

Supplementary Data

for

Depletion of hemoglobin transcripts and long read sequencing improves the transcriptome annotation of the polar bear (*Ursus maritimus*)

Ashley Byrne^{1,2}, Megan A. Supple³, Roger Volden^{2,4}, Kristin Laidre⁵, Beth Shapiro^{3,6}, Christopher Vollmers^{2,4}

Includes:

Figures S1-S4

Tables S1 and S2

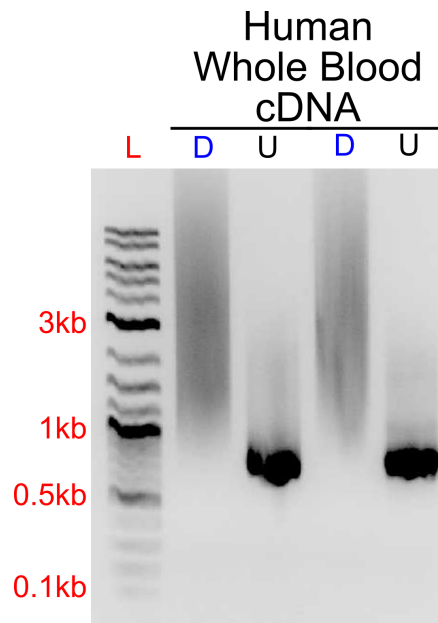


Figure S1: Long-DASH sgRNA also depletes human hemoglobin transcripts from full-length cDNA. Technical replicates of depleted (D) or undepleted (U) human whole blood cDNA were visualized on an agarose gel. DNA ladder (L) suggests highly abundant cDNA species - putatively hemoglobin around ~700bp.

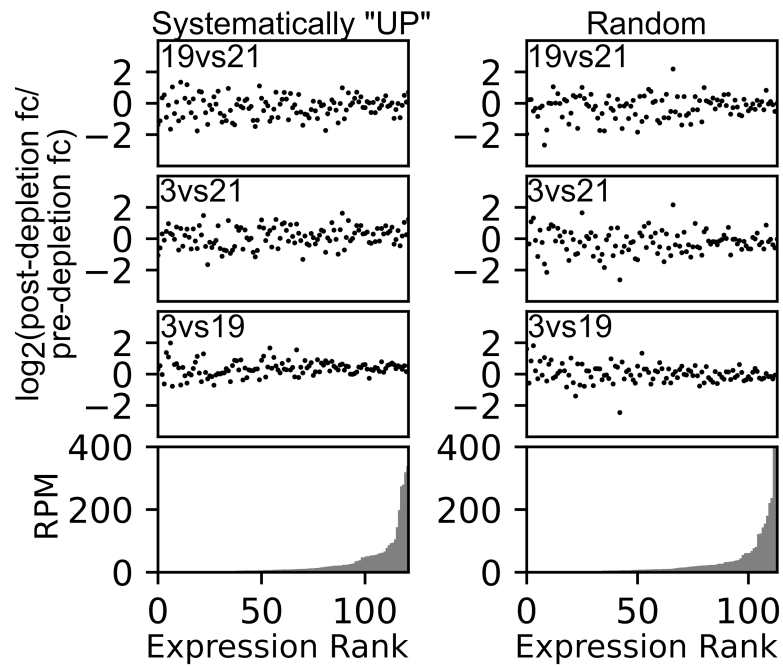


Figure S2: Depletion of hemoglobin affects expression levels within but not fold-change between samples. Expression levels (bottom) as well as the changes in differential expression between polar bears pre- and post-depletion are shown for genes that are upregulated in all 3 polar bears post-depletion (Systematically "UP") and a random selection of genes with similar expression distribution (Random).

A) Hemoglobin, alpha (HBA)

XM_008696690.1	-----CCGCCCGCACATTTCTGGTCTCACAGACTCAGAAAGAA	40
NM_000558.4	CATAAACCTGGCGGCTCGCGGCCGGGCACTTCTGGTCCCCACAGACTCAGAGAGAA	60
NM_000517.4	CATAAACCTGGCGGCTCGCGGCCGGGCACTTCTGGTCCCCACAGACTCAGAGAGAA	60
	* * * * *	
XM_008696690.1	GCCACCATGGTGCTGTCTCCCGCCGACAAGAGCAACGTCAAGGCCACCTGGGATAAGATC	100
NM_000558.4	CCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCTGGGTAAGGTC	120
NM_000517.4	CCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCTGGGTAAGGTC	120

XM_008696690.1	GGCAGCCACGCTGGCGAGTATGGCGGCGAGGCTCTGGAGAGGACCTTCGCGTCTTCCCC	160
NM_000558.4	GGCGGCGACGCTGGCGAGTATGGTGGCGAGGCCCTGGAGAGGATGTTCTGCTTCCCC	180
NM_000517.4	GGCGGCGACGCTGGCGAGTATGGTGGCGAGGCCCTGGAGAGGATGTTCTGCTTCC	180
	*** *****	
XM_008696690.1	ACCACCAAGACCTACTTCCCCACTTCGACCTGAGCCCTGGCTCCGCCAGGTCAAGGCC	220
NM_000558.4	ACCACCAAGACCTACTTCCCCACTTCGACCTGAGCCACGGCTCTGCCAGGTTAAGGGC	240
NM_000517.4	ACCACCAAGACCTACTTCCCCACTTCGACCTGAGCCACGGCTCTGCCAGGTTAAGGGC	240

XM_008696690.1	CACGGCAAGAAGGTGGCCGACGCCCTGACCACCGCCGAGGCCACCTGGACGACCTGCCG	280
NM_000558.4	CACGGCAAGAAGGTGGCCGACGCCCTGACCACCGCCGAGGCCACCTGGACGACATGCC	300
NM_000517.4	CACGGCAAGAAGGTGGCCGACGCCCTGACCACCGCCGAGGCCACCTGGACGACATGCC	300

XM_008696690.1	GGCGCCTGTCCGCTCTGAGGACCTGCACGCGCACAAAGCTGCGAGTGGACCCGGTCAAC	340
NM_000558.4	AACGCGCTGTCCGCTCTGAGGACCTGCACGCGCACAAAGCTTGGGTGGACCCGGTCAAC	360
NM_000517.4	AACGCGCTGTCCGCTCTGAGGACCTGCACGCGCACAAAGCTTGGGTGGACCCGGTCAAC	360
	*** *****	
XM_008696690.1	TTCAAGTTCCTGAGCCACTGCCTGCTGGTGACCCTGGCCAGCCACCACCCGCGGAGTTC	400
NM_000558.4	TTCAAGTTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCAGCCACCACCCGCGGAGTTC	420
NM_000517.4	TTCAAGTTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCAGCCACCACCCGCGGAGTTC	420

XM_008696690.1	ACCCCTGCCGTCCACGCCTCCCTGGACAAGTTCCTCAGCGCCGTGAGCACCGTCTCACC	460
NM_000558.4	ACCCCTGCCGTGCACGCCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTCTGACC	480
NM_000517.4	ACCCCTGCCGTGCACGCCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTCTGACC	480

XM_008696690.1	TCCAAATACCGTTAAGCTGGAGCCGCGGACCCCTCCCGCTCCCGGCTGGGGCCTCTTGC	520
NM_000558.4	TCCAAATACCGTTAAGCTGGAGCCCTGGTGGCCATGCTTCTTGGCCCTTGGGCCTCCCC	540
NM_000517.4	TCCAAATACCGTTAAGCTGGAGCCCTGGTGGCCATGCTTCTTGGCCCTTGGGCCTCCCC	540

XM_008696690.1	GC-----TCCACGCGCTGAACCTCCCGATCTTTGAATAAAGTCTGAGTGG	566
NM_000558.4	CAGCCCCTCCTCCCCTTCTGCACCCGTACCCCGTGGTCTTTGAATAAAGTCTGAGTGG	600
NM_000517.4	CGGGCCCTCCTCCCCTCCTTGACC - GGCCCTTCTGGTCTTTGAATAAAGTCTGAGTGG	599
	* * * * *	
XM_008696690.1	GCTGCAG-----	573
NM_000558.4	GCGCAAAAAAAAAAAAAAAAAAAAAA	627
NM_000517.4	GCAGCAAAAAAAAAAAAAAAAAAAA---	622
	** **	

A) Construction of sgRNAs

#1 tracrRNA oligo (in Reverse Orientation/Anti-sense)

```
<-----tracrRNA-----> <-----Primer----->
5'-AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTATTTAACTT GCTATTTCTAGCTCTAAAAC-3'
```

The oligo above is a universal tracrRNA template which allows you to generate full sgRNA templates with any target sequence oligo as long as the target sequence oligo meets the following requirements below:

1. The oligo contains the reverse complement of the primer sequence on the 3' end.
2. the oligo contains 'GG' on the 3' end of the target sequence for T7 RT.

#1 HBA1/HBA2 target oligos based upon mRNA prediction of Polar Bear/Human genes.

```
<-----T7-----><--target_sequence--><-----Primer----->
oligo 1 5'-GAAATTAATACGACTCACTATAGG AAGGCCACGGCAAGAAGG GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 2 5'-GAAATTAATACGACTCACTATAGG CACTGCCTGCTGGTGACCC GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 3 5'-GAAATTAATACGACTCACTATAGG GGTYAAGGSCCAGGCAAGA GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 4 5'-GAAATTAATACGACTCACTATAGG ACCTCCAAATACCGTAAAGC GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 5 5'-GAAATTAATACGACTCACTATAGG GCCGACAAGASCAACGTCA GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 6 5'-GAAATTAATACGACTCACTATAGG GGGAAGTAGTCTTGGTGGTG GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 7 5'-GAAATTAATACGACTCACTATAGG TCCTRAGCCACTGCCTGC GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 8 5'-GAAATTAATACGACTCACTATAGG CAGGTCGCTCAGRGGGACA GTTTTAGAGCTAGAAATAGC-3'
```

#2 HBB target oligos based upon mRNA prediction of Polar Bear/Human gene.

```
<-----T7-----><--target_sequence--><-----Primer----->
oligo 1 5'-GAAATTAATACGACTCACTATAGG CACTGTGACAAGCTGCACG GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 2 5'-GAAATTAATACGACTCACTATAGG GAAGTTGGTGGTGAGGCCCT GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 3 5'-GAAATTAATACGACTCACTATAGG CAGGCTGCCTATCAGAARG GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 4 5'-GAAATTAATACGACTCACTATAGG GCAACWCWCAASAGACACCA GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 5 5'-GAAATTAATACGACTCACTATAGG GAGGTTCTTTGABTCCTTTG GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 6 5'-GAAATTAATACGACTCACTATAGG AAGGTGAACGTGGATGAAGT GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 7 5'-GAAATTAATACGACTCACTATAGG GCTCCTGGCAACGTGC GTTTTAGAGCTAGAAATAGC-3'
<-----T7-----><--target_sequence--><-----Primer----->
oligo 8 5'-GAAATTAATACGACTCACTATAGG CAGAARGTGGTGGCTGGTG GTTTTAGAGCTAGAAATAGC-3'
```

B) Target sequences and oligo sequences for hemoglobin depletion

HBA targets	HBB targets
(1) AAGG S CCACGGCAAGAAGG Human- AAGGGCCACGGCAAGAAGG Polar- AAGGCCACGGCAAGAAGG	(1) CACTGTGACAAGCTGCACG Human- CACTGTGACAAGCTGCACG Polar- CACTGTGACAAGCTGCACG
(2) CACTGCCTGCTGGTGACCC Human- CACTGCCTGCTGGTGACCC Polar- CACTGCCTGCTGGTGACCC	(2) GAAGTTGGTGGTGAGGCCCT Human- GAAGTTGGTGGTGAGGCCCT Polar- GAAGTTGGTGGTGAGGCCCT
(3) GGT V AAGGSCCAGCAAGA Human- GGTTAAGGGCCACGGCAAGA Polar- GGTC AAGGCCACGGCAAGA	(3) CAGGCTGCCTATCAGAA R G Human- CAGGCTGCCTATCAGAAAG Polar- CAGGCTGCCTATCAGAAG
(4) ACCTCCAAATACCGTTAAGC Human- ACCTCCAAATACCGTTAAGC Polar- ACCTCCAAATACCGTTAAGC	(4) GCAAC C CAAA S AGACACCA Human- GCAACCTCAAACAGACACCA Polar- GCAACCACAAAGAGACACCA
(5) GCCGACAAGAS S CAACGTCA Human- GCCGACAAGACCAACGTCA Polar- GCCGACAAGAGCAACGTCA	(5) GAGGTTCTTTGAB T CTCTTTG Human- GAGGTTCTTTGAGTCTCTTTG Polar- GAGGTTCTTTGACTCTCTTTG
(6) GGGAAGTAGGTCTTGGTGGTG (r) Human- GGGAAGTAGGTCTTGGTGGTG Polar- GGGAAGTAGGTCTTGGTGGTG	(6) AAGGTGAACGTGGATGAAGT Human- AAGGTGAACGTGGATGAAGT Polar- AAGGTGAACGTGGATGAAGT
(7) TCCT R AGCCACTGCCTGC Human- TCCTAAGCCACTGCCTGC Polar- TCCTGAGCCACTGCCTGC	(7) GCTCCTGGGCAACGTGC Human- GCTCCTGGGCAACGTGC Polar- GCTCCTGGGCAACGTGC
(8) CAGGTCGCTCAG R GCGGACA (r) Human- CAGGTCGCTCAGGGCGGACA (Lost PAM sequence) Polar- CAGGTCGCTCAGAGCGGACA	(8) CAGAA R GTGGTGGCTGGTG Human- CAGAAAGTGGTGGCTGGTG Polar- CAGAAGTGGTGGCTGGTG

Figure S4. sgRNA design and construction.

Oligonucleotides designed for hemoglobin depletion from full-length cDNA. Oligos were chosen to deplete Hemoglobin mRNA transcripts from Human and Polar Bear whole blood. A) To generate sgRNAs a template free PCR was performed to anneal the tracrRNA oligo to an oligo containing the target sequence to generate a full-length oligo. The full-length oligos were converted into sgRNA templates using in-vitro transcription. B) Target oligos used for generating sgRNAs. Degenerate bases are highlighted in grey. (r) indicates reverse orientation

Sample Name	Sequencer	Run Type	Library Prep	Read Number	Median Accuracy
PB3_depleted	Illumina HiSeqX	2X151	Smart-seq2	22322746	N/A
PB19_depleted	Illumina HiSeqX	2X151	Smart-seq2	19418607	N/A
PB21_depleted	Illumina HiSeqX	2X151	Smart-seq2	22467660	N/A
PB3_undepleted	Illumina HiSeqX	2X151	Smart-seq2	58088942	N/A
PB19_undepleted	Illumina HiSeqX	2X151	Smart-seq2	105936701	N/A
PB21_undepleted	Illumina HiSeqX	2X151	Smart-seq2	63096050	N/A
PB19_undepleted	ONT MinION	9.4.1/LSK109	R2C2	5313	93%
PB3_depleted_R1	ONT MinION	9.4.1/LSK109	R2C2	390526	93%
PB3_depleted_R2	ONT MinION	9.4.1/LSK109	R2C2	1691780	94%
PB19_depleted_R1	ONT MinION	9.4.1/LSK109	R2C2	59097	93%
PB19_depleted_R2	ONT MinION	9.4.1/LSK109	R2C2	866087	97.5%
PB21_depleted	ONT MinION	9.4.1/LSK109	R2C2	830952	92%

Table S1: High-throughput sequencing runs and characteristics

For R2C2/ONT MinION runs, fully processed R2C2 read numbers and median accuracies are given. Some R2C2/ONT MinION runs were multiplexed, sometimes with samples unrelated to this study. Samples PB19_depleted_R2 and PB3_depleted_R2 represent the current output of the R2C2/ONT MinION combination

RT

```
>Oligo-dT-smartseq2  
/5Me-isodC/AAGCAGTGGTATCAACGCAGAGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTVN  
>TSO_Smartseq2  
AAGCAGTGGTATCAACGCAGAGTACATrGrGrG
```

Primers for amplifying cDNA

```
>ISPCR  
AAGCAGTGGTATCAACGCAGAGT
```

Tn5 Oligos (Smart-seq2 library prep)

```
>Tn5ME-R  
[phos]CTGTCTCTTATACACATCT  
>Tn5ME-A  
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG  
>Tn5ME-B  
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
```

Primers for amplifying Tn5 Product

```
>Nextera_Primer_A  
AATGATACGGCGACCACCGAGATCTACAC [8bp i5 index] TCGTCGGCAGCGTCAGATG  
>Nextera_Primer_B  
CAAGCAGAAGACGGCATACGAGAT [8bp i7 index] GTCTCGTGGGCTCGGAGATGTGTAT
```

R2C2 Splint_Oligos

```
>UMI_Splint_1_Forward_ISPCR (Matches ISPCR Primer)  
ACTCTGCGTTGATACCACTGCTTTGAGGCTGATGAGTTCCATANNNNNATATANNNNNATCACTACTTAGTTTTTGGATAGCTTC  
AAGCCAGAGTTGTCTTTTTCTCTTTGCTGGCAGTAAAAG  
>UMI_Splint_1_Reverse_ISPCR (Matches ISPCR Primer)  
ACTCTGCGTTGATACCACTGCTTAAAGGGATATTTTCGATCGCANNNNNATATANNNNNTTAGTGCATTTGATCCTTTACTCCTC  
CTAAAGAACAACCTGACCCAGCAAAAGGTACACAATACTTTACTGCCAGCAAAGAG  
>UMI_Splint_2_Forward_ISPCR (Matches ISPCR Primer)  
ACTCTGCGTTGATACCACTGCTTTGCCGTTGGGTATCAATAANNNNNATATANNNNNATTGCCTTTATTCTATCTACTTAGTTT  
TGGCGATGTAGTCTACCTATCCTGATGCTGAATAAAGGC  
>UMI_Splint_2_Reverse_ISPCR (Matches ISPCR Primer)  
ACTCTGCGTTGATACCACTGCTTAATTAGTTCTAGGATCACGNNNNNATATANNNNNCTGCCATCGAAAATTTTACCCGTAA  
CAAGAACTTACAACCTCTGACGCCTATATCATGAAGGCCTTTATTTCAGCATCAGGA
```

Table S2 Oligos used in the manuscript

All oligos are shown 5'->3' and were ordered from Integrated DNA Technologies (IDT). Lower case 'r' indicates RNA bases. Spaces in sequence