], this pocket-sized device is widely used both for scientific research and clinical applications [36]. Different transcriptomic studies exploited the ability of this technology to uncover the diversity of alternative splicing isoforms and their expression levels [12,14,15]. Using MinION sequencing, Bolisetty et al. [37] identified over 7000 isoforms for *Dscam1*, the most complicated alternatively spliced gene known in nature, with an average identity of full-length alignments > 90%, by MinION sequencing. Moreover, thanks to its long-read capability, *de novo* sequencing of microbial, viral, and eukaryotic whole genomes is more easily obtainable [18,38,39]. For these reasons, MinION has proven to be a valuable support for the ongoing worldwide pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [20]. Its capability to perform rapid long-read sequencing analysis, with flexible scalability and accurate consensus-level sequence determination [21], offered key knowledge of virus transmission and evolution as well as for vaccine development [22].

Focusing on RNA studies, to date, ONT sequencing has achieved relevant results in terms of uncovered transcripts, quantification of expression levels, and differential expression analysis, comparable with Illumina technology [6,40]. One of the most important factors for the proper design of RNA-Seq experiments is the sequencing depth. This parameter represents the number of reads collected during each run for a given sample, and in general, its increment leads to an improvement of the sequencing results. However, a unique optimal number of required reads cannot be claimed, and a high depth is normally required to study novel or less abundant transcripts. Nonetheless, a higher depth of sequencing inevitably involves an increment in costs.

Identification of the optimal read depth in function of the aim of the experiments and the complexity of the target transcriptome is a crucial aspect [41]. Indeed, several studies on RNA-Seq performance that involved changing the depth were carried out for short-read technologies [26–30]. Conversely, few references only for the earlier versions of ONT technology are present in the literature, leaving the experimental design somewhat sketchy [6,15,34,35]. Furthermore, these studies were not focused on the evaluation of the optimal read depth, but they were performed to compare this new technology with Illumina platforms. Thus, using the latest available PCR-cDNA kit (i.e., SQK-PCS109), we assessed the ability of the ONT MinION sequencing platform to identify and accurately quantify annotated genes by changing the number of reads, in human primary cells.

With our approach (i.e., poly-A RNA-Seq), more than 12,000 genes were detected with about two million reads, where about 22% belonged to the non-coding biotype. Signs of saturation were obtained at about 14 million reads, representing the typical amount of data achieved in a single run. Similar outcomes were achieved in terms of quantifications, where all the evaluated synthetic datasets had a comparable gene expression. Taking into account the biotype classification, we confirmed that an accurate sequencing of non-coding genes is particularly challenging, since they are typically expressed in low levels [42].

In conclusion, our study is the first, to the best of our knowledge, to show how many genes can be accurately identified and quantified as a function of sequencing depth employing the latest PCR-cDNA ONT kit. We intended to provide new users approaching ONT RNA-Seq a guideline on the optimal read depth to be reached, obtaining a good compromise between accuracy of results, costs, and processing time.

That being said, our study has some limitations to be pointed out. The aim of our study was to analyze the performance of MinION RNA-Seq when varying the number of reads; thus, we did not compare our results with any short-read technologies. Additionally, our data were generated employing human primary cardiac fibroblasts; thus, the number of genes related to the identified depth may not be applicable to other cell types.

4. Materials and Methods

4.1. Sample and Library Preparation

Valve interstitial cells (VICs) were isolated from three human stenotic valves, and RNA was extracted using the Total RNA Purification Plus Kit (Norgen Biotek, Thorold, ON, Canada), according to the manufacturer's instructions. We pooled samples and prepared three cDNA libraries following the recommendations of the Nanopore cDNA-Seq protocol for the SQK-PCS109 kit. Briefly, we employed RT primers to convert only poly-adenylated RNA into cDNA. For the multiplex run, we used six different ad hoc designed barcoded sequences (Supplementary Table S6). cDNA synthesis was performed using 50 ng of total RNA per sample. RT and strand-switching primers were provided by ONT with the SQK-PCS109 kit. Following RT, PCR amplification was performed using the LongAmp Taq 2X Master Mix (New England Biolabs, Ipswich, MA, USA) and the following cycling conditions: 1 cycle (95 °C for 30 s), 18 cycles (95 °C for 15 s, 62 °C for 15 s, and 65 °C for 3 min), and 1 last cycle (65 °C for 15 min). PCR products were purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). The cDNA sequencing libraries were prepared using a total of 200 fmol of cDNA each.

4.2. MinION Sequencing

Nanopore libraries were sequenced using a MinION Mk1B sequencing device with R9.4 flow cells. Sequencing was controlled and data were generated using ONT MinKNOW software (v3.4.12). Runs were terminated after 48 h and FAST5 files were generated.

4.3. Data Processing

DNA bases were called from FAST5 files using ONT Guppy GPU (v3.4.5) in high accuracy mode [43]. Reads with an average Phred quality score, which measures the confidence based on the estimated error rate, lower than 7 were discarded. The first three runs were combined in a single dataset, named DS100, and used as the reference for studying the sequencing performance. Then, 10 additional datasets were generated by randomly sampling 90% (DS90), 80% (DS80), 70% (DS70), 60% (DS60), 50% (DS50), 40% (DS40), 30% (DS30), 20% (DS20), 10% (DS10), and 5% (DS5) of DS100. To assess the reliability of the sampling, this procedure was implemented 10 times for each dataset.

4.4. Bioinformatic Analysis

Reads were aligned to the 22 diploid chromosomes of the GRCh38 human genome reference with minimap2 (v2.1, default parameters except for -ax splice) [44]. SAM-to-BAM format conversion as well as an assessment of the alignment quality were performed using Samtools (v 1.10) [45]. The FeatureCounts software (v2.0.0) [46], included in the Subread package, was used to count the mapped reads. Finally, the expression of each gene was reported as counts per million transformed in logarithmic scale ($\log_2(\text{CPM})$) [47]. For the multiplex run, the quality-checked reads were de-multiplexed and trimmed for barcodes using the Cutadapt function (v1.15) [48], before the alignment and counting procedures. Genes with a read count greater than 3 were deemed as expressed.

4.5. Performance Evaluation

Computational analyses were implemented using the R software environment (v 3.6.0) [49]. Pearson's correlation coefficients (r_p) were computed to check the reproducibility of the quantifications. The average number of genes detected and the coefficient of variation (CV) between the replicates across each subset were computed to evaluate the dispersion generated by subsetting. Correlations between each of the subsets and with the entire dataset were assessed. The biotype was assigned based on the Ensembl 97 database (GRCh38.p12). Genes were grouped into protein-coding and non-coding, the latter including long non-coding RNA (lncRNA), non-coding RNA (ncRNA), and pseudogenes, following the Ensembl Genome Browser annotation (<https://www.ensembl.org/info/genome/genebuild/biotypes.html>, accessed date 28 January 2021).

To evaluate the results as a function of read depth, we compared the number and expression level of genes detected among each subset with respect to DS100. These analyses were repeated for coding and non-coding genes separately.

The gene expression variations (%GEV) of the $\log_2(\text{CPM})$ values obtained in the subsets (from DS90 to DS5) were compared to that calculated in the total dataset (DS100) as follows:

$$\%GEV_{DSi,j} = (\log_2(\text{CPM}_{TOT,j}) - \log_2(\text{CPM}_{DSi,j}) / \log_2(\text{CPM}_{TOT,j})) \times 100 \quad (1)$$

where $\%GEV_{DSi,j}$ is the gene expression variation (expressed as a percentage) relative to the i -th subset (from DS90 to DS5) and the j -th gene, $\log_2(\text{CPM}_{TOT,j})$ is the expression value obtained with the total dataset (DS100) for the j -th gene, and $\log_2(\text{CPM}_{DSi,j})$ is the expression value using the i -th subset (from DS90 to DS5) for the j -th gene. These values were computed only for the genes identified at least from the 80% of replicates of the i -th subset.

The variation distribution was depicted by violin plots. In addition, we repeated the same analysis after dividing the detected coding and non-coding genes into quartiles, which represented high (Q4), medium-high (Q3), medium-low (Q2), and low expression (Q1) transcripts.

4.6. Quantitative Real-Time PCR

We tested the robustness of our results correlating gene expression levels obtained from the DS100, DS50, and DS5 datasets with values detected by qPCR, performed on an ABI Prism 7900 HT (Applied Biosystems, Foster City, CA, USA) with SYBR Green dye (New England BioLabs, Ipswich, MA, USA), according to the manufacturers' instructions. To consider expression levels spread over a broad range, we selected glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as the reference gene and metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), collagen type I alpha 1 chain (COL1A1), decorin (DCN), matrix metalloproteinase 2 (MMP2), H19 imprinted maternally expressed transcript (H19), catalase (CAT), superoxide dismutase 3 (SOD3), BCL2 apoptosis regulator (BCL2), bone morphogenetic protein 2 (BMP2), and interleukin 4 (IL4) as the target genes to be evaluated (Supplementary Table S7). For each gene, the cycle threshold (Ct) value was determined and the dCt value was calculated (target Ct–GAPDH Ct).

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22126317/s1>. Table S1: Average length and the coefficient of variation (CV) among the 10 replicates are reported for each dataset. Table S2: Number of total, coding and, non-coding genes identified from each dataset. Table S3: Quantile values of gene expression variations of each subset respect to DS100 considering total genes. Table S4: Quantile values of gene expression variations of each subset respect to DS100 considering coding genes. Table S5: Quantile values of gene expression variations of each subset respect to DS100 considering non-coding genes. Table S6: Barcoded sequences, Table S7: qPCR primers. Figure S1: Length distributions of sequenced reads composing (A) the total dataset DS100 and (B) each replicate of DS5 (as an example). Figure S2: Correlation analysis of the gene expression values obtained from the different datasets. Figure S3: Variation of gene detection among the eleven datasets (DS5–DS100), in terms of percentage, in function of the average expression levels ($\log_2(\text{CPM})$). Figure S4: Distributions of gene expression variation for coding genes with high, mid-high, mid-low, and low expression. Figure S5: Distributions of gene expression variation for non-coding genes with high, mid-high, mid-low, and low expression.

Author Contributions: P.P. and Y.D. conceived the study. V.A.M. enrolled patients and collected the informed consent and specimens. P.S., V.V., D.M., and V.A. isolated the human primary cells, extracted the RNA, prepared the cDNA-Seq libraries, and performed qPCR analysis. I.M., Y.D., and M.S. designed the barcoded sequences. I.M. performed the in silico analysis. M.C. supervised the in silico analysis. I.M. and P.P. drafted the manuscript. P.S., M.C., V.V., D.M., V.A., V.A.M., M.S., L.C., G.I.C., and Y.D. substantially revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Italian Ministry of Health funds (Ricerca Finalizzata: GR-2018-12366423; ERA-CVD: PICASSO-JTC-2018-042) and by Fondazione Gigi e Pupa Ferrari ONLUS (FPF-14).

Institutional Review Board Statement: The study protocol was approved by the Institutional Review Board and by the Ethical Committee of Centro Cardiologico Monzino IRCCS, following the principles of the Declaration of Helsinki (1964).

Informed Consent Statement: Written informed consent to participate in this study was obtained from patients undergoing aortic valve replacement due to aortic stenosis.

Data Availability Statement: The data presented in this study and R code are openly available in Zenodoo at [doi:10.5281/zenodo.4767610](https://doi.org/10.5281/zenodo.4767610).

Conflicts of Interest: The authors declare no conflict of interest.

References

- Collins, F.S. Implications of the Human Genome Project for Medical Science. *JAMA* **2001**, *285*, 540–544. [[CrossRef](#)] [[PubMed](#)]
- Ashley, E.A. Towards precision medicine. *Nat. Rev. Genet.* **2016**, *17*, 507–522. [[CrossRef](#)]
- Byron, S.A.; Van Keuren-Jensen, K.R.; Engelthaler, D.M.; Carpten, J.D.; Craig, D.W. Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. *Nat. Rev. Genet.* **2016**, *17*, 257–271. [[CrossRef](#)] [[PubMed](#)]
- Stark, R.; Grzelak, M.; Hadfield, J. RNA sequencing: The teenage years. *Nat. Rev. Genet.* **2019**, *20*, 631–656. [[CrossRef](#)] [[PubMed](#)]
- Rhoads, A.; Au, K.F. PacBio Sequencing and Its Applications. *Genom. Proteom. Bioinform.* **2015**, *13*, 278–289. [[CrossRef](#)]
- Sessegolo, C.; Cruaud, C.; Da Silva, C.; Cologne, A.; Dubarry, M.; Derrien, T.; Lacroix, V.; Aury, J.-M. Transcriptome profiling of mouse samples using nanopore sequencing of cDNA and RNA molecules. *Sci. Rep.* **2019**, *9*, 1–12. [[CrossRef](#)] [[PubMed](#)]
- Lu, H.; Giordano, F.; Ning, Z. Oxford Nanopore MinION Sequencing and Genome Assembly. *Genom. Proteom. Bioinform.* **2016**, *14*, 265–279. [[CrossRef](#)]
- Martin, J.A.; Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* **2011**, *12*, 671–682. [[CrossRef](#)] [[PubMed](#)]
- Jain, M.; Olsen, H.E.; Paten, B.; Akeson, M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* **2016**, *17*, 1–11. [[CrossRef](#)]
- Oikonomopoulos, S.; Wang, Y.C.; Djambazian, H.; Badescu, D.; Ragoussis, J. Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. *Sci. Rep.* **2016**, *6*, 31602. [[CrossRef](#)] [[PubMed](#)]
- Workman, R.E.; Tang, A.D.; Tang, P.S.; Jain, M.; Tyson, J.R.; Razaghi, R.; Zuzarte, P.C.; Gilpatrick, T.; Payne, A.; Quick, J.; et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods* **2019**, *16*, 1297–1305. [[CrossRef](#)]
- Byrne, A.; Beaudin, A.E.; Olsen, H.E.; Jain, M.; Cole, C.; Palmer, T.; Dubois, R.M.; Forsberg, E.C.; Akeson, M.; Vollmers, C. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* **2017**, *8*, 16027. [[CrossRef](#)]
- Kim, D.; Lee, J.-Y.; Yang, J.-S.; Kim, J.W.; Kim, V.N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921.e10. [[CrossRef](#)] [[PubMed](#)]
- Križanović, K.; Echchiki, A.; Roux, J.; Šikić, M. Evaluation of tools for long read RNA-seq splice-aware alignment. *Bioinformatics* **2018**, *34*, 748–754. [[CrossRef](#)] [[PubMed](#)]
- Soneson, C.; Yao, Y.; Bratus-Neuenschwander, A.; Patrignani, A.; Robinson, M.D.; Hussain, S. A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat. Commun.* **2019**, *10*, 1–14. [[CrossRef](#)] [[PubMed](#)]
- Uszczyńska-Ratajczak, B.; Lagarde, J.; Frankish, A.; Guigó, R.; Johnson, R. Towards a complete map of the human long non-coding RNA transcriptome. *Nat. Rev. Genet.* **2018**, *19*, 535–548. [[CrossRef](#)] [[PubMed](#)]
- Quick, J.; Ashton, P.; Calus, S.T.; Chatt, C.; Gossain, S.; Hawker, J.; Nair, S.; Neal, K.; Nye, K.; Peters, T.; et al. Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of Salmonella. *Genome Biol.* **2015**, *16*, 1–14. [[CrossRef](#)] [[PubMed](#)]
- Quick, J.; Loman, N.J.; Duraffour, S.; Simpson, J.T.; Severi, E.; Cowley, L.; Bore, J.A.; Koundouno, R.; Dudas, G.; Mikhail, A.; et al. Real-time, portable genome sequencing for Ebola surveillance. *Nat. Cell Biol.* **2016**, *530*, 228–232. [[CrossRef](#)]
- Quick, J.; Grubaugh, N.D.; Pullan, S.T.; Claro, I.M.; Smith, A.D.; Gangavarapu, K.; Oliveira, G.; Robles-Sikisaka, R.; Rogers, T.F.; Beutler, N.A.; et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **2017**, *12*, 1261–1276. [[CrossRef](#)] [[PubMed](#)]
- Li, J.; Wang, H.; Mao, L.; Yu, H.; Yu, X.; Sun, Z.; Qian, X.; Cheng, S.; Chen, S.; Chen, J.; et al. Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing. *Sci. Rep.* **2020**, *10*, 1–10. [[CrossRef](#)] [[PubMed](#)]
- Bull, R.A.; Adikari, T.N.; Ferguson, J.M.; Hammond, J.M.; Stevanovski, I.; Beukers, A.G.; Naing, Z.; Yeang, M.; Verich, A.; Gamaarachchi, H.; et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. *Nat. Commun.* **2020**, *11*, 1–8. [[CrossRef](#)]

22. Moore, S.C.; Penrice-Randall, R.; Alruwaili, M.; Randle, N.; Armstrong, S.; Hartley, C.; Haldenby, S.; Dong, X.; Alrezaihi, A.; Almsaud, M.; et al. Amplicon-Based Detection and Sequencing of SARS-CoV-2 in Nasopharyngeal Swabs from Patients With COVID-19 and Identification of Deletions in the Viral Genome That Encode Proteins Involved in Interferon Antagonism. *Viruses* **2020**, *12*, 1164. [[CrossRef](#)] [[PubMed](#)]
23. Wick, R.R.; Judd, L.M.; Gorrie, C.L.; Holt, K.E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom.* **2017**, *3*, e000132. [[CrossRef](#)] [[PubMed](#)]
24. King, J.; Harder, T.; Beer, M.; Pohlmann, A. Rapid multiplex MinION nanopore sequencing workflow for Influenza A viruses. *BMC Infect. Dis.* **2020**, *20*, 1–8. [[CrossRef](#)]
25. Tyler, A.D.; Mataseje, L.; Urfano, C.J.; Schmidt, L.; Antonation, K.S.; Mulvey, M.R.; Corbett, C.R. Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications. *Sci. Rep.* **2018**, *8*, 1–12. [[CrossRef](#)]
26. Baccarella, A.; Williams, C.R.; Parrish, J.Z.; Kim, C.C. Empirical assessment of the impact of sample number and read depth on RNA-Seq analysis workflow performance. *BMC Bioinform.* **2018**, *19*, 1–12. [[CrossRef](#)] [[PubMed](#)]
27. Liu, Y.; Ferguson, J.; Xue, C.; Silverman, I.M.; Gregory, B.; Reilly, M.P.; Li, M. Evaluating the Impact of Sequencing Depth on Transcriptome Profiling in Human Adipose. *PLoS ONE* **2013**, *8*, e66883. [[CrossRef](#)] [[PubMed](#)]
28. Desai, A.; Marwah, V.S.; Yadav, A.; Jha, V.; Dhaygude, K.; Bangar, U.; Kulkarni, V.; Jere, A. Identification of Optimum Sequencing Depth Especially for De Novo Genome Assembly of Small Genomes Using Next Generation Sequencing Data. *PLoS ONE* **2013**, *8*, e60204. [[CrossRef](#)] [[PubMed](#)]
29. Sims, D.W.; Sudbery, I.; Illott, N.E.; Heger, A.; Ponting, C.P. Sequencing depth and coverage: Key considerations in genomic analyses. *Nat. Rev. Genet.* **2014**, *15*, 121–132. [[CrossRef](#)]
30. Tarazona, S.; García-Alcalde, F.; Dopazo, J.; Ferrer, A.; Conesa, A. Differential expression in RNA-seq: A matter of depth. *Genome Res.* **2011**, *21*, 2213–2223. [[CrossRef](#)]
31. Orsini, P.; Minervini, C.F.; Cumbo, C.; Anelli, L.; Zagaria, A.; Minervini, A.; Coccaro, N.; Tota, G.; Casieri, P.; Impera, L.; et al. Design and MinION testing of a nanopore targeted gene sequencing panel for chronic lymphocytic leukemia. *Sci. Rep.* **2018**, *8*, 1–10. [[CrossRef](#)] [[PubMed](#)]
32. Jain, M.; Koren, S.; Miga, K.H.; Quick, J.; Rand, A.C.; Sasani, T.A.; Tyson, J.R.; Beggs, A.D.; Diltthey, A.T.; Fiddes, I.T.; et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* **2018**, *36*, 338–345. [[CrossRef](#)] [[PubMed](#)]
33. Bowden, R.; Davies, R.W.; Heger, A.; Pagnamenta, A.T.; De Cesare, M.; Oikkonen, L.E.; Parkes, D.; Freeman, C.; Dhalla, F.; Patel, S.Y.; et al. Sequencing of human genomes with nanopore technology. *Nat. Commun.* **2019**, *10*, 1–9. [[CrossRef](#)] [[PubMed](#)]
34. Seki, M.; Katsumata, E.; Suzuki, A.; Sereewattanawoot, S.; Sakamoto, Y.; Mizushima-Sugano, J.; Sugano, S.; Kohno, T.; Frith, M.C.; Tsuchihara, K.; et al. Evaluation and application of RNA-Seq by MinION. *DNA Res.* **2018**, *26*, 55–65. [[CrossRef](#)] [[PubMed](#)]
35. Byrne, A.; Cole, C.; Volden, R.; Vollmers, C. Realizing the potential of full-length transcriptome sequencing. *Philos. Trans. R. Soc. B Biol. Sci.* **2019**, *374*, 20190097. [[CrossRef](#)] [[PubMed](#)]
36. Kono, N.; Arakawa, K. Nanopore sequencing: Review of potential applications in functional genomics. *Dev. Growth Differ.* **2019**, *61*, 316–326. [[CrossRef](#)] [[PubMed](#)]
37. Bolisetty, M.T.; Rajadinakaran, G.; Graveley, B.R. Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol.* **2015**, *16*, 1–12. [[CrossRef](#)] [[PubMed](#)]
38. Loman, N.J.; Quick, J.; Simpson, J.T. A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nat. Methods* **2015**, *12*, 733–735. [[CrossRef](#)]
39. Istace, B.; Friedrich, A.; D'Agata, L.; Faye, S.; Payen, E.; Beluche, O.; Caradec, C.; Davidas, S.; Cruaud, C.; Liti, G.; et al. de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *GigaScience* **2017**, *6*, 1–13. [[CrossRef](#)] [[PubMed](#)]
40. Sahlin, K.; Medvedev, P. Error correction enables use of Oxford Nanopore technology for reference-free transcriptome analysis. *Nat. Commun.* **2021**, *12*, 1–13. [[CrossRef](#)]
41. Conesa, A.; Madrigal, P.; Tarazona, S.; Gomez-Cabrero, D.; Cervera, A.; McPherson, A.; Szczesniak, M.W.; Gaffney, D.J.; Elo, L.L.; Zhang, X.; et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**, *17*, 1–19. [[CrossRef](#)]
42. Cabili, M.N.; Trapnell, C.; Goff, L.; Koziol, M.J.; Tazon-Vega, B.; Regev, A.; Rinn, J.L. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **2011**, *25*, 1915–1927. [[CrossRef](#)] [[PubMed](#)]
43. Wick, R.R.; Judd, L.M.; Holt, K.E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **2019**, *20*, 1–10. [[CrossRef](#)]
44. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
45. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
46. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2013**, *30*, 923–930. [[CrossRef](#)] [[PubMed](#)]
47. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2009**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]

-
48. Kechin, A.; Boyarskikh, U.; Kel, A.; Filipenko, M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. *J. Comput. Biol.* **2017**, *24*, 1138–1143. [[CrossRef](#)] [[PubMed](#)]
 49. Dean, C.B.; Nielsen, J.D. Generalized linear mixed models: A review and some extensions. *Lifetime Data Anal.* **2007**, *13*, 497–512. [[CrossRef](#)] [[PubMed](#)]