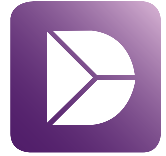
Data in Brief 40 (2022) 107736

Contents lists available at ScienceDirect

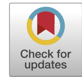
Data in Brief

journal homepage: www.elsevier.com/locate/dib



Data Article

The comprehensive transcriptome of human ductus arteriosus smooth muscle cells (hDASMC)



Rachel E.T. Bentley[a,∗], Charles C.T. Hindmarch[a,b,∗], Kimberly J. Dunham-Snary[a,c], Brooke Snetsinger[b], Jeffrey D. Mewburn[c], Arthur Thébaud[d], Patricia D.A. Lima[b], Bernard Thébaud[e,f,g,h], Stephen L. Archer[a,b,∗∗]

[a] Department of Medicine, Queen's University, Kingston, Ontario, Canada
[b] QCPU, Queen's Cardiopulmonary Unit, Translational Institute of Medicine (TIME), Department of Medicine, Queen's University, Canada
[c] Department of Biomedical and Molecular Science, Queen's University, Canada
[d] Department of Kinesiology and Health Studies, Queen's University, Canada
[e] Ottawa Hospital Research Institute, Regenerative Medicine Program, Ottawa, Ontario, Canada
[f] Department of Pediatrics, Division of Neonatology, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada
[g] Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada
[h] Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Ontario, Canada

A R T I C L E I N F O

Article history:
Received 23 September 2021
Revised 15 December 2021
Accepted 16 December 2021
Available online 20 December 2021

Keywords:
Patent ductus arteriosus
RNAsequencing
Normoxia
Hypoxia
Oxygen sensing
Smooth muscle cells

A B S T R A C T

The Ductus Arteriosus (DA) is a fetal vessel that connects the aorta to the pulmonary artery ensuring that placental oxygenated blood is diverted from the lungs to the systemic circulation. Following exposure to oxygen (O_2), in the first few days of life, the DA responds with a functional closure that is followed by anatomical closure. Here, we study human DA smooth muscle cells (DASMC) taken from 10 term infants during congenital heart surgery. Purification of these cells using flow cytometry ensured a pure population of DASMCs, which we confirmed as responsive to O_2. An oxygen-induced increase in intracellular calcium of 18.1%±4.4% and SMC constriction (-27%±1.5% shortening) occurred in all cell lines

DOI of original article: 10.1016/j.ygeno.2021.07.006
∗∗ Corresponding author(s).
E-mail address: stephen.archer@queensu.ca (S.L. Archer).
∗ These authors contributed equally and are co-first authors.

https://doi.org/10.1016/j.dib.2021.107736
2352-3409/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

within five minutes. These cells were maintained in either hypoxia (2.5% O_2), mimicking *in utero* conditions or in normoxia (19% O_2) mimicking neonate conditions. We then used 3' RNAsequencing to identify the transcriptome of DASMCs in each condition [1]. In this paper, we present the full differentially regulated gene list from this experiment.

© 2021 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications Table

Subject	Omics: Transcriptomics
Specific subject area	Human disease, Molecular biology,
Type of data	Datafile (.txt)
How data were acquired	RNAsequencing of highly purified ductus arteriosus smooth muscle cells from infants undergoing congenital heart surgery.
Data format	Raw (.txt file)
Parameters for data collection	1. Human DASMC extracted from human infants undergoing surgery for non-Patent ductus arteriosus (non-PDA) congenital heart defects 2. Cells were stored at -80C until cells were woken, and grown under hypoxic conditions 3. Cells were sorted for smooth muscle cells (SMA) to ensure a pure population 4. Cells were grown in either hypoxia (to mimic in utero physiology) and normoxia (to mimic neonate physiology). 5. Cells were processed (RNA extraction) and sequenced (Illumina) to create counts files for each transcript in each sample so that differential analysis can be performed (see co-submission, Bentley et al 2021).
Description of data collection	Following RNA extraction, 3' libraries (Lexogen) were constructed and sequenced using an Illumina Nextseq550
Data source location	Queen's University, Kingston, Ontario, Canada. Professor, Head Department of Medicine, Queen's University, Kingston, Ontario, Canada, K7L 3N6.
Data accessibility	Repository name: Mendeley Data Data identification number: http://dx.doi.org/10.17632/z42wpkbb8k.1 Direct link to the dataset: https://data.mendeley.com/datasets/z42wpkbb8k/1
Related research article	R.E.T. Bentley, C.C.T. Hindmarch, K.J. Dunham-Snary, B. Snetsinger, J.D. Mewburn, A. Thébaud, P.D.A. Lima, B. Thébaud, S.L. Archer, The molecular mechanisms of oxygen-sensing in human ductus arteriosus smooth muscle cells: A comprehensive transcriptome profile reveals a central role for mitochondria, Genomics, 113 (2021) 3128-3140 [1]. PMID: 34245829; DOI: 10.1016/j.ygeno.2021.07.006

Value of the Data

- These data profile the transcriptome of the DASMC under hypoxic and normoxic conditions and provide counts data for each gene in each sample.
- These files can be used to identify differentially regulated genes that exist between these conditions and identify the underlying molecular mechanisms of oxygen sensing in this tissue.
- This data will be relevant to researchers wishing to understand the molecular signature of the DASMCs and understand the oxygen sensing capacity of these cells

• Further integration with other 'omic' data such as proteomics will offer a more comprehensive understanding of the DA oxygen sensor.

1. Data Description

Data file 1: The data file contains all data associated with our co-submission to Genomics [1] where we only reported genes satisfying a adjusted p-value <0.05 were presented. Here, we do not filter our data, which will ensure the scientific community has access to sub-significant data. The data file column definitions is defined below:

Column	Definition
hgnc_symbol	Human Genome Official Gene Symbol.
baseMean	The effect size estimate. This value indicates how much the gene or transcript's expression seems to have changed between the comparison and control groups. This value is reported on a logarithmic scale to base 2.*
log2FoldChange	log 2 fold change*
lfcSE	The standard error estimate for the log2 fold change estimate.*
stat	The value of the test statistic for the gene or transcript.*
pvalue	P-value of the test for the gene or transcript.*
padj	Adjusted P-value for multiple testing for the gene or transcript.*
###_Hx	### = Sample name, Hx = Hypoxia
###_Nx	### = Sample name, Nx = Normoxia

*Definitions derived from Illumina

2. Experimental Design, Materials and Methods

RNA isolation and sequencing of purified hDASMC cells under hypoxic or normoxic environmental conditions

Human DASMC cells grown for 96-hours in either or normoxia (19.6% O_2, 5% CO_2, balance N_2) or hypoxia (2.5% O_2, 5% CO_2, balance N_2). RNA from each of these 10 cell lines were extracted using 1mL of TRI Reagent® (Sigma-Aldrich, cat# T9424-100ML, Lot # MKCK9023) and processed in batch using the Zymo Direct-zol RNA MiniPrep kit (cat# R2050, Zymo Research, California, USA). The samples in TRI Reagent® were centrifuged at 12 000 *g* for 30 seconds to remove particulate debris, collecting the supernatant, and transferring it to an RNase-free tube. An equal volume of anhydrous ethanol was added to the samples and mixed thoroughly, transferring the mixture into a Zymo-Spin™ Column in a collection tube and centrifuging at 12 000 *g* for 30 seconds, discarding the flow through. The columns were transferred to new collection tubes and 400 µL of wash buffer was added to the columns, centrifuging the columns and discarding the flow through. 80 µL of a DNase I treatment (5 µL DNase I, 75 µL DNA Digestion Buffer per sample) was added directly to each column matrix, incubating for 15 minutes at room temperature. Following DNase I treatment Direct-zol™ RNA PreWash (400 µL) was added to each column and centrifuged, discarding the flow-through; this step was repeated. 700 µL of RNA Wash buffer was added to each column, centrifuging for 2 minutes at 12 000 *g*. Following the wash step, the columns were transferred to new RNase-free tubes. 50 µL DNase/RNase-free water was added to the matrix of each column, centrifuging for 30 seconds at 12,000 *g*.

To ensure the highest quality RNA, quality control was performed using the Qubit spectrophotometer (Thermo Fisher Scientific, Waitham, Massachusetts, USA), which allowed us to capture the absorbance at 260nm, 280nm and 230nm and calculate that the RNA was pure and free of contaminants like ethanol. Libraries for each sample of pure DASMC were then built using the QuantSeq 3' FWD mRNA-Seq Library Prep Kit for Illumina (Lexogen, Austria). All steps were performed at room temperature, filter tips were used and samples were mixed well and centrifuged to collect at each step unless stated otherwise. All mastermixes were prepared to account for all samples +10%. For each cell line, 450ng of RNA was prepared in a volume of 5 µL

and combined with 5 µL of First Strand cDNA synthesis Mix 1 (FS1); 96-well plates were used to simplify batch library construction. Samples were then denatured using 5 µL of the RNA-FS1 mix for 3 minutes at 85°C followed by 20 minutes at 42°C. A first strand mastermix was prepared using 0.5 µL E1 enzyme and 9.5ul First Strand cDNA Synthesis Mix 2 (FS2) per reaction. This FS2 mastermix was then incubated for 2-3 minutes at 42°C before 10ul of this was added to each reaction, and the plate was incubated for a further 15 minutes at 42°C. Following this step, 5ul of RNA removal reagent was added and mixed before incubation for 10 minutes at 95°C and then 25°C for 25 minutes. Second strand synthesis was achieved through the application of 10 µL SS1 reagent to each sample followed by incubation for 1 minute at 98°C and then down to 25°C using a modified ramp speed of 0.5°C/second. Samples were then incubated at this temperature for 30 minutes. While this incubation was in process, a mastermix containing 4 µL Second Strand Synthesis Mix 2 (SS2) and 1 µL Enzyme mix (E2) per sample was prepared, and 5ul of this mastermix was added to each well prior to 15 minutes incubation at 25°C. Libraries were purified by mixing 16 µL of Purification beads with each sample and placing the plate onto a magnet which binds the magnetic beads to the wall of each sample well. Following 5 minutes incubation, the supernatant was aspirated without disturbing the bound beads. While the plate was still on the magnet, 40ul of Elution buffer was added, and then the plate removed from the magnet and the beads mixed with the Elution buffer and incubated for 2 minutes. To each sample, 56 µL of Purification solution was then incubated for 5 minutes before the plate was returned to the magnet for a further 5 minutes. At this time, the supernatant was once again aspirated and disposed, and the pellets from each sample washed twice with 80% ethanol for 30 seconds each. Purified pellets were then allowed to dry (for no more than 10 minutes to avoid overdrying), and a final 20 µL of Elution buffer was used to resuspend the bead pellets. Following a 2 minute incubation, the plate was placed back onto the magnet for 5 minutes and 17 µL of purified sample was removed by pipette and placed into a separate fresh 96 well plate. In order to establish the number of cycles to optimally amplify our libraries, a test PCR was performed using the PCR add-on module from Lexogen. A mastermix was prepared using 7 µL of PCR Mix, 5 µL of P7 primer, 1 µL of Enzyme mix, 1.2 µL of 2.5x SYBR Green I and 14.1 µL of Elution buffer. To 1.7 µL of each sample, we added 28.3 µL of mastermix and then incubated at 98°C for 30 seconds before running in a thermocycler for 35 cycles of 98°C for 30 seconds, 65°C for 20 seconds and 72°C for 10 seconds. Finally, the samples were incubated for 1 minute at 72°C. We used this data to determine the optimal number of cycles to use in the final library amplification, based upon the number of cycles required to achieve 50% of the maximum signal, subtracting 3 samples to determine 16 cycles as optimal. A final PCR mastermix was then prepared using 7 µL PCR mix and 1 µL of Enzyme mix 3 and 8 µL of this applied to each well. Following this, 5 µL of index was added to each library being careful not to contaminate libraries with the wrong index, or the indexes with any nuclear material. These final libraries were then amplified for 16 cycles of final PCR amplification using the same conditions stated for the PCR add on protocol above. Samples were then cleaned using bead purification following the same purification steps as stated above, but using 30 µL each of Purification beads, Elution buffer, and Purification buffer. The 17 µL of library eluted at the end of this purification was subject to quality control using the Qubit so that they could be pooled at a concentration of 35fM each. Pooled libraries were diluted to 2nM in Elution buffer, and then denatured using 0.2N NaOH for 5 minutes, followed by incubation in 200mM Tris-HCl (pH7.0) for 5 minutes.

Libraries were sequenced using the Illumina NextSeq 550 using the Mid V2.5 chemistry; libraries were sequenced using 75 cycles. Quality control on the sequencer revealed 87.74% of reads >Q30 and 76.68% passed filter. Each sample yielded between 5-10Million reads. All bioinformatics were performed either on the Centre for Advanced Computing (CAC) Frontenac server, or on local machines running R (R Foundation for Statistical Computing, Vienna, Austria). The data analysis follows the following pipeline:

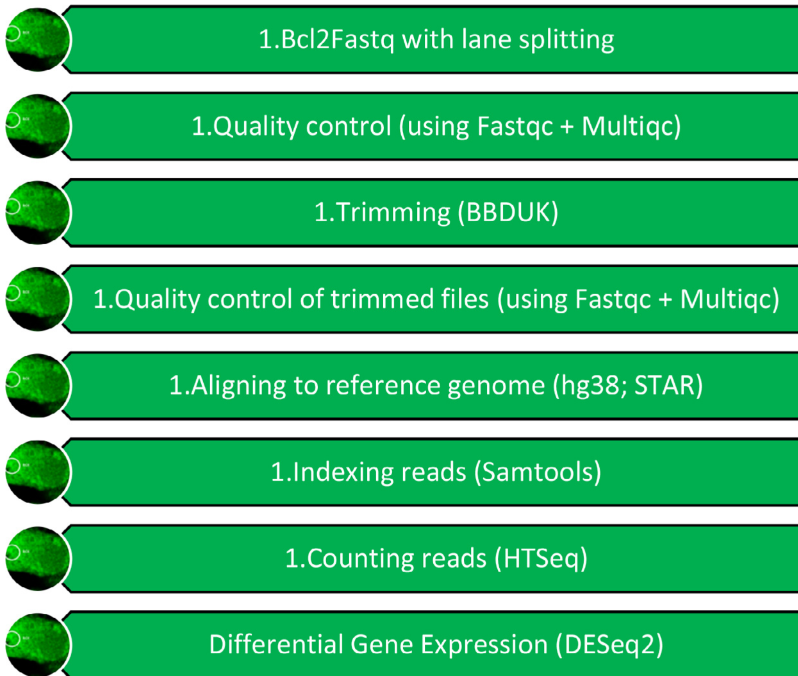


Table 1. Flow diagram to describe the processing of raw data from the Illumina Sequencing through to the counted reads presented here.

Briefly, BCL2 files from the Illumina were reconstructed using Bcl2fastq per sample based on Index ID provided in a SampleSheet. Fastq files were assembled from bcl2 output without lane split, and subject to quality control per-sample with fastqc and across samples using multiqc. Low quality reads and adapters were trimmed from reads using BBDUK (https://jgi.doe.gov/) [2], using the following parameters:

```
in=$sample\.fastq.gz \
out=$sample\trimmed_clean.fastq.gz \
ref=polyA.fa.gz \
ref=truseq_rna.fa.gz \
k=13 \
ktrim=r \
useshortkmers=t \
mink=5 \
qtrim=r \
trimq=10 \
minlength=20 \
gcbins=auto\
```

Each trimmed sample was then subject to quality control (using fastqc), and compared to original output using multiqc. Subsequent to this, each sample was aligned to the NCBI human genome (HG38) using STAR [3] using the following parameters:

```
-runThreadN 24 \
-genomeDir / human/index \
-readFilesIn sample.fastq \
-outFilterType BySJout \
```

```
-outFilterMultimapNmax 20 \
-alignSJoverhangMin 8 \
-alignSJDBoverhangMin 1 \
-outFilterMismatchNmax 999 \
-outFilterMismatchNoverLmax 0.6 \
-alignIntronMin 20 \
-alignIntronMax 1000000 \
-alignMatesGapMax 1000000 \
-outSAMattributes NH HI nM MD \
-outSAMtype BAM SortedByCoordinate
```

Reads were indexed using Samtools prior to counting with HTSeq-count [4]. Subsequent analysis (presented in our co-submission) of this data was performed by piping the counts files through DESeq2 using Liklihood ratio test (LRT).

Ethics Statement

Ethics approval was obtained at each university where this research was performed. Human DA samples were isolated during the course of congenital heart surgery at either University of Chicago (IRB number A3523-01) or University of Nebraska (IRB number 100-11-EP). Ethics approval from Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board (HSREB) for the continued use of the human cell lines in ongoing research (TRAQ #6007784).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

CRediT Author Statement

Rachel E.T. Bentley: Methodology, Formal analysis, Writing – original draft, Writing – review & editing; **Charles C.T. Hindmarch:** Methodology, Formal analysis, Writing – original draft, Writing – review & editing; **Kimberly J. Dunham-Snary:** Methodology, Writing – review & editing; **Brooke Snetsinger:** Methodology, Writing – review & editing; **Jeffrey D. Mewburn:** Methodology, Writing – review & editing; **Arthur Thébaud:** Methodology, Writing – review & editing; **Patricia D.A. Lima:** Methodology, Writing – review & editing; **Bernard Thébaud:** Writing – review & editing; **Stephen L. Archer:** Conceptualization, Writing – original draft, Writing – review & editing.

Acknowledgments

This work was supported by a Foundation grant from the Canadian Institutes for Health Research (CIHR) and grants from the National Institutes of Health (R01-HL071115, 1RC1HL099462), Canada Foundation for Innovation (229252 and 33012), a Tier 1 Canada Research Chair in Mitochondrial Dynamics and Translational Medicine (950-229252), a grant from the William M. Spear Endowment Fund in Pulmonary Research and the Richard K. Start Memorial Fund and a grant from the William J Henderson Foundation. The research is supported by the Queen's Cardiopulmonary Unit (QCPU).

References

[1] R.E.T. Bentley, C.C.T. Hindmarch, K.J. Dunham-Snary, B. Snetsinger, J.D. Mewburn, A. Thébaud, P.D.A. Lima, B. Thébaud, S.L. Archer, The molecular mechanisms of oxygen-sensing in human ductus arteriosus smooth muscle cells: a comprehensive transcriptome profile reveals a central role for mitochondria, Genomics 113 (2021) 3128–3140.
[2] BBTools User Guide, in, Joint Genome Institute.
[3] B. van de Geijn, G. McVicker, Y. Gilad, J.K. Pritchard, WASP: allele-specific software for robust molecular quantitative trait locus discovery, Nat. Methods 12 (2015) 1061–1063.
[4] S. Anders, P.T. Pyl, W. Huber, HTSeq–a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169.