


RESEARCH ARTICLE

Next-generation sequencing protocol of hematopoietic stem cells (HSCs). Step-by-step overview and troubleshooting guide

Justyna Jarczak¹ , Kamila Bujko¹, Katarzyna Brzeźniakiewicz-Janus², Mariusz Ratajczak³, Magdalena Kucia^{1*}

1 Laboratory of Regenerative Medicine at Medical University of Warsaw, Warsaw, Poland, **2** Department of Hematology, University of Zielona Gora, Multi-Specialist Hospital Gorzow Wlkp., Zielona Gora, Poland, **3** Stem Cell Institute at Graham Brown Cancer Center, University of Louisville, Louisville, KY, United States of America

* magdalena.kucia@wum.edu.pl



Abstract

Populations of very small embryonic-like stem cells (VSELs) (CD34+lin-CD45- and CD133+lin-CD45-), circulating in the peripheral blood of adults in small numbers, have been identified in several human tissues and together with the populations of hematopoietic stem cells (HSCs) (CD34+lin-CD45+) and CD133+lin-CD45+ constitute a pool of cells with self-renewal and pluripotent stem cell characteristics. Using advanced cell staining and sorting strategies, we isolated populations of VSELs and HSCs for bulk RNA-Seq analysis to compare the transcriptomic profiles of both cell populations. Libraries were prepared from an extremely small number of cells; however, their good quality was preserved, and they met the criteria for sequencing. We present here a step-by-step NGS protocol for sequencing VSELs and HSC with a description of troubleshooting during library preparation and sequencing.

OPEN ACCESS

Citation: Jarczak J, Bujko K, Brzeźniakiewicz-Janus K, Ratajczak M, Kucia M (2025) Next-generation sequencing protocol of hematopoietic stem cells (HSCs). Step-by-step overview and troubleshooting guide. PLoS ONE 20(1): e0313009. <https://doi.org/10.1371/journal.pone.0313009>

Editor: Francesco Bertolini, European Institute of Oncology, ITALY

Received: August 27, 2024

Accepted: October 16, 2024

Published: January 9, 2025

Copyright: © 2025 Jarczak et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data sets supporting this article's results are available in the Sequence Read Archive (SRA) repository (<https://www.ncbi.nlm.nih.gov/sra>) and assigned unique persistent identifier: SUB14759502.

Funding: This research was funded by National Science Center NCN OPUS 2021/41/B/NZ6/01590 grant from the National Science Center to Magdalena Kucia.

1. Introduction

Among the increasingly frequent transcriptome analyses at the single cell resolution, bulk RNA-Seq is still widely employed, especially because of its relatively low cost and greater simplicity of proceeding. Several examples of RNA-Seq experiments in hematopoiesis studies [1–7] and the developmental journey of hematopoietic stem cells [8–11]. Using RNA-Seq, we can analyze the gene expression level across the entire transcriptome in any experimental setup conditions. Considering the accuracy and affordability of RNA-Seq, it is recognized as the standard technology, especially in molecular biology and genetics [12–14]. The importance of experimental hematology is indisputable, considering the role of hematopoietic stem cells in bone marrow transplantations [3,15–21], immunotherapies [22–24], and in anti-pathogenic and anti-infection processes [24–27]. The dynamics and complexity of the hematopoiesis process involving self-renewal, differentiation, and proliferation require precise, innovative methods such as RNA-Seq to elucidate its molecular basis [7,28–31]. Hematopoietic stem/progenitor cells (HSPCs) are specified into cells from all hematopoietic lineages including

Competing interests: The authors have declared that no competing interests exist.

erythrocytes, granulocytes, monocytes, lymphocytes, and blood platelets. The major obstacle for RNA Seq in stem cell research is the limited number of primary cells.

The study aimed to evaluate the suitability of HSCs (CD34+lin-CD45+) and VSELs (CD34+lin-CD45-) for use in RNA sequencing and optimize the method of library preparation based on the low number and imperfect quality of RNA samples, isolated from HSC and very small embryonic-like stem cells (VSELs) [32–37]. We wanted to check whether the material with the low number of cells is still sufficient to prepare RNA-seq libraries and to obtain good quality NGS results enabling the analysis of the transcriptomic profile of these cells. In addition, we wanted to analyze the differences in transcriptomic profiles between VSELs and HSC, considering their origin and functions.

2. Materials and methods

2.1. Samples

Blood samples were obtained from 9 adult COVID patients at the Department of Hematology, Multi-Specialist Hospital Gorzow Wlkp, between 08.01.2023 and 31.03.2023. The information about age, sex, as well as clinical details is presented in Table 1. This study was performed based on guidelines and approval of the Medical University of Warsaw Bioethics Committee (permission number KB/50/2022). All procedures were performed in accordance with the Declaration of Helsinki (ethical principles for medical research involving human subjects). All participants gave written consent to participate in the study.

2.2. Isolation of HSCs and VSELs

Samples were constantly handled at 4 °C, avoiding cell disruption. Blood units, containing between 15–20 mL, were processed with the use of Lysis Buffer (BD) to perform erythrocyte lysis. Samples were incubated at 23 °C for 10 minutes and centrifuged for 30 min at 400x g at 4 °C. The procedure was repeated two times. The mononuclear phase was collected, washed, and used for further analysis.

2.3. Fluorescence-activated cell sorting

The methodology of fluorescence activated cell sorting has been described previously [32,38,39]. MNCs were stained with the following antibodies: Lineage (Lin) cocktail of antibodies, each FITC-conjugated: CD235a (clone GA-R2 [HIR2]), anti-CD2 (clone RPA-2.10), anti-CD3 (clone UCHT1), anti-CD14 (clone M5E2), anti-CD16 (clone 3G8), anti-CD19 (clone HIB19), anti-CD24 (clone ML5), anti-CD56 (clone NCAM16.2) and anti-CD66b (clone

Table 1. Characterization of the sample donors including age, sex, and medical diagnosis.

Donor ID	Age	Sex	Infection	Diagnosis
1	67	M	Influenza	Multiple myeloma
2	61	K	COVID	Ovarian cancer
3	69	M	-	-
4	63	K	COVID	-
5	77	M	COVID	Myelodysplastic syndrome
6	39	K	COVID	Hodgkin lymphoma
7	69	M	COVID	Multiple myeloma
8	59	M	COVID	Chronic lymphocytic leukemia
9	76	M	-	Chronic lymphocytic leukemia

<https://doi.org/10.1371/journal.pone.0313009.t001>

G10F5) (all BD Biosciences, San Jose, CA, United States); PE-Cy7-conjugated anti-CD45 (clone HI30, BioLegend, San Diego, CA, United States); and PE-conjugated anti-CD34 (clone 581, BioLegend, San Diego, CA, United States). Antibodies were used in the manufacturer's recommended concentration. Cells were stained in the dark, placed on ice for 30 min, then washed and resuspended in RPMI-1640 medium containing 2% FBS (Corning Inc, Corning, NY, United States). Cells were sorted according to the following strategy: first, small events (2–15 μ m in size) were included in the “lymphocyte like” gate and then further analyzed for the expression of Lin markers and CD45 and CD34 antigens (S1 Fig). Populations of CD34 + VSELs (CD34+lin–CD45–) and HSCs (CD34+lin–CD45+) were sorted on the MoFlo As-trios EQ cell sorter (Beckman Coulter, Brea, CA, United States).

2.4. RNA isolation and quality and quantity assessment

Sorted cells of four cell populations (CD34+lin–CD45– and CD34+lin–CD45+, as well as mononuclear cells (MNC), were used for RNA isolation by RNeasy Micro Kit (Qiagen, Germany). The isolation was supported by the DNA digestion step with the use of the RNase-Free DNase Set (Qiagen, Germany). RNA was eluted in the volume of 15 μ L. The isolated RNA samples have been subjected to a qualitative and quantitative assessment using a Quantus Fluorometer (Promega, USA) and TapeStation 4100 (Agilent Technologies, USA), respectively. Universal Human RNA Standard (Agilent Technologies, USA) was used as an internal control of the library preparation process. Information on RNA quantity and quality is presented in Table 2.

2.5. Library preparation and sequencing

Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus Kit (Illumina, USA) was used to prepare libraries; then they were pooled and run on Illumina NextSeq 1000/2000

Table 2. Sample characteristics including the number of cells, RNA concentration, and reference mapping results.

L. p.	Sample_ID	Cell type	The number of cells after sorting	RNA conc. (Quantus) (ng/ μ L)	The concentration of undiluted library (nM)	Number of reads (mln)	Percent of reads aligned to exons (%)
1	1	HSC	2300	1.2	-	-	-
2	2	VSEL	300	1.3	-	84	-
3	3	MNC	-	14.2	24.7	114	63.77
4	4	HSC	7000	1.3	78.9	60	46.64
5	5	VSEL	1800	1.2	-	118	-
6	6	MNC	-	16.2	-	8	-
7	7	HSC	11000	2.1	67.9	30	33.93
8	8	VSEL	1500	2.1	30.5	53	36.67
9	9	MNC	-	2.2	79.5	90	53.74
10	10	HSC	20000	0.79	134.5	67	58.79
11	11	VSEL	2000	0.42	-	33	-
12	12	MNC	-	78.8	1671.5	47	71.39
13	13	HSC	10000	0.91	128.2	37	46.51
14	14	VSEL	2000	0.37	-	52	-
15	15	MNC	-	38.7	4980.8	63	45.02
16	16	VSEL	1000	0.17	3.1	0	-
17	17	VSEL	1800	0.58	14.3	30	42.71
18	18	VSEL	350	0.4	-	-	-
19	19	VSEL	2200	lower than blank	-	-	-

<https://doi.org/10.1371/journal.pone.0313009.t002>

(Illumina, USA) in P2 flow cell chemistry (200 cycles) with paired-end sequencing mode assuming 30 millions of reads per sample. Quantity of libraries was measured with the use of the KAPA Library Quantification Kit (Roche, Switzerland), while quality was verified with a High-Sensitivity DNA Kit 1000 on TapeStation 4150 (Agilent Technologies, USA).

2.6. Data analysis

The pipeline described by Cebola et al., (<https://github.com/CebolaLab/RNA-seq>) employed several open-source algorithms for data analysis. Raw BCL files from the Illumina platform were demultiplexed and converted into raw sequences (FASTQ format) with the use of bcl2fastq [40]. Subsequently, the quality of sequences was assessed using fastqc [40], and the reports encompassing sequence quality, GC content, duplication rates, length distribution, K-mer content, and adapter contamination were compiled for all the samples. To remove poor quality sequences and adapter contamination, FASTQ files were trimmed using fastp [41]. Files after trimming were used for alignment to the reference genome (GRCh38.p13), along with a reference transcriptome with the use of a STAR aligner [42]. Files after mapping were sorted, indexed, and flagged with the use of samtools [42] to be ready for final post-alignment quality control with Qualimap [43]. Next, a matrix of gene counts was prepared based on previously aligned BAM files, with the use of Salmon [44]. Count data containing information about the transcripts per million (TPM) values were then imported into R using tximport [45]. The final process of tximport includes combining the transcript-level counts to gene-level for all samples. The use of DESeq2 enabled the analysis of differential gene expression in these data [46]. Data were normalized by the library size and differentially expressed genes (DEGs) were estimated with the use of a Generalized Linear Model (GLM). Data visualization was performed with the use of the ggplot2 R package (version 3.4.4) [47].

First steps of data analysis: pre-processing, quality control, filtering, alignment and reads quantification were done in a Python environment [48]; while differential gene expression and data visualization with the use of R [47], implemented in Anaconda distribution (Anon, 2020. Anaconda Software Distribution, Anaconda Inc. Available at: <https://docs.anaconda.com/>).

3. Results

3.1 Quantitative and qualitative assessment of libraries

First, we wanted to evaluate whether having a low number of cells and a minimal amount of RNA with imperfect quality due to the limited number of sorted cells would allow us to prepare libraries meeting the criteria for next generation sequencing. According to the manufacturer's instructions, for intact RNA samples, the average fragment length is ~375–475 bp and the expected insert size is ~190 bp. Among 20 libraries, with quality assessed with TapeStation 4150 (Agilent Technologies, USA), 12 (including Human Universal RNA Standard) had a fragment length of more than 300bp and met the standards for sequencing (Fig 1). These were libraries prepared from all three isolated and sorted cell populations: HSCs, VSELs, and MNCs. However, the best results came from libraries prepared from the increased amount of RNA, regardless of the cell type: 23-1-02-MNC, 21-MNC, 14-HSC, and 23-1-02-HSC (Fig 1). On the other hand, we also obtained libraries with bad quality, which were not included in the analysis (Fig 2). Therefore, good quality libraries were next employed for quantitative measurements by KAPA Library Quantification Kit (Roche, Switzerland). The mean library concentration was 1181.69 nM, reaching the maximum for the Human Universal RNA Standard (6966.3 nM), and the lowest value for 4-VSEL sample (3.1 nM). More detailed information on libraries concentration is summarized in Table 2.

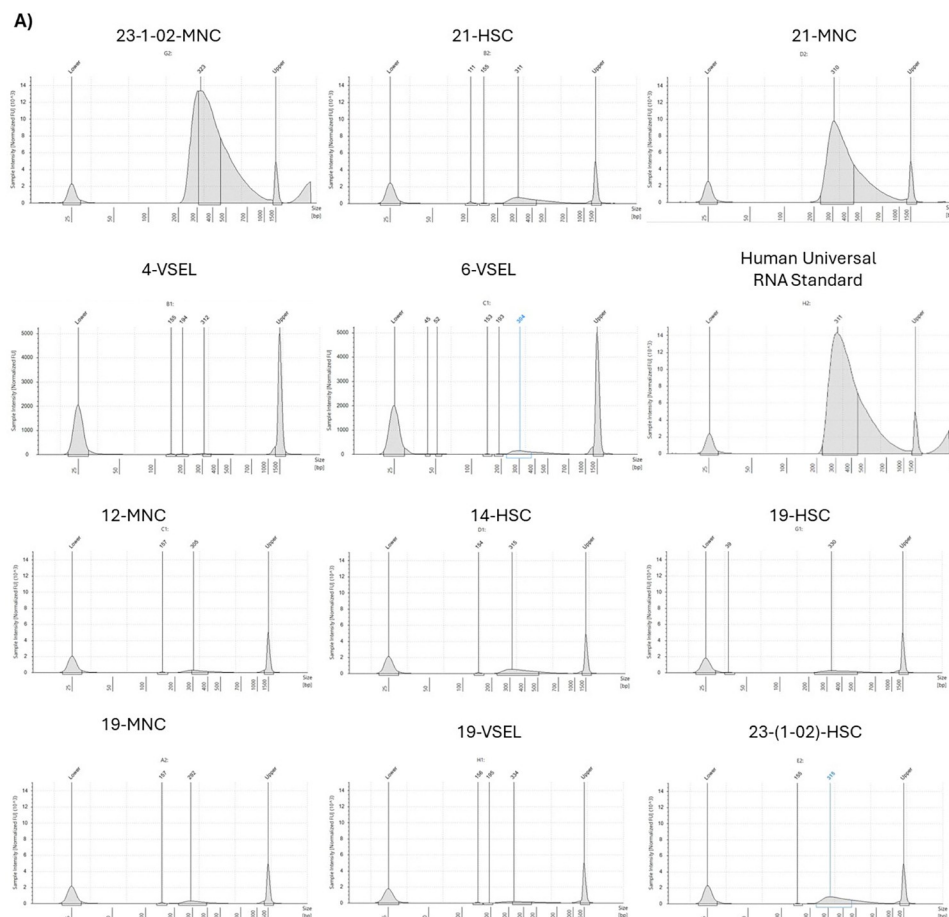


Fig 1. Electropherograms of “good quality” final libraries prepared with the use of Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus Kit (Illumina, USA), assessed with the D1000 ScreenTape assay. DNA shows a peak size of around 300 bp, which matches the expected size range between 300 and 400 bp.

<https://doi.org/10.1371/journal.pone.0313009.g001>

3.2 Quality parameters of sequencing runs and samples after sequencing

Libraries were sequenced in two runs (8 samples for 1 run). For the first run, we selected the best samples meeting both qualitative and quantitative criteria: 12-MNC, 14-HSC, 19-HSC, 19-VSEL, 19-MNC, 21-HSC, 21-MNC, and 6-VSEL. The second run included samples with insufficient quality. These were mostly libraries prepared from VSELS: 12-VSEL, 14-VSEL, 21-VSEL, 23(1-02)-VSEL and 4-VSEL as well as MNC: 14-MNC. We also added several other libraries: 23(1-02)-HSC and 23(1-02)-MNC. Both runs had an excellent quality parameter of sequencing with a QC30 parameter greater than 90%. Details are presented in [Table 3](#). Despite having good sequencing results, not all samples could be included in the analysis due to their poor-quality parameters ([Fig 3B](#)—lower panel, low-quality samples). However, for most samples, we achieved correct sequencing data allowing subsequent analysis ([Fig 3A](#)—upper panel, good quality samples). Interestingly, the number of reads for all samples from both runs was close to or greater than the assumed 30 million, except for 4-VSEL (0 reads) and 14-MNC (8 mln of reads) ([Table 3](#)). Furthermore, we did not find the statistically significant differences in the number of reads for all cell types ([Fig 4](#)).

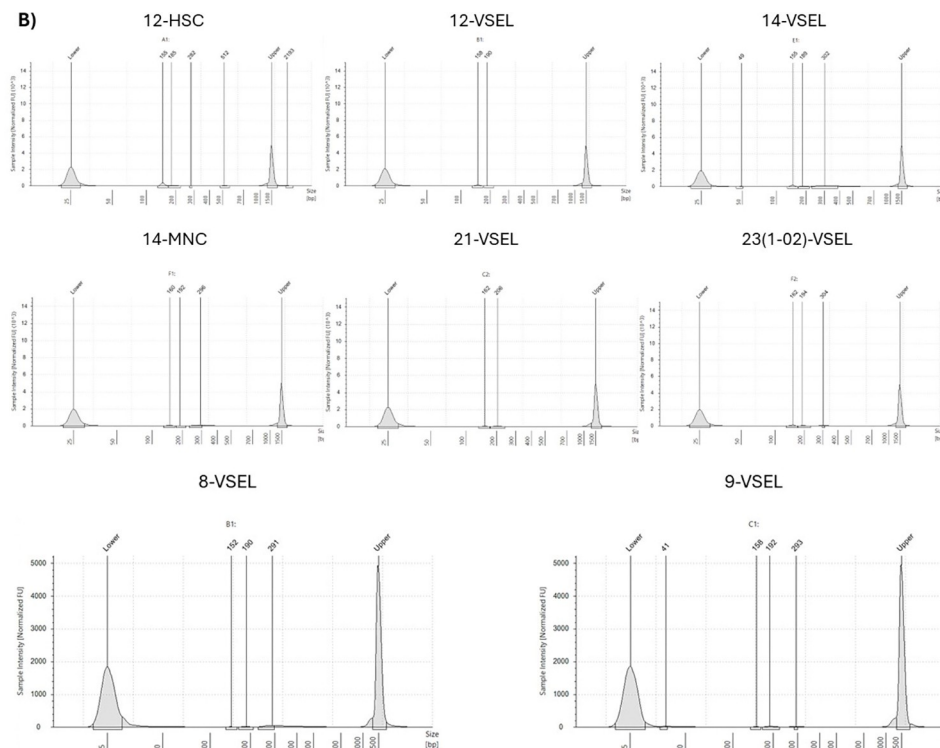


Fig 2. Electropherograms of “bad-quality” final libraries prepared with the use of Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus Kit (Illumina, USA), assessed with the D1000 ScreenTape assay. The 300bp peaks, which matches the expected size range are not observed.

<https://doi.org/10.1371/journal.pone.0313009.g002>

3.3 Mapping parameters

We selected 10 samples for the differential gene expression analysis. Reads were aligned to the human genome reference (GRCh38.p13) and mapping results are presented in Table 2. Approximately, 70–90% of the reads in a high-quality RNA-Seq dataset from a well-annotated genome should align to exons. The mean value evaluated for our samples was equal to 49.9%. The highest values were obtained for 21-MNC (71%) and 12-MNC (63%). This was expected since for those libraries we detected the high RNA concentration and the largest input taken for its preparation (Table 3). Conversely, the lowest values were obtained for 19-HSC (33.93%) and 19-VSEL (36.67). The RNA concentration and input utilized for library preparation were not the lowest. Therefore, the low mapping findings are likely attributable to other factors, such as inadequate RNA quality. However, we did not find any statistical significance in the percentage of reads aligned to exons for all cell types (Fig 4). This high percentage indicates effective mRNA enrichment and minimal contamination. Nonetheless, the precise percentage may fluctuate contingent upon various factors, including the choice of library preparation method, sequencing depth, genome annotation, and biological variability. Of note, it is also crucial to evaluate additional metrics, such as the percentage of reads aligned to genes, introns, and intergenic regions, to comprehensively assess the quality of RNA-Seq data.

Table 3. Quality parameters of sequencing runs.

Run	Parameter Q30 (%)	Yield (Gbp)	Aligned (%)	Error Rate (%)
1	92.93	90.40	5	0.11
2	91.73	88.64	4.38	0.27

<https://doi.org/10.1371/journal.pone.0313009.t003>



Fig 3. FastQC reports showing per base sequences quality in good (A) and bad (B) quality samples.

<https://doi.org/10.1371/journal.pone.0313009.g003>

3.4 Statistical analysis

The number of cells, RNA concentration, library concentration, number of reads, and percentage of reads aligned to exons were compared for HSCs, VSELs, and MNCs. We confirmed the differences in cell number between HSCs and VSELs (MNCs were not included in the comparison), which were isolated from the same blood volume, indicating individual variability (Fig 4, p-value = 0.001). However, the increased number of HSCs did not result in higher RNA concentration (Fig 4). We also did not observe any statistically significant differences in library concentration (except for MNCs), the number of reads, and the percentage of reads aligned to exons between HSCs, VSELs and MNC (Fig 4).

In addition, we did not find any statistically significant correlations between all variables (Fig 5) indicating that even samples with low RNA concentration may be considered for analysis and may not achieve low sequencing parameters.

3.5 Differential gene expression

Finally, we run differential gene expression analysis using DESeq2, to find the differences in transcriptomic profiles between VSELs and HSCs. MNCs were not included in the analysis. Considering the results of the mapping, it was possible to include 2 VSELs samples and 4

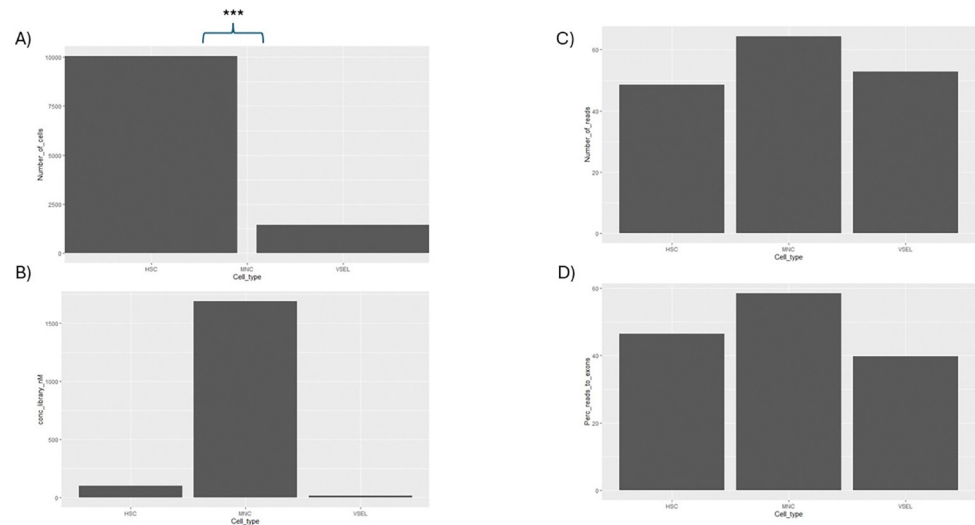


Fig 4. Differences in the cell number (A), library concentration (B), number of reads (C), and percentage of reads aligned to exons (D) between HSCs, VSELs, and MNCs.

<https://doi.org/10.1371/journal.pone.0313009.g004>

HSCs samples. According to this, we unfortunately observed weak statistical power, the presence of outliers, and results bias. However, this was mostly related to the clinical reasons (patients with CLL) rather than the quality of the prepared libraries.

4. Discussion

Using well-established methods with little sequence-specific sample manipulation produces the greatest outcomes in terms of response linearity and reproducibility when sample quantity is not the limiting issue. On the other hand, if gathering such comparatively big samples is not feasible "low-input RNA-seq" is often utilized; however, there are specific challenges and strategies to ensure high-quality data [7,49]. Major challenges are related to the low RNA yield, amplification bias, and dropout events [50].

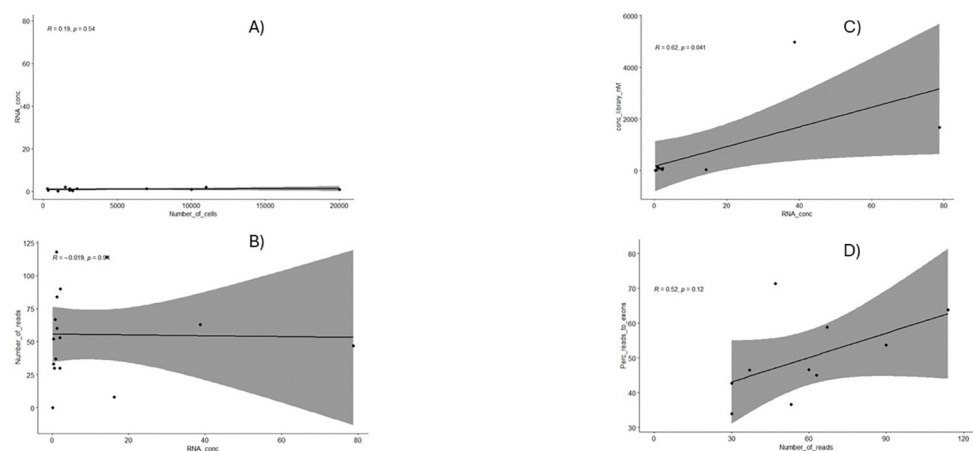


Fig 5. Pearson correlations between RNA concentration and the cell number (A), RNA concentration and number of reads (B), RNA concentration and library concentration (C), and number of reads and percentage of reads aligned to exons (D).

<https://doi.org/10.1371/journal.pone.0313009.g005>

Although there is some discernible volatility introduced by preamplification processes, the data quality is not significantly affected, and no clear sequence-specific biases are introduced [51,52].

Next, although there are indicators that can provide high levels of HSC purity, HSC populations continue to be functionally diverse in terms of their ability to self-renew after transplantation [53,54]. To methodically describe each cell's global transcriptional landscape, we employed RNA-seq after multiparameter fluorescence-activated cell sorting (FACS). Principal component analysis was carried out to identify the primary causes of variance (PCA).

The specific research question that an RNAseq experiment is intended to address may or may not be impacted by the fact that bulk tissues may have a mixture of several different cell types. About 0.58 ng of total RNA is the lowest limit of dependable library formation that we have found using the Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus kit, and we have used this quantity of RNA in as-yet-unpublished investigations on peripheral blood sorted population of HSC and VSELs.

The quantity of material available for profiling, determines which RNAseq procedures should be used. We discovered that by utilizing the standard Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus according to manufacturer procedure, excellent quality libraries may be created with a little less than the producer's recommended minimum. We found that the Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus methodology yields data with a superior linear response to the growing concentration of any given gene when there is enough material available to use it, as opposed to several other "single cell" protocols that perform relatively similarly in this metric.

The techniques and algorithms for RNA-seq are still developing. Many novel applications of RNAseq-based analysis are now feasible as a result of advancements. There are differences in the essential statistical methodologies and concerns for sample-level, gene-level, transcript-level, and exon-level applications. Given that various statistical models and data distribution assumptions are frequently employed by tools to address identical biological questions. The agreement between the results varies so frequently. The final interpretation may differ depending on the parameters and tool selection. The results can potentially be affected by bias resulting from the sequencing library, such as bias resulting from short reads or insufficient small non-coding RNA detection bias caused by RNA modification [42].

To sum up, we outlined the developing process for analysis based on bulk RNAseq and emphasized the information gaps. Obtaining agreement on the most appropriate instruments or pipelines is difficult because biological goals and situations vary. We offer a useful guide of key limitations and quality control analysis that should be addressed when working with low input RNA Seq.

5. Conclusions

The article discusses techniques and strategies for optimizing RNA sequencing in scenarios with limited cell numbers and compromised sample quality. It highlights the importance of careful sample handling and robust data analysis approaches to mitigate the challenges posed by low-input and degraded RNA samples. Importantly, the study emphasizes that even working with stem cells and their low number do not have to disqualify samples from RNA-Seq analysis.

Supporting information

S1 Fig. Cell sorting strategy for the isolation of HSCs and VSELs from peripheral whole blood. Immunostained cells were first visualized by dot plot showing forward scatter (FSC) vs.

side scatter (SSC) signals, where small events ranging from 2–15 μm were gated (P1) (A) and further analyzed for the expression of Lineage markers. Lineage negative events were gated (Lin-) (B) and analyzed for the expression of CD45 and CD34 antigens. The populations of CD34+Lin-CD45+ HSCs (C) and CD34+Lin-CD45- VSELs (D) were separately sorted on the MoFlo Astrios EQ cell sorter. MNCs isolation and staining was described in the Materials and Methods section. Representative dot plots saved during the sample acquisition are shown. (TIF)

Author Contributions

Conceptualization: Magdalena Kucia.

Data curation: Kamila Bujko.

Formal analysis: Justyna Jarczak.

Investigation: Justyna Jarczak.

Methodology: Justyna Jarczak, Kamila Bujko, Katarzyna Brzeźniakiewicz-Janus.

Project administration: Magdalena Kucia.

Resources: Kamila Bujko, Katarzyna Brzeźniakiewicz-Janus.

Software: Justyna Jarczak.

Supervision: Mariusz Ratajczak.

Writing – original draft: Justyna Jarczak.

Writing – review & editing: Mariusz Ratajczak, Magdalena Kucia.

References

- McKinney-Freeman S, Cahan P, Li H, Lacadie SA, Huang HT, Curran M, et al. The transcriptional landscape of hematopoietic stem cell ontogeny. *Cell Stem Cell*. 2012; 11(5):701–14. <https://doi.org/10.1016/j.stem.2012.07.018> PMID: 23122293; PubMed Central PMCID: PMC3545475.
- Mikkola HK, Orkin SH. The journey of developing hematopoietic stem cells. *Development*. 2006; 133(19):3733–44. Epub 2006/09/14. <https://doi.org/10.1242/dev.02568> PMID: 16968814.
- Papayannopoulou T. Bone marrow homing: the players, the playfield, and their evolving roles. *Curr Opin Hematol*. 2003; 10(3):214–9. <https://doi.org/10.1097/00062752-200305000-00004> PMID: 12690289.
- Ratajczak MZ, Bujko K, Brzeźniakiewicz-Janus K, Ratajczak J, Kucia M. Hematopoiesis Revolves Around the Primordial Evolutional Rhythm of Purinergic Signaling and Innate Immunity—A Journey to the Developmental Roots. *Stem Cell Rev Rep*. 2024; 20(3):827–38. Epub 20240216. <https://doi.org/10.1007/s12015-024-10692-9> PMID: 38363476; PubMed Central PMCID: PMC10984895.
- Rossi L, Challen GA, Sirin O, Lin KK, Goodell MA. Hematopoietic stem cell characterization and isolation. *Methods Mol Biol*. 2011; 750:47–59. https://doi.org/10.1007/978-1-61779-145-1_3 PMID: 21618082; PubMed Central PMCID: PMC3621966.
- Zeng Y, He J, Bai Z, Li Z, Gong Y, Liu C, et al. Tracing the first hematopoietic stem cell generation in human embryo by single-cell RNA sequencing. *Cell Res*. 2019; 29(11):881–94. Epub 20190909. <https://doi.org/10.1038/s41422-019-0228-6> PMID: 31501518; PubMed Central PMCID: PMC6888893.
- Zhang Y, Ye T, Gong S, Hong Z, Zhou X, Liu H, et al. RNA-sequencing based bone marrow cell transcriptome analysis reveals the potential mechanisms of E'jiao against blood-deficiency in mice. *Biomed Pharmacother*. 2019; 118:109291. Epub 20190808. <https://doi.org/10.1016/j.biopha.2019.109291> PMID: 31401395.
- Tang Q, Iyer S, Lobbardi R, Moore JC, Chen H, Lareau C, et al. Dissecting hematopoietic and renal cell heterogeneity in adult zebrafish at single-cell resolution using RNA sequencing. *J Exp Med*. 2017; 214(10):2875–87. Epub 20170906. <https://doi.org/10.1084/jem.20170976> PMID: 28878000; PubMed Central PMCID: PMC5626406.

9. Granot N, Storb R. History of hematopoietic cell transplantation: challenges and progress. *Haematologica*. 2020; 105(12):2716–29. Epub 20201201. <https://doi.org/10.3324/haematol.2019.245688> PMID: 33054108; PubMed Central PMCID: PMC7716373.
10. Gschweng E, De Oliveira S, Kohn DB. Hematopoietic stem cells for cancer immunotherapy. *Immunol Rev*. 2014; 257(1):237–49. <https://doi.org/10.1111/imr.12128> PMID: 24329801; PubMed Central PMCID: PMC3901057.
11. Nasiri K, Mohammadzadehsalini S, Kheradjoo H, Shabestari AM, Eshaghizadeh P, Pakmehr A, et al. Spotlight on the impact of viral infections on Hematopoietic Stem Cells (HSCs) with a focus on COVID-19 effects. *Cell Commun Signal*. 2023; 21(1):103. Epub 20230508. <https://doi.org/10.1186/s12964-023-01122-3> PMID: 37158893; PubMed Central PMCID: PMC10165295.
12. Hrdlickova R, Toloue M, Tian B. RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip Rev RNA*. 2017; 8(1). Epub 20160519. <https://doi.org/10.1002/wrna.1364> PMID: 27198714; PubMed Central PMCID: PMC5717752.
13. Bivens NJ, Zhou M. RNA-Seq Library Construction Methods for Transcriptome Analysis. *Curr Protoc Plant Biol*. 2016; 1(1):197–215. <https://doi.org/10.1002/cppb.20019> PMID: 31725988.
14. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: a method for comprehensive transcriptome analysis. *Curr Protoc Mol Biol*. 2010; Chapter 4: Unit 4 11 1–3. <https://doi.org/10.1002/0471142727.mb0411s89> PMID: 20069539.
15. Shin DM, Liu R, Klich I, Wu W, Ratajczak J, Kucia M, et al. Molecular signature of adult bone marrow-purified very small embryonic-like stem cells supports their developmental epiblast/germ line origin. *Leukemia*. 2010; 24(8):1450–61. Epub 2010/05/29. <https://doi.org/10.1038/leu.2010.121> PMID: 20508611.
16. Lapidot T, Dar A, Kollet O. How do stem cells find their way home? *Blood*. 2005; 106(6):1901–10. Epub 20050512. <https://doi.org/10.1182/blood-2005-04-1417> PMID: 15890683.
17. Orkin SH, Zon LI. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*. 2008; 132(4):631–44. <https://doi.org/10.1016/j.cell.2008.01.025> PMID: 18295580; PubMed Central PMCID: PMC2628169.
18. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. *Cell Stem Cell*. 2012; 10(2):120–36. <https://doi.org/10.1016/j.stem.2012.01.006> PMID: 22305562.
19. Jenq RR, van den Brink MR. Allogeneic haematopoietic stem cell transplantation: individualized stem cell and immune therapy of cancer. *Nat Rev Cancer*. 2010; 10(3):213–21. Epub 20100219. <https://doi.org/10.1038/nrc2804> PMID: 20168320.
20. Gluckman E. Cord blood transplantation. *Biol Blood Marrow Transplant*. 2006; 12(8):808–12. <https://doi.org/10.1016/j.bbmt.2006.05.011> PMID: 16864050.
21. Martelli MF, Di Ianni M, Ruggeri L, Falzetti F, Carotti A, Terenzi A, et al. HLA-haploidentical transplantation with regulatory and conventional T-cell adoptive immunotherapy prevents acute leukemia relapse. *Blood*. 2014; 124(4):638–44. Epub 20140612. <https://doi.org/10.1182/blood-2014-03-564401> PMID: 24923299.
22. Alexander T, Greco R. Hematopoietic stem cell transplantation and cellular therapies for autoimmune diseases: overview and future considerations from the Autoimmune Diseases Working Party (ADWP) of the European Society for Blood and Marrow Transplantation (EBMT). *Bone Marrow Transplant*. 2022; 57(7):1055–62. Epub 20220516. <https://doi.org/10.1038/s41409-022-01702-w> PMID: 35578014; PubMed Central PMCID: PMC9109750.
23. Juric MK, Ghimire S, Ogonek J, Weissinger EM, Holler E, van Rood JJ, et al. Milestones of Hematopoietic Stem Cell Transplantation—From First Human Studies to Current Developments. *Front Immunol*. 2016; 7:470. Epub 20161109. <https://doi.org/10.3389/fimmu.2016.00470> PMID: 27881982; PubMed Central PMCID: PMC5101209.
24. Guillaume T, Rubinstein DB, Symann M. Immune reconstitution and immunotherapy after autologous hematopoietic stem cell transplantation. *Blood*. 1998; 92(5):1471–90. PMID: 9716573.
25. Appelbaum FR. Haematopoietic cell transplantation as immunotherapy. *Nature*. 2001; 411(6835):385–9. <https://doi.org/10.1038/35077251> PMID: 11357147.
26. Till BG, Jensen MC, Wang J, Chen EY, Wood BL, Greisman HA, et al. Adoptive immunotherapy for indolent non-Hodgkin lymphoma and mantle cell lymphoma using genetically modified autologous CD20-specific T cells. *Blood*. 2008; 112(6):2261–71. Epub 20080528. <https://doi.org/10.1182/blood-2007-12-128843> PMID: 18509084; PubMed Central PMCID: PMC2532803.
27. Maus MV, Grupp SA, Porter DL, June CH. Antibody-modified T cells: CARs take the front seat for hematologic malignancies. *Blood*. 2014; 123(17):2625–35. Epub 20140227. <https://doi.org/10.1182/blood-2013-11-492231> PMID: 24578504; PubMed Central PMCID: PMC3999751.

28. Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.* 2015; 25(12):1860–72. Epub 20151001. <https://doi.org/10.1101/gr.192237.115> PMID: 26430063; PubMed Central PMCID: PMC4665007.
29. Choi J, Baldwin TM, Wong M, Bolden JE, Fairfax KA, Lucas EC, et al. Haemopedia RNA-seq: a database of gene expression during haematopoiesis in mice and humans. *Nucleic Acids Res.* 2019; 47(D1): D780–D5. <https://doi.org/10.1093/nar/gky1020> PMID: 30395284; PubMed Central PMCID: PMC6324085.
30. Laurenti E, Gottgens B. From haematopoietic stem cells to complex differentiation landscapes. *Nature.* 2018; 553(7689):418–26. <https://doi.org/10.1038/nature25022> PMID: 29364285; PubMed Central PMCID: PMC6555401.
31. Nam AS, Kim KT, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature.* 2019; 571(7765):355–60. Epub 20190703. <https://doi.org/10.1038/s41586-019-1367-0> PMID: 31270458; PubMed Central PMCID: PMC6782071.
32. Bujko K, Ciechanowicz AK, Kucia M, Ratajczak MZ. Molecular analysis and comparison of CD34(+) and CD133(+) very small embryonic-like stem cells purified from umbilical cord blood. *Cytometry A.* 2023; 103(9):703–11. Epub 20230607. <https://doi.org/10.1002/cyto.a.24767> PMID: 37246957.
33. Domingues A, Rossi E, Bujko K, Detrich G, Richez U, Blandinieres A, et al. Human CD34(+) very small embryonic-like stem cells can give rise to endothelial colony-forming cells with a multistep differentiation strategy using UM171 and nicotinamide acid. *Leukemia.* 2022; 36(5):1440–3. Epub 20220215. <https://doi.org/10.1038/s41375-022-01517-0> PMID: 35169243; PubMed Central PMCID: PMC9061289.
34. Guerin CL, Loyer X, Vilar J, Cras A, Mirault T, Gaussem P, et al. Bone-marrow-derived very small embryonic-like stem cells in patients with critical leg ischaemia: evidence of vasculogenic potential. *Thromb Haemost.* 2015; 113(5):1084–94. Epub 2015/01/23. <https://doi.org/10.1160/TH14-09-0748> PMID: 25608764.
35. Kassmer SH, Jin H, Zhang PX, Bruscia EM, Heydari K, Lee JH, et al. Very small embryonic-like stem cells from the murine bone marrow differentiate into epithelial cells of the lung. *Stem Cells.* 2013; 31(12):2759–66. Epub 2013/05/18. <https://doi.org/10.1002/stem.1413> PMID: 23681901; PubMed Central PMCID: PMC4536826.
36. Havens AM, Shiozawa Y, Jung Y, Sun H, Wang J, McGee S, et al. Human very small embryonic-like cells generate skeletal structures, in vivo. *Stem Cells Dev.* 2013; 22(4):622–30. Epub 2012/10/02. <https://doi.org/10.1089/scd.2012.0327> PMID: 23020187; PubMed Central PMCID: PMC3564465.
37. Parte S, Bhartiya D, Telang J, Daithankar V, Salvi V, Zaveri K, et al. Detection, characterization, and spontaneous differentiation in vitro of very small embryonic-like putative stem cells in adult mammalian ovary. *Stem Cells Dev.* 2011; 20(8):1451–64. Epub 2011/02/05. <https://doi.org/10.1089/scd.2010.0461> PMID: 21291304; PubMed Central PMCID: PMC3148829.
38. Suszynska M, Zuba-Surma EK, Maj M, Mierzejewska K, Ratajczak J, Kucia M, et al. The proper criteria for identification and sorting of very small embryonic-like stem cells, and some nomenclature issues. *Stem Cells Dev.* 2014; 23(7):702–13. Epub 2013/12/05. <https://doi.org/10.1089/scd.2013.0472> PMID: 24299281; PubMed Central PMCID: PMC3967357.
39. Abdelbaset-Ismael A, Suszynska M, Borkowska S, Adamiak M, Ratajczak J, Kucia M, et al. Human haematopoietic stem/progenitor cells express several functional sex hormone receptors. *J Cell Mol Med.* 2016; 20(1):134–46. Epub 20151030. <https://doi.org/10.1111/jcmm.12712> PMID: 26515267; PubMed Central PMCID: PMC4717849.
40. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
41. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018; 34(17):i884–i90. <https://doi.org/10.1093/bioinformatics/bty560> PMID: 30423086; PubMed Central PMCID: PMC6129281.
42. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013; 29(1):15–21. Epub 20121025. <https://doi.org/10.1093/bioinformatics/bts635> PMID: 23104886; PubMed Central PMCID: PMC3530905.
43. Garcia-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Gotz S, Tarazona S, et al. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics.* 2012; 28(20):2678–9. Epub 20120822. <https://doi.org/10.1093/bioinformatics/bts503> PMID: 22914218.
44. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods.* 2017; 14(4):417–9. Epub 20170306. <https://doi.org/10.1038/nmeth.4197> PMID: 28263959; PubMed Central PMCID: PMC5600148.

45. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2021. <https://www.R-project.org>.
46. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15(12):550. <https://doi.org/10.1186/s13059-014-0550-8> PMID: 25516281; PubMed Central PMCID: PMC4302049.
47. Team RC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. 2022. <https://www.R-project.org>.
48. Mullender S. X Python reference manual. Department of Computer Science [CS]. CWI. 1995.
49. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 2016; 17:13. Epub 20160126. <https://doi.org/10.1186/s13059-016-0881-8> PMID: 26813401; PubMed Central PMCID: PMC4728800.
50. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced Applications of RNA Sequencing and Challenges. *Bioinform Biol Insights.* 2015; 9(Suppl 1):29–46. Epub 20151115. <https://doi.org/10.4137/BBI.S28991> PMID: 26609224; PubMed Central PMCID: PMC4648566.
51. Yalamanchili HK, Wan YW, Liu Z. Data Analysis Pipeline for RNA-seq Experiments: From Differential Expression to Cryptic Splicing. *Curr Protoc Bioinformatics.* 2017; 59:11 5 1–5 21. Epub 20170913. <https://doi.org/10.1002/cpbi.33> PMID: 28902396; PubMed Central PMCID: PMC6373869.
52. Jiang G, Zheng JY, Ren SN, Yin W, Xia X, Li Y, et al. A comprehensive workflow for optimizing RNA-seq data analysis. *BMC Genomics.* 2024; 25(1):631. Epub 20240624. <https://doi.org/10.1186/s12864-024-10414-y> PMID: 38914930; PubMed Central PMCID: PMC11197194.
53. Haas S, Trumpp A, Milsom MD. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell.* 2018; 22(5):627–38. <https://doi.org/10.1016/j.stem.2018.04.003> PMID: 29727678.
54. Crisan M, Dzierzak E. The many faces of hematopoietic stem cell heterogeneity. *Development.* 2016; 143(24):4571–81. <https://doi.org/10.1242/dev.114231> PMID: 27965438; PubMed Central PMCID: PMC5201026.