SUPPORTING INFORMATION

Digital encoding of cellular mRNAs enables precise and absolute gene expression measurement by single-molecule counting

Glenn K Fu*, Julie Wilhelmy, David Stern, Christina Fan & Stephen P A Fodor*

Cellular Research Inc., 3183 Porter Drive, Palo Alto, CA 94304, USA

Corresponding Author

*Email: gfu@cellular-research.com or sfodor@cellular-research.com. Tel: (+1) 650 7526144

Contents:

Experimental methods

Cell culture and single-cell isolation

Human leukemia K562 cells (ATCC) were grown and maintained in RPMI 1640 culture medium (GIBCO) supplemented with 10% heat inactivated fetal bovine serum (GIBCO) and 1X antibiotic-antimycotic (GIBCO). The cultures were incubated at 37°C in 5% CO_2. For hemin induction, K562 cells were added to fresh culture media supplemented with 0.05 mM hemin (Sigma-Aldrich) and incubated for 45 - 72 hours. Single cells were washed and isolated by limiting dilution to approximately 1 cell per µL in calcium-magnesium free phosphate buffered saline (GIBCO). One (1) µL of this dilution was added to several 0.2 mL PCR tubes and the presence of a single cell was verified by visual inspection under an inverted microscope. Each single-cell sample was frozen at -80°C until further processing. Prior to cDNA synthesis, 8 µL of 0.7% Triton X-100 was added to each cell and the reaction tube was incubated at 65°C for 5 min for cell lysis. The sample was mixed by vortexing and placed on ice.

RNA samples and controls

Human liver total RNA was purchased from Ambion and human normal lymphocyte total RNA was purchased from Biochain. RNA experiments were carried out using low nucleic acid binding tubes (Ambion) and tips (Rainin). To reduce the non-specific binding even further, each tube was pre-rinsed with a solution of 10 ng/µL yeast RNA (Ambion). All RNA dilutions were performed in a solution of 1 ng/µL *E.coli* total RNA (Roche) as a carrier. Volumes below 5 µL were avoided to reduce pipetting errors. For control experiments, a 1.2 kb synthetic Kanamycin polyadenylated RNA (Promega) was used. Control spike-in RNAs for sequencing experiments were a mixture of polyadenylated *B. subtilis* transcripts, *Dap, Phe, Thr* and *Lys* (Affymetrix).

Barcode and primer design

Barcodes consist of 21-base oligonucleotide tags (Supporting information, Table S3)[1]. A set of 960 deoxyuridine modified oligo dT primers each containing a distinct barcode tag and a universal PCR binding sequence (AGCACGACAGACGCCTGAT), was synthesized and pooled in equimolar concentration (Cellular Research). A second strand (SS) cDNA synthesis primer was selected 200 to 500 bp from the 3' end of each mRNA tested. A forward gene-specific PCR primer (F1) was selected 50 to 150 bp downstream of the SS primer. A nested primer (F2) was selected 50 to 150 bp downstream of the F1 primer for the detection reaction. To design primers using the NCBI Primer3 and primer BLAST program (http://www.ncbi.nlm.nih.gov/tools/primer-blast/), the reverse complement to the universal PCR binding site was appended to the 3' end of each transcript and the entire sequence was input into the program. Gene-specific forward primers were selected using the universal PCR binding sequence as the reverse primer. Default primer selection parameters were applied except for the following modifications: melting temperatures ranged from 59°C to 65°C, primer GC content ranged from 40 - 60%. One of the

high abundance genes (RPLP0) was assayed in duplicate using an alternate set of primers which resulted in equivalent counts. All gene assays, except for ABCF1, were successful with the initial set of selected primers. Two 4 bp runs of primer dimer pairs were present in the original ABCF1 F1 primer, which when omitted resulted in a successful assay. Although initial experiments included the use of deoxyuridine substitutions on the second strand primer, its use was subsequently found to be unnecessary and has been excluded in later experiments. This optional strategy may be reserved for troubleshooting difficult situations where primer interference with downstream steps is suspected. Genes for multiplex targeted sequencing were chosen from popular human qPCR gene expression assays. Resulting F1 PCR amplicons were 362 to 450 bp in length, and F2 PCR amplicons were 210 to 320 bp in length. All gene IDs and primer sequences are listed (Supporting information, Table S1 and S4).

cDNA synthesis and molecular indexing

Eight (8) µL of sample (isolated RNA or crude cell lysate) was mixed into a tube containing 1 µg of *E. coli* carrier total RNA (Roche), 1 µL of 10 mM dNTP mix (NEB), 1 µL of 4 µM SS primer and 1 µL of 4 µM barcoded oligo-dT primer pool (Cellular Research) and incubated at 65°C for 5 min. The reaction was placed on ice and the following reverse transcription master mix was added: 2 µL 10X MMLV buffer (NEB), 10 units SUPERase-In (Ambion), 200 units MMLV reverse transcriptase (NEB) and 2 units Titanium Taq polymerase (Clontech) for a total reaction volume of 20 µL. The reaction was incubated at 37°C for 20 min followed by second strand synthesis: 94°C for 2 min, 55°C for 2 min, 68°C for 2 min. To maximize efficiency, this second strand reaction is repeated twice. The high annealing temperature of the second strand reaction (55°C) prevents un-intended excess barcode incorporation (dT_{17} T_m = 44°C). This procedure copies deoxyuridines on the barcoded dT primer into adenines on the second strand cDNA. All of the above reactions occur in a single tube by modulating temperature, and do not require additional pipetting steps. The reverse transcription reaction contains both the reverse transcriptase enzyme and the thermophillic DNA polymerase, but the latter is initially inactivated by antibody binding and is not active during the cDNA synthesis step. After heat inactivation of the mesophilic reverse transcriptase, the DNA polymerase is hot-start activated and begins to carry out second strand synthesis. To inactivate excess primers, 2.5 units of Uracil DNA Glycosylase (UDG) (NEB) were added and the reaction was mixed and transferred to a new tube and incubated at 37°C for 15 min. No purification procedure is necessary because abasic primers are prevented from participating and interfering in downstream steps, thereby increasing the specificity and efficiency of those steps.

Singleplex mRNA transcript counting

A 5 µL portion of the cDNA reaction was added to a 15 µL PCR master mix consisting of 0.75X Titanium Taq buffer, 0.15 µM dNTP mix, 0.05 µM gene specific F1 primer, 0.05 µM universal reverse primer and 0.75X Titanium Taq DNA polymerase. The reaction was incubated at 94°C for 2 min, followed by 30 cycles of: 94°C for 20 sec, 58°C for 20 sec, 68°C for 20 sec; followed by 68°C for 4 min. One (1) µL of the

PCR product was added into a 50 µL detection reaction consisting of 1X Taq buffer (NEB), 0.2 mM dNTP mix, 0.2 µM gene specific F2 primer, 0.2 µM Tye-563 labelled universal reverse primer and 2.5 units Taq DNA polymerase (NEB). The reaction was incubated at 94°C for 2 min, followed by 30 cycles of: 94°C for 20 sec, 55°C for 20 sec, 72°C for 20 sec; followed by 72°C for 4 min. Four (4) µL of the product was analyzed on a polyacrylamide 4-20% gradient TBE gel (Invitrogen) to assess size and purity.

Multiplex mRNA transcript counting

To simultaneously assay multiple genes in a single reaction, the molecular indexing PCR protocol was performed with the following changes: Primers for all genes were combined and used together in each step. The annealing step in the first amplification reaction was extended from 20 sec to 90 sec. After multiplex PCR, detection reactions for individual genes were then performed independently in separate F2 PCR reactions.

Imaging system

The PIXEL instrument (Cellular Research) is an automated fluorescence imaging system for 16-well slides (Figure 1B-C). The imaging optical train consists of an achromatic cemented doublet lens (85 mm focal length, Edmund Optics 47640), an emission bandpass filter (Semrock FF01-593/40-25), a camera lens (25 mm focal length, Fujinon HF25HA-1B), and a 1.2-megapixel CCD camera (Point Grey Research CMLN-13S2M-CS). CCD size is 4.8 mm x 3.6 mm, of which the central 4.8 mm x 2.4 mm area is used. The magnification of the imaging optical train is 0.3. Therefore the instrument's field of view is 16 mm x 8 mm, and two adjacent wells of a 16-well slide can be imaged simultaneously. Exposure time for fluorescence images is typically 1 sec. Fluorescence excitation is provided by an off-axis (45 degree incidence) Kohler illuminator consisting of a 525 nm, 10 watt LED (LedEngin LZ4-40G100), an aspheric collimating lens (20 mm focal length, Thorlabs ACL2520-A), a rectangular aperture (19 mm x 7.5 mm), a bandpass filter (Semrock FF01-531/40-25), and three plano-convex lenses (60 mm focal length, Edmund Optics 47347). The 16-well slide is mounted on a single-axis motorized translation stage constructed from a stepping-motor-driven linear actuator (Haydon Kerk Z26441-05-151ENG) and a linear slide with recirculating ball bearings (Misumi SSELBW9-170).

Barcode detection and counting

Barcode detectors (Cellular Research) were constructed by inkjet printing of 5’ Amine modified oligonucleotides of complementary sequence to the oligo dT barcodes onto epoxysilane coated glass slides or plates. Sixteen (16) or 96 well chambers were applied to partition samples for hybridization. Dye labelled PCR product was added to an 80 µL solution of 100 pM Cy3 hybridization control oligo (Cellular Research) in Wash A buffer (Affymetrix) and incubated at 95°C for 5 min then placed on ice. The denatured PCR product was transferred into a well chamber on the detector slide for incubation at 37°C for 3 hrs followed by a single rinse in Wash B buffer (Affymetrix) at room temperature. The remaining

liquid was vacuum aspirated and the detector was either imaged on the FLAIR scanner (Sensovation AG) or the PIXEL instrument (Cellular Research). For each image, a detection threshold is computed to classify barcodes as either present or absent. The algorithm applies a lower (LL) and upper (UL) intensity threshold limit, and *w*, a window size. Each was set to default values that were determined from a reference experimental dataset. We let $y = [y_1, y_2, \ldots, y_n]$ represent the set of *n* background subtracted probe intensities where each *y* satisfies $LL \leq y \leq UL$, and sort the intensities in increasing order to obtain $y^* = [y_{(1)}, y_{(2)}, \ldots, y_{(n)}]$. The distance between each neighboring sorted probe intensity, $d = [d_1, d_2, d_i, \ldots, d_{n-1}]$ is calculated, where $d_i = y^*_{i+1} - y^*_i$. We then calculate a corresponding threshold statistic for each of the distances, *d*:

$$x_j = \frac{\sum_{i=j-w}^{j+w} d_i}{2w+1} \qquad \textit{Equation 1}$$

The threshold *C* is the probe intensity corresponding to the largest x_j statistic, i.e. y_{j_max}, where j_max corresponds to $x_{j\_max} = \max(x_j)$. The number of barcodes detected, *k*, corresponds to probes with intensities greater than the threshold *C*. Once *k* is determined, we can compute the number of molecules measured as described previously[2].

Barcode counting by sequencing

Poly-A RNA containing Dap, Lys, Phe and Thr controls (Affymetrix) was diluted $1{:}1x10^8$ into 0.3 fg per µL Kan control RNA (Promega). One (1) µL was added to 5 – 20 pg of various human total RNA samples mixed with 0.5 µg *E. Coli* RNA (added as a carrier) in each of 96 PCR plate wells. Five (5) nMoles of dNTP and 2 pMoles of barcoded oligo-dT primers were added into a final volume of 4.9 µL and reactions were incubated at 65°C for 3 min and then placed on ice. Two (2) µL water, 0.5 µL 0.1M DTT, 2 µL 5X first strand buffer, 2 units SUPERase-In (Life) and 60 units MMLV (NEB) were added and the reactions were incubated at 37°C for 30 min and 65°C for 20 min. One (1) µL of 2 µM each of pooled F1 primers, 1 µL of 2 µM sample indexing primer and 0.2 µL of Titanium Taq DNA polymerase (Clontech) were added and the reactions were incubated at 94°C for 2 min, 55°C for 3 min and 68°C for 2 min, and repeated for 2 more cycles. All 96 reactions were then pooled into a single tube, and 20 units of UDG (Affymetrix) was added and incubated for 30 min at 37°C. One (1) mL AMPure beads (Beckman Coulter) was added to purify the cDNAs, and air-dried beads were re-suspended in 250 µL of multiplex PCR reaction containing 1X Titanium Taq PCR buffer, 0.3 mM dNTPs, 50 nM each of pooled F1 primer, 1 µM iL_B_PCR reverse primer (Cellular Research) and 5 µL of Titanium Taq polymerase. PCR was carried out in 5 tubes of 50 µL each for 94°C 2min, and 15 cycles of 94°C for 30 sec, 55°C for 3 min, 68°C for 1 min; and 68°C for 7 min. An equal volume of AMPure was added to purify the PCR product, and DNA was eluted in 40 µL 10 mM Tris-HCl, pH 8. Ten (10) µL was used for a nested PCR in 100 µL of 1X Titanium PCR buffer, 50 nM of each F2 primer, 1 µM of iL_B_PCR primer and 2 µL of Titanium Taq polymerase. PCR condition was the same as above, except only 12 cycles were used, and products were purified as before. To prepare for sequencing, 5 µL purified product was amplified by PCR in 100 µL1X Taq PCR buffer, 0.2 mM dNTPs,

0.2 μM each IDX D0 and P1_long primers (Cellular Research) and 1 μL Taq polymerase (Affymetrix). PCR conditions were 94°C for 2 min, 3 cycles of 94°C for 15 sec, 52°C for 30 sec, 72°C for 30 sec; and 11 cycles of 94°C for 15 sec, 60°C for 30 sec, 72°C for 30 sec; followed by 72°C for 7 min. Products were purified with an equal volume of AMPure. Sequencing was performed on the Illumina MiSeq instrument. Reads were mapped to the 960 oligo-dT barcodes and to the 96 gene targets and 4 spike-in controls using BWA[3] or Bowtie 2[4]. No read minimums were applied as inclusion criteria for counting barcodes. A bash script was used to count the number of barcodes detected in the sequencing reads for each gene.

RNA-Seq analysis of bulk RNA sample

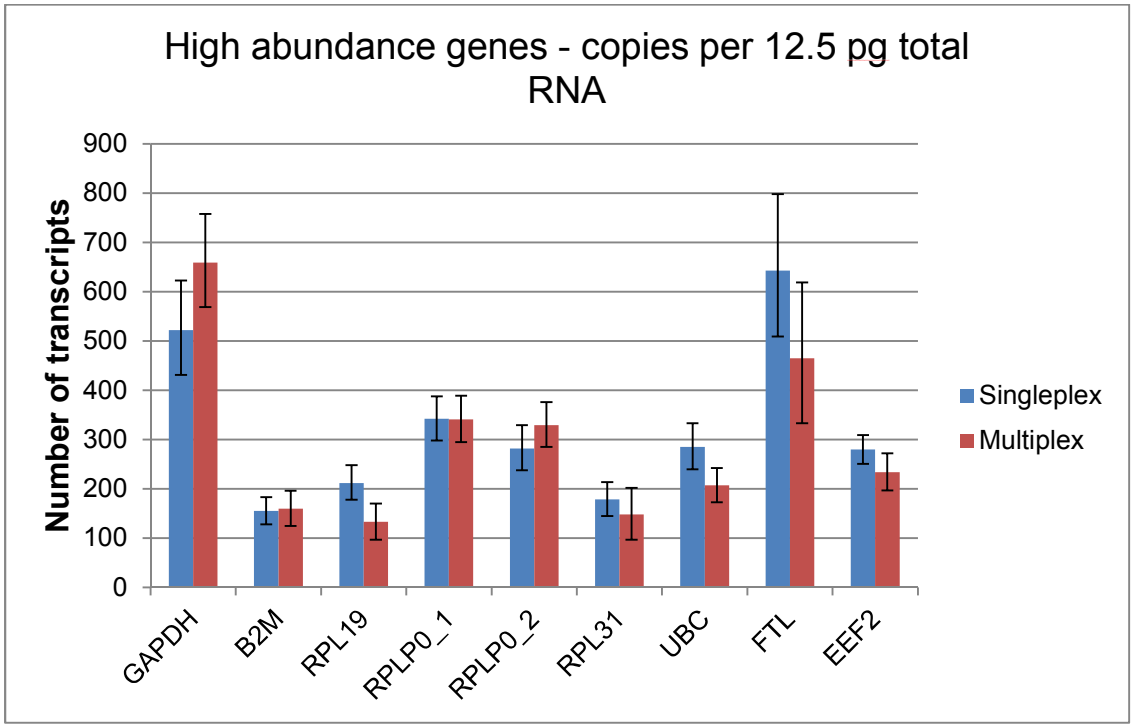
cDNAs were generated using the Illumina mRNA-Seq sample preparation kit through the polyadenylation step before using the NEXTflex qRNA-Seq kit (Cellular Research) to prepare sequencing libraries. Five hundred (500) ng of human normal lymphocyte total RNA was used as starting material and External RNA Controls Consortium (ERCC) RNAs (Ambion) were spiked into the input RNA sample to serve as an internal control. DNA sequencing was performed on a MiSeq instrument and paired-end reads were mapped using BWA to the HG19 Refseq RNA sequences downloaded from the UCSC Genome Browser (http://genome.ucsc.edu). RPKM values were calculated following published methods[5].

Determination of Template Switching cDNA synthesis efficiency

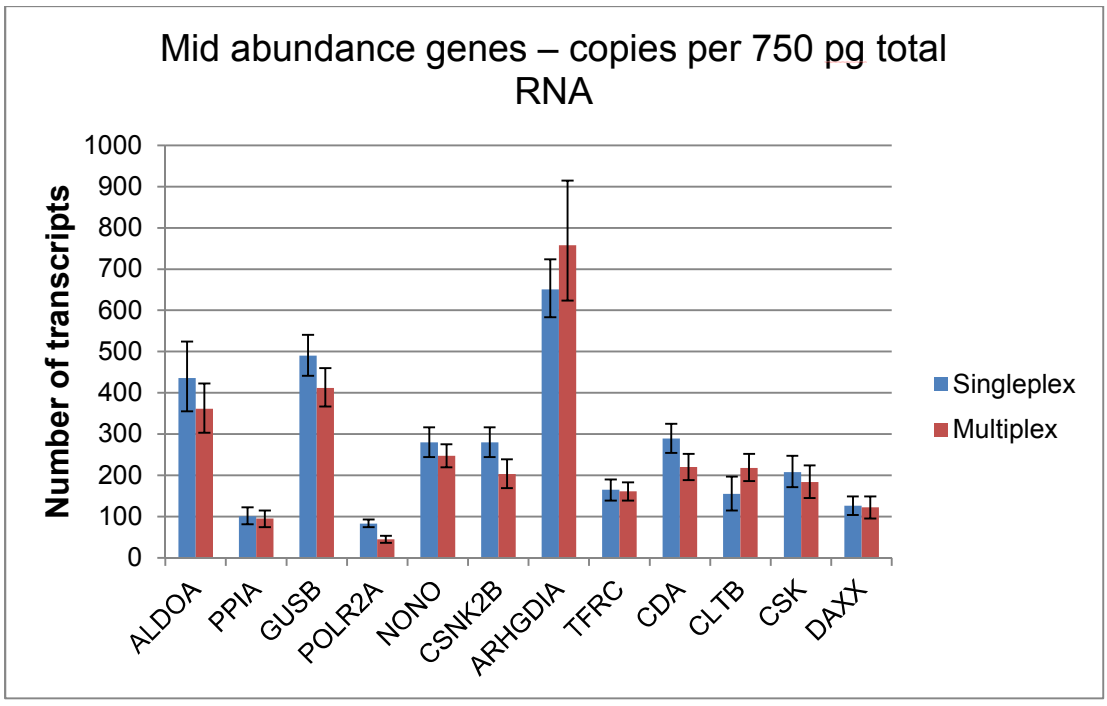
Synthetic poly-adenylated RNAs were used as input for strand switch cDNA synthesis[6]. Each 1,092 nt RNA molecule contains one of 960 possible 21 nucleotide molecular barcode tags between positions 728 – 748 (Supporting Information, Figure S6). Ten-fold serial dilutions of the barcoded RNA was reverse transcribed in 20 μL in 1 mM dNTP, 1 μM T20KanR1 RT primer, 1 μM Template switch oligonucleotide (TGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNACGCrGrGrG), 50 mM Tris-HCl pH 8.3, 6 mM MgCl2, 75 mM KCl, 5 mM DTT, 0.1 mg/ml BSA, 1 unit SUPERaseIn and 200 units SuperScript II (Life technologies) for 42°C for 90 min and inactivated at 70°C for 15 min. cDNAs were purified by Ampure XP Beads after TS and one-tenth of the cDNA was PCR amplified with either forward + reverse gene-specific primers, or strand switch primer + reverse gene-specific primer (Supporting Information, Figure S6). Amplification reactions were performed in 20 μL in 0.2 mM dNTP, 0.2 μM each primer, 1 unit Taq DNA polymerase in 1X PCR buffer (Affymetrix) for 30 cycles. Nested PCR was performed using a Cy3 labeled forward primer and a nested gene-specific primer, and barcodes were detected and counted as described.

Supporting Information Figures

A



B



C

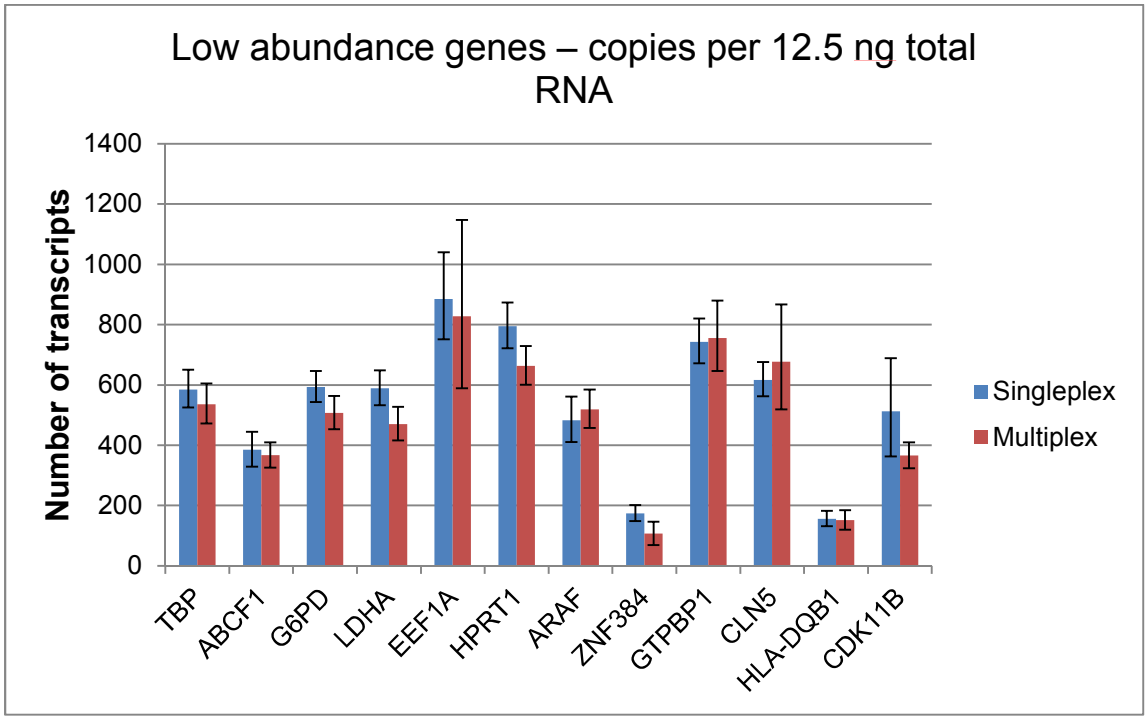


Figure S1. Singleplex (blue bars) and multiplex (red bars) molecular indexing PCR was performed on sets of high (**a**), medium (**b**) and low (**c**) abundance genes. In singleplex assays, one gene was amplified and measured at a time. In multiplex assays, all genes were amplified together before the detection of individual genes. Increasing amounts of liver total RNA was used in each experiment as indicated. Error bars show the 95% confidence intervals of the measurements.

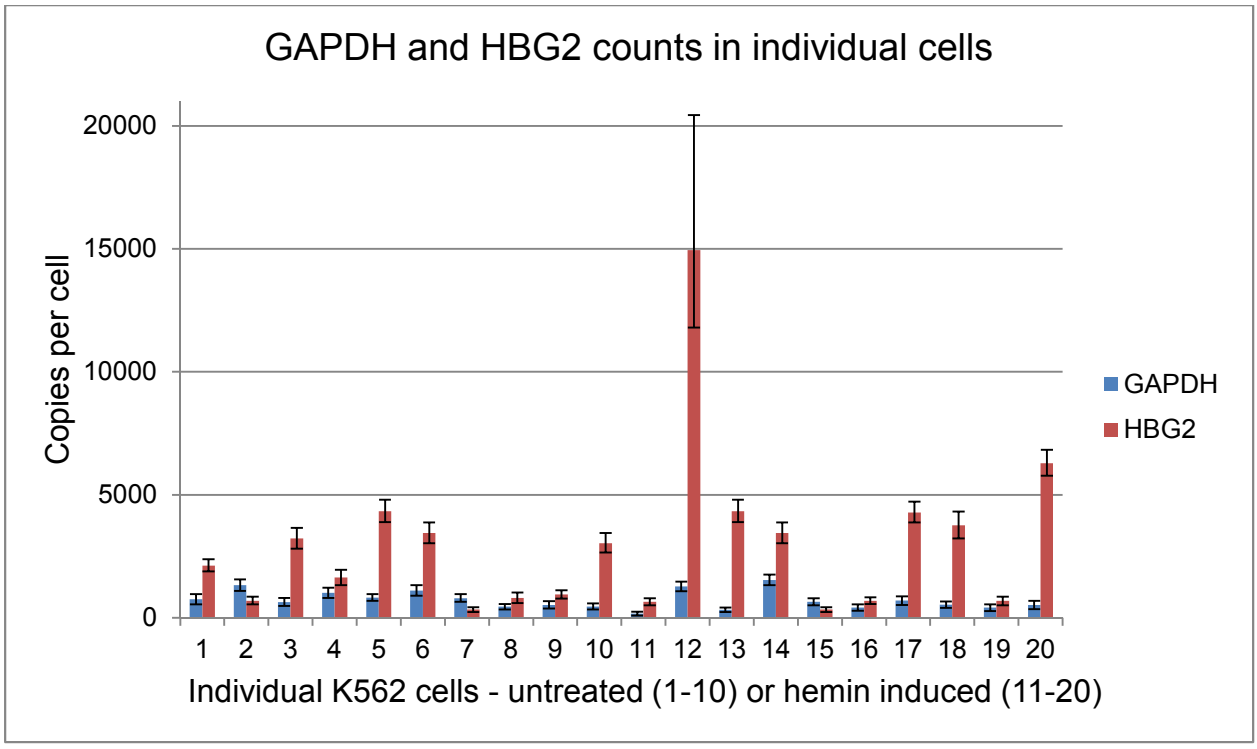


Figure S2. Direct gene expression measurements in single cells. K562 cells were grown either without, or with hemin treatment to induce HBG2 expression. Ten individual cells from each condition were picked into PCR tubes and lysed for measurements of GAPDH (blue bars) and HBG2 (red bars) expression levels. The absolute transcript copies detected in each cell is shown in the chart. Error bars indicate 95% measurement confidence intervals.

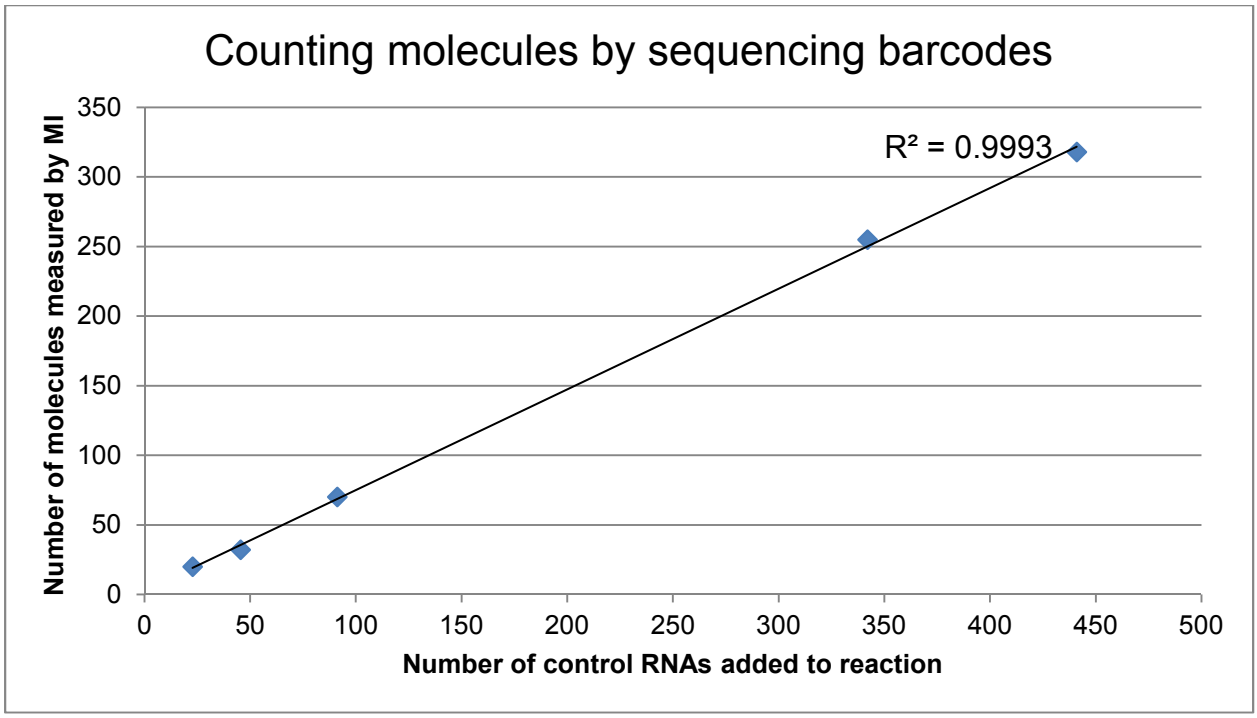


Figure S3. Average measured number of copies of control RNAs added to each of 96 molecular indexing reactions determined by counting barcodes in sequenced reads. Details are shown in Figure S5.

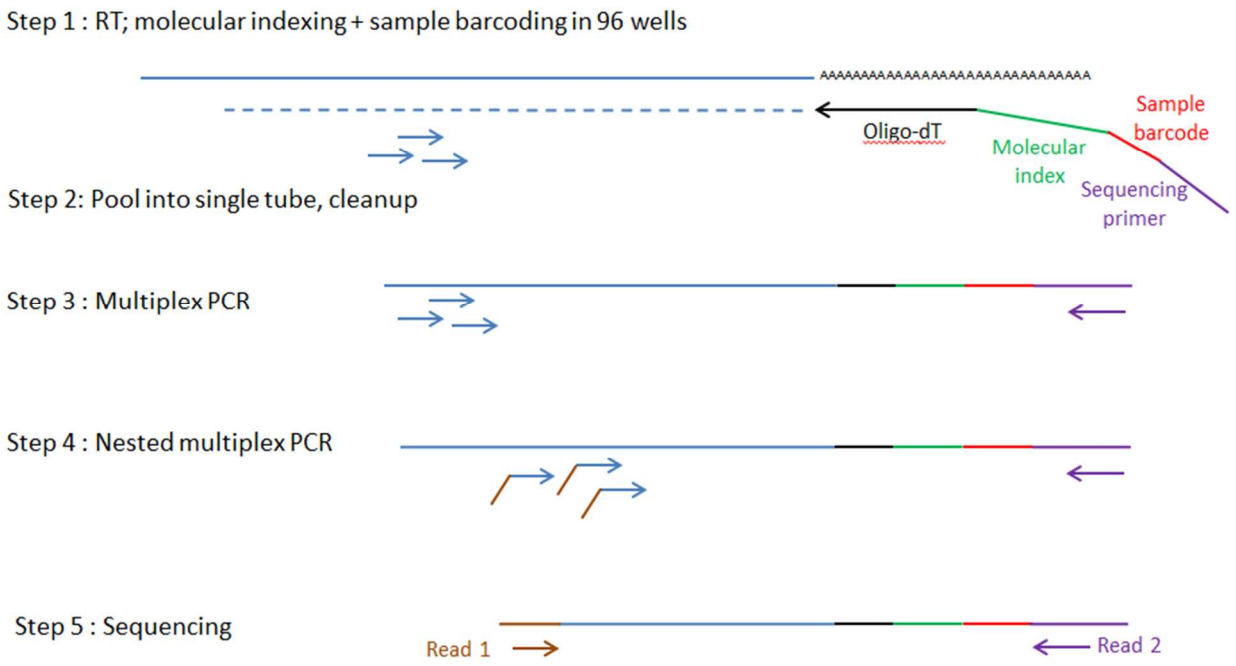
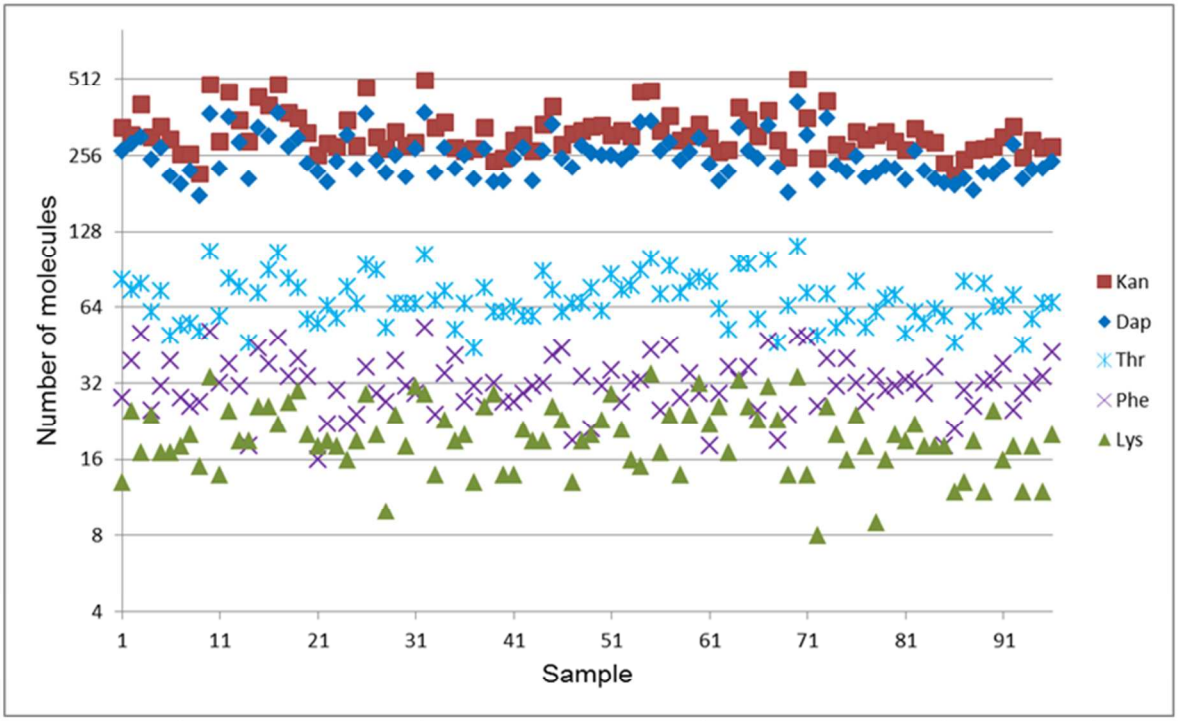
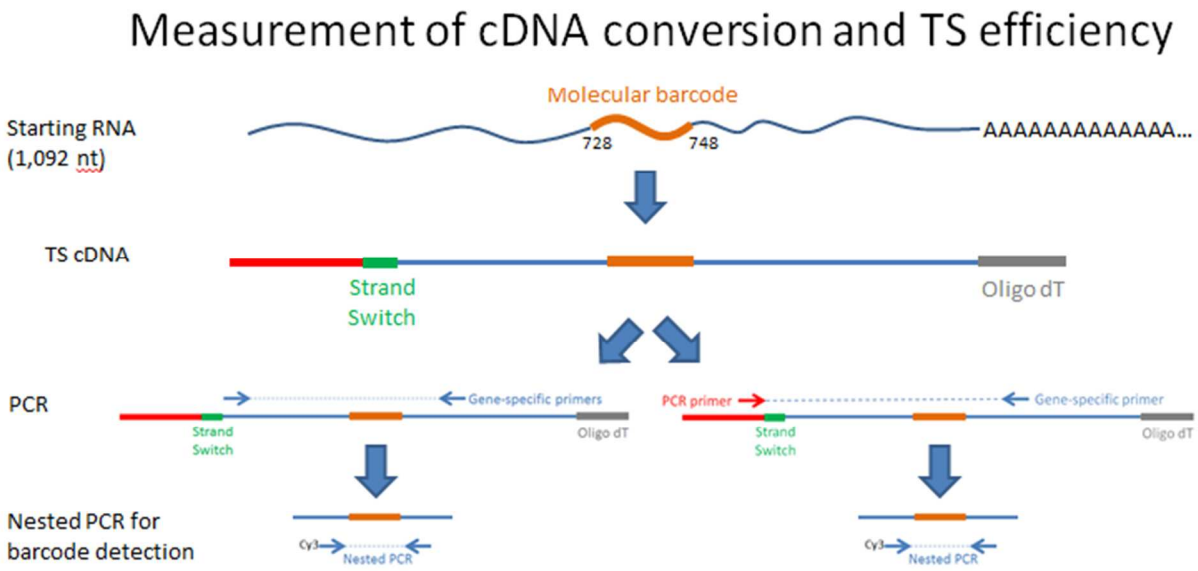


Figure S4. Workflow for combining sample barcoding with molecular indexing. 8 nt barcodes are added to identify pooled cDNAs from samples in a 96 well plate. After reverse transcription, samples are pooled into a single tube, and excess primers are inactivated by UDG treatment and purified using Ampure. Multiplex PCR is applied to amplify cDNAs for 96 genes and 5 spike-in controls, followed by nested PCR to add sequencing priming sites. PCR products are sequenced on the Illumina MiSeq instrument.



Control	Input # RNA	Average # detected	Sensitivity	StDev	CV
Kan	441	318	72.1%	64.5	20.3%
Dap	342	255	74.6%	51.4	20.2%
Thr	91	70	76.8%	15.4	22.0%
Phe	46	32	70.2%	8.0	25.0%
Lys	23	20	87.7%	6.0	30.0%

Figure S5. Absolute counts of molecules for 5 spike-in RNA controls across 96 reactions. Plot above shows the number of molecules detected in sequencing, and the table below shows the number of molecules added per well for each control. Detection sensitivity is expressed as the average number of cDNA molecules detected divided by the number of RNA molecules added.



Input no. of RNAs	No. of cDNAs detected	RT efficiency	No. of TS cDNAs detected	RT + TS efficiency
0	0	NA	0	NA
8	0	0.00%	0	0.00%
80	13	16.36%	3	3.76%
800	65	8.15%	9	1.13%
8000	745	9.31%	126	1.57%

Figure S6. Measurement of Template Switching (TS) efficiency. Barcoded synthetic RNAs were used as input for reverse transcription and TS. cDNAs synthesized were split and either amplified by PCR using a pair of gene-specific primers (left panel) or using the primer incorporated by TS (right panel). Barcodes were counted by nested PCR with dye labelled primers followed by hybridization detection. RT efficiency is calculated as the number of cDNA molecules detected divided by the number of input RNA templates used. The combined efficiency of RT and TS is calculated by dividing the number of resulting cDNAs containing the strand switch oligo with the number of input RNAs.

GGGAGAAATACAAGCTTGGGCGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATCATCATGAA
CAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAACGTCTTGCTC
GAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAATGTCGGGCAATC
AGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTGAAACATGGCAAAGGTAGCGT
TGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCAT
TTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCAGGTATTAGAAG
AATATCCTGAGTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGT
AATTGTCCTTTTAACAGCGATCGCGTATTTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGC
GAGTGATTTTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGCTTTTGCCATTC
TCACCGGATTCAGTCGGCGCGCC**AGCACGACAGACGCCTGAT**Axxxxxxxxxxxxxxxxxxxxx**GCCTGCGA**GCGGCCG
CGATGTTGGACGAGTCGGAATCGCAGAC**CGATACCAGGATCTTGCCAT**CCTATGGAACTGCCTCGGTGAGTTTTC
TCCTTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCATTTGAT
GCTCGATGAGTTTTTCTAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGAGCATTACGCTGACTTGACGGGAC
GGCGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGATCACGCATCTTCCCGACAACGCAGACCG
TTCCAAAAAAAAAAAAAAAAAAA

Figure S7. Sequence of the barcoded spike-in RNA.

Notes:

xxxxxxxxx represents 21 nt barcode

Reverse transcription primer:

T20KanR1	TTTTTTTTTTTTTTTTTTTTGGAACGGTCTGCGTTGTC

Gene-specific primers:

RseqQCKF1	ATTCTCACCGGATTCAGTCG
RseqQCKR1	CGAGGCAGTTCCATAGGATG

Nested Primers:

Cy3PCR004	/5Cy3/ATTATG**AGCACGACAGACGCCTGAT**
RseqQCKR2	**ATGGCAAGATCCTGGTATCG**

Table S1. PCR primers for targeted cDNA sequencing.

Name	F1 Sequence	F2 Sequence
NM_144646.3	TTGACTTTGCCTTGGAGAGC	ATATTTGGACATAACAGACTTGGAA
NR_015342.1	TTTTTCTTACAGTGTCTTGGCATA	TGCTGACTTTTAAAATAAGTGATTCG
NM_000193.2	CGTGACCCTAAGCGAGGAG	GCGGCAGAGTAGCCCTAAC
NM_001777.3	TTTGCAGTGATTTGAAGACCA	TGGGCTATTTCTATTGCTGCT
NM_000600.3	GGCATTCCTTCTTCTGGTCA	AATGGAAAGTGGCTATGCAG
NM_021127.2	CTGGGCTATATACAGTCCTCAAA	GGTTGTAGTCACTTTAGATGGAAAA
NM_004318.3	GGGGTGATTATGACCAGTTGA	TTTGTTTGACTTTGAGCACCA
NM_002467.4	TGCATGATCAAATGCAACCT	AATGTTTCTCTGTAAATATTGCCATT
NM_001773.2	TCTTCCGAAAAATCCTCTTCC	CACCCCCATATGGTCATAGC
NM_001770.5	CTGGGGTCCCAGTCCTATG	AGCACCAGGTGATCCTCAG
NM_001718.4	TGTACTGGGAAGGCAATTTCA	TGTTTTGCTGTAACATTGAAGGA
NR_023920.1	GAGCCGCTGGGGTTACTC	TAATGCCACAGTGGGGATG
NM_000267.3	CAGTTAGTTGCTGCACATGGA	GGGCCTAAACTTTGGCAGTT
NM_000633.2	TTGCATTTCTTTTGGGGAAG	TTTTACCTTCCATGGCTCTTTT
NM_000314.4	GTCATGCATGCAGATGGAAG	GCCTTACTCTGATTCAGCCTCTT
NM_021151.3	GCTGCAGTGAGCTGTGATGT	CGTAACAAAATTCATTGTGGTGT
NM_002415.1	GTTCCTCTCCGAGCTCACC	AGAACCGCTCCTACAGCAAG
NM_004985.3	TCCGAAAGTTTCCAATTCCA	GTGCTTTCTTTTGTGGGACA
NM_005375.2	TTGTTTGGGAGACTCTGCATT	GGGAGTTCTGCATTTGATCC
NM_000555.3	GACCCCACTTGGACTGGTAG	TGGGTCAGAGGACTTCAAGG
NM_001668.3	GTGATCTTGATTGCGGCTTT	AGGGTTCTGATCACATTGCAC
NM_025237.2	GGGGGAAAAACTACAAGTGC	CTGCAGGACTGGTCGTTTTT
NM_021117.3	TGATTCCTTTTCCTGCCTGT	AGGGCAGGGTAGAGAGGGTA
NM_016316.2	AAAAACCTCCAGGCCAGACT	TTCTTCCATGCGGAGAAATC
NM_021975.3	AATCAAAATAACGCCCCAGA	CATGGCTGAAGGAAACCAGT
NM_004333.4	TTGCTAAAAATTGGCAGAGC	TTGCCAGCTATCACATGTCC
NM_001621.4	TTGTTAAGTGCCAAACAAAGGA	TCTTTTCCTGTACCAGGTTTTTC
NM_005239.5	AAGCTGGGAAGAGCAAAGC	TGACTGGGAACATCTTGCTG
NM_000485.2	AGGACAGAGGGTGGTCGTC	TGGCACCTGTACCCTTCTTC
NM_004048.2	TGAGTGCTGTCTCCATGTTTG	TTCAATCTCTTGCACTCAAAGC
NM_001657.2	CCTCACAGCTGTTGCTGTTATT	TGGAGTCACTGCCAAGTCAT
NM_012238.4	AAAACACCCAGCTAGGACCA	TTTGCATGATGTTTGTGTGC
NM_002055.4	AACTGAGGCACGAGCAAAGT	GCACCCACTCTGCTTTGACT
NM_002392.4	GCTTTATGGGTGGATGCTGA	ACCATGTAGCCAGCTTTCAA
NM_001625.3	ATAATATCGCCAGCCTCAGC	GCAACTGGGCATGAGTACCT
NM_002110.3	TCCAGAGTGTGCTGGATGAC	CCACACCCCCTTCCTACTC
NM_002943.3	TGCAAGCCATTTATGGGAAT	AGTCTGCTTATTTCCAGCTGTTT
NM_000059.3	TGGAATGAGGTCTCTTAGTACAGTT	TCCTGTTCAAAAGTCAGGATGA
NM_018136.4	TCCCAGAAACACCTGTAAGGA	AAATCACAAATCCCCTGCAA
NM_003467.2	TGTCTAGGCAGGACCTGTGG	CTGAACATTCCAGAGCGTGT
NM_004958.3	AGTGATGCTGCGACTCACAC	CAGTGGGACCACCCTCACT

NM_006139.3	GGCTCAGAAAGTCTCTCTTTCC	TCTGTAGATGACCTGGCTTGC
NM_002693.2	CTCCCAAACTCAGGCTTTCA	TCAGAACCAAGATGCCAACA
NM_001080432.2	AAAGCGCTGGGATTACAGG	CATGACCCAGCCTATGGTTT
NM_005954.2	CGTCCAGTTGCTTGGAGAAG	ACCTCCTGCAAGAAGAGCTG
NM_024865.2	AATAACCTTGGCTGCCGTCT	TTGGGAGGCTTTGCTTATTTT
NM_001905.2	GGGAATTCTCAGTGCCAACT	CTGGGAAACACTCCTTGCAT
NM_002046.4	GCATCCTGGGCTACACTGAG	CAACGAATTTGGCTACAGCA
NM_002253.2	TGCTGGGAACAATGACTATAAGA	CAAAGGTCATAATGCTTTCAGC
NM_002356.5	GCCTAAAACACTTTGGGTGGT	TTTGACGTATCTTTTCATCCAA
NM_000189.4	GGGTGCCCACAAAATAGAGA	TGTTGTTGGTTTCCAAAAAGG
NM_000546.5	GAGACTGGGTCTCGCTTTGT	GCCAACTTTTGCATGTTTTG
NM_152860.1	TGGGGAAGGCTTTCTCTAGG	CCCAAGCTGATCTGGTGGT
NM_016231.4	TTCAACTTGAGTGATCTGAGCTG	TGCTGTGAAAGAAACAAACATTG
NM_000518.4	TATGGGCAACCCTAAGGTGA	GCACGTGGATCCTGAGAACT
NM_000905.3	CGCTGCGACACTACATCAAC	CCAGCCCAGAGACACTGATT
NM_005038.2	TGGAGTCTTGCTCTGTCACC	CACGCCCAGCTAATTTTTGT
NM_000041.2	ACGAGGTGAAGGAGCAGGT	CCTGGTGGAAGACATGCAG
NM_005957.4	CGATGCCTTTGGGTAGAGAG	TCACACCTGTAATCCCAGCA
NR_002785.2	ACTGATCGTCCAAGGACTGG	CAGAGCTCCGCCTCATTAGT
NM_000321.2	AAAAAGAAATCTGGTCTTGTTAGAAAA	TCCATTTCATCATTGTTTCTGC
NM_152756.3	TTGAAAAGTGGTAAGGAATTGTGA	TGGTGTTTGTAGGTCACTGAACA
NM_000610.3	CACCAAGAATTGATTTTGTAGCC	AACATGGTCCATTCACCTTTATG
NR_033314.1	AAAAATGGGGGAAAATGGTG	AGAGCGAGACTCCGTCTCAA
NM_017460.5	CATGGTTGAAACCCCATCTC	AGTGAGCTGAGATTGCACCA
NR_002196.1	TTCAAAGCCTCCACGACTCT	AGACGGCCTTGAGTCTCAGT
NM_000591.3	GCTGGAACAGGTGCCTAAAG	GGGAATCCCTTCCTGGTC
NM_000106.5	CCCTAAGGGAACGACACTCA	CTTCCTGCCTTTCTCAGCAG
NM_138712.3	ACCTGCTACAAGCCCTGGA	TGCAGGTGATCAAGAAGACG
NM_004304.4	GGATCCCTAAGACCGTGGAG	GGTTTTGAGCATGGGTTCAT
NM_000754.3	CCACCTCAGAGGCTCCAA	CCAGCCCACTCCTATGGAT
NM_000492.3	TGCTGTATTTTAAAAGAATGATTATGA	AAACTGGGACAGGGGAGAAC
NM_000444.4	GTAGCTGGGACGCTGGTTTA	TTTGGGTAGGTGACCTGCTT
NM_002463.1	ATTCCCTTCCCCCTACAAGA	TCACTGAACGAATGAGTGCTG
NM_000552.3	CCTGAGTGCAACGACATCAC	ACGATGTGCAGGACCAGTG
NM_005430.3	GGGGGAACCAGCAGAAAT	AATTTGCACTGAAACGTGGA
NM_003150.3	GACCTAGGGCGAGGGTTC	CTGTTGTGGCCCATTAAAGAA
NM_000388.3	AATTCCTGAAGCCAGATCCA	TTCCCTCCAGCAGTGGTATT
NM_007294.3	AAAATGTTTATTGTTGTAGCTCTGG	CACCAGGAAGGAAGCTGTTG
NM_005933.3	TTTCAAGAGCTCAACAGATGACA	TTTCCTTGTGTTCTTCCAAGC
NM_002343.3	GACTGCCCGGACAAGTTTT	TCGCAGGCATTACTAATCTGAA
NM_000376.2	GAGAAGGTGCCCCAAAATG	CTCTGGCTGGCTAACTGGAA
NM_002462.3	AGCCACTGGACTGACGACTT	AGAGCCCCACCCTCAGAT
NM_021005.3	GGAGGACTAGTGAGGGAGGTG	TGTGCAGAGTTCTCCATCTGA

NM_012343.3	GGCAAGTGATGTGGCAATTA	TGCCTGTTACAAATATCAAGGAA
NM_001741.2	GTTGGAGCACCTGGAAAGAA	TTTCCCTTCTTGCATCCTTC
NM_014417.4	ATGCCTGCCTCACCTTCAT	TGTGACCACTGGCATTCATT
NM_014009.3	ACAGGGGCACTGTCAACAC	CTCACACACACGGCCTGTTA
NM_006908.4	AAAAATCATGTGTTGCAGCTTT	CACTTGACCAATACTGACCCTCT
NM_005228.3	TGCTTTCACAACATTTGCAG	GTGTGTGCCCTGTAACCTGA
NM_013994.2	AATGTTTCCTTGTGCCTGCT	CCACTTCCCACTTGCAGTCT
NM_000639.1	ATATCCTGAGCCATCGGTGA	TGTGTGTGTGTGTGTGTGTGT
NM_002701.4	TTTTGGTACCCCAGGCTATG	TCTCCCATGCATTCAAACTG
NM_000268.3	ACCCCGTGGCATTACATAAC	TCTAAGTGTTCCTCACTGACAGG
NM_003140.1	CTTCCAGGAGGCACAGAAAT	TACTCTGCAGCGAAGTGCAA
NM_000551.3	CTAACCTGGGCGACAGAGTG	CCAAGATCACACCATTGCAC

Table S2. Sample barcodes for sequencing.

TAGCTTGT
CGATGTTT
GCCAATGT
ACAGTGGT
ATCACGTT
GATCAGCG
CAGATCTG
TTAGGCAT
GGCTACAG
CTTGTACT
ACTTGATG
TGACCACT
TGGTTGTT
TCTCGGTT
TAAGCGTT
TCCGTCTT
TGTACCTT
TTCTGTGT
TCTGCTGT
TTGGAGGT
TCGAGCGT
TGATACGT
TGCATAGT
TTGACTCT
TGCGATCT
TTCCTGCT
TAGTGACT
TACAGGAT
TCCTCAAT
TGTGGTTG
TAGTCTTG
TTCCATTG
TCGAAGTG
TAACGCTG
TTGGTATG
TGAACTGG
TACTTCGG
TCTCACGG
TCAGGAGG
TAAGTTCG
TCCAGTCG
TGTATGCG

TCATTGAG
TGGCTCAG
TATGCCAG
TCAGATTC
TACTAGTC
TTCAGCTC
TGTCTATC
TATGTGGC
TTACTCGC
TCGTTAGC
TACCGAGC
TGTTCTCC
TTCGCACC
TTGCGTAC
TCTACGAC
TGACAGAC
TAGAACAC
TCATCCTA
TGCTGATA
TAGACGGA
TGTGAAGA
TCTCTTCA
TTGTTCCA
TGAAGCCA
TACCACCA
TGCGTGAA
GGTGAGTT
GATCTCTT
GTGTCCTT
GACGGATT
GCAACATT
GGTCGTGT
GAATCTGT
GTACATCT
GAGGTGCT
GCATGGCT
GTTAGCCT
GTCGCTAT
GGAATGAT
GAGCCAAT
GCTCCTTG
GTAAGGTG
GAGGATGG

GTTGTCGG
GGATTAGG
GATAGAGG
GTGTGTCG
GCAATCCG
GACCTTAG
GCCTGTTC
GCACTGTC
GCTAACTC
GATTCATC
GTCTTGGC

Table S4. cDNA synthesis and PCR primers.

Gene	SS primer	F1 primer	F2 primer
GAPDH	CACATGGCCUCCAAGGAGUAA	AAGAGAGAGACCCTCACTGCTG	GTTGCCATGTAGACCCCTTG
B2M	CTTCCAATTTACATACUCTGCTUAGA	TTGACAGGATTATTGGAAATTTGTT	GGCATGGTTGTGGTTAATCTG
RPL19	AAGAAGCUCCTGGCUGACC	GAGGCCCGCAGGTCTAAG	CCAAGGAGGAAGAGACCAAG
RPLP0_1	CGAGTCCTGGCCTTGTCTGT	AAGCCAAGGAAGAGTCGGAGG	GGAGTCGGACGAGGATATGGG
RPLP0_2	ACGGGTACAAACGAGTCCTG	AAGGTTGAAGCCAAGGAAGAGT	AAGCAACCAACTTAGCCAGTTTT
RPL31	ACCTCGGGCACTCAAAGAGATT	GGCTCAACAAAGCTGTCTGGG	GCGGCTGTCCAGAAAACGTAA
UBC	ACTCTGCACTTGGTCCTGCG	TTGGTCCTGCGCTTGAGGGG	GGGGGTGTCTAAGTTTCCCCTTT
FTL	TTGGATCTTCATGCCCTGGGT	CCTTCTGAGCCCAGCGACTT	GCTTCTGCCTAAGCCTCTCCC
EEF2	CAGCAGACACGCCCTCTTAGT	TTTGGGTACATGGCCCGCAG	TGTCACCATGAGCACCTCCAG
ALDOA	AGCGAGTCCCTCTTCGTCTCT	CTTCCCTCGTGACAGTGGTG	CATCACCCTTTCCGGCACAC
PPIA	CCCTTCTCTGCCCACCTTAACA	TGTCTGGTTCCTTCTGCGTGA	CAGACCTGAGTGCAAGGTGGA
GUSB	AGACTGATACCACCUGCGUGT	TTCCACAGCAGCAGAACAAG	CCTGGGTTTTGTGGTCATCT
POLR2A	GGTGTTCTCACTCCTACCCCAC	ATCCTTCCTTGCCTGTGGCTT	GGTAGTTCCTGCTGTGAGTGGT
NONO	GGACACCCAGGAATGACCCTTT	CTGACAGTAGCTCTTAGACTCGCC	GAGCAGTAATGCAGCATCAACCTAT
CSNK2B	GCTCTTCATGGTGCATCCCG	TGCCCAGGCTCTACGGTTTC	CCCACCTGTCCTGCAGTCTTT
ARHGDIA	GTGCACACATTGCTGAGAGCC	GGTCTTCGTGTTGTGTCTGCC	GCCTGTCCCTGCTTCTTTTCTG
TFRC	CAGACTAGTGACAAGCTCCTGGT	TGGGCCTTTTGGAGGTAAAGGA	GTGCAACAGTGTGGAGATTCCTTG
CDA	CCCTCCTAGCAACCTGCCTT	CTGCCCCACCTTTCCTTTCCT	AGTCTGGACTGCTTCCCCATC
CLTB	GCTTTGGTCGGGGTGGAGAC	CAGCCTATTCCACTCCTCCCC	CTCACGCCTCCTCTCAGTCTAC
CSK	TTTCCTGTCCTGCCCGTGAG	GCGCTTGACCATGTTGCACT	ACCCGCCTTGTGAGATGGAA
DAXX	GCCTCCTCGGCCTGGTACTT	TCTGATTAGCTGCCTCCCCTTC	TTTGGAGGAAGGTGGGAAGCA
TBP	CCATGAACCACAGUTTTTATAUTTC	TTCTGTGCCAGACACATTCC	TGTTAATGAAAATGAATGGCTGT
ABCF1	GGAGATTGAUGGGAGTUGGA	CTGTTCTCAGTCTTGATTTGCTT	TTTGCAGTCTTGCTGACAGTG
G6PD	GTGAGAGAATCUGCCTGUGG	TCCTTGTCACCAGCAACATC	ACCCGTAGGCAGCCTCTCT
LDHA	TGGGCAACCCTGCAACGATT	GTTGTATTGGTACCACTTCCATTGT	ATCTTGTCCTCTGGAAGCTGG
EEF1A	AGATGGCCAATAGATGCCCTGAT	TTCCCACTGGAAGCAGGAATGAG	CAACTGTGGCTTGAGCACCAC
HPRT1	TGTCAGTTGCTGCATTCCTAAAC	GTTCCCTTTGGGCGGATTGT	TGAATCTTTGTCAGCAGTTCCCT
ARAF	TTTGTGCCTGATGTGCCTTCC	TTGTGTCTCCCCCGCCATTC	GAATTCTGCTGGGCCTTTGCT
ZNF384	GAGCTGTGGGAGGCTTTGTG	CCACAGTCAGATGCCGGAC	CTCCACCCCTGTCCTCTAGC
GTPBP1	AGGGTGGGGTTGGATCTGTG	CTACAGCCCTCCCAGCTAAACA	ACACCAGCCCTTTTGTAGTGGA
CLN5	AAGGCTAAGTGATCTTTCTCTGGCT	TAATGAAGTCAGGTCTTGAACCCCG	TCTGCTGACTGAATTGGATGCACTA
HLA-DQB1	TTGGGGTGTGAAGTAAGGGTGG	TGGCCTTAGAATATCCCAGCCTC	CTCTTCTGGTTTGGTGAGTGCTG
CDK11B	CCCATCGACCCCTCCATGTT	AGACGGGCTTCCACCTTACC	AAGGTCAGAGTGGACCCCGT
HBG2	GAGAAACCCUGGGAAGGCUC	CTTCAAGCTCCTGGGAAATG	CCTGGCAGAAGATGGTGACT
Kanamycin	CCAUUCUCACCGGAUUCAGU	GATGTTGGACGAGTCGGAAT	CTGCCTCGGTGAGTTTTCTC

REFERENCES

(1) Hardenbol, P.; Yu, F.; Belmont, J.; Mackenzie, J.; Bruckner, C.; Brundage, T.; Boudreau, A.; Chow, S.; Eberle, J.; Erbilgin, A.; Falkowski, M.; Fitzgerald, R.; Ghose, S.; Iartchouk, O.; Jain, M.; Karlin-Neumann, G.; Lu, X.; Miao, X.; Moore, B.; Moorhead, M.; Namsaraev, E.; Pasternak, S.; Prakash, E.; Tran, K.; Wang, Z.; Jones, H. B.; Davis, R. W.; Willis, T. D.; Gibbs, R. A. *Genome research* **2005**, *15*, 269.

(2) Fu, G. K.; Hu, J.; Wang, P. H.; Fodor, S. P. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 9026.

(3) Li, H.; Durbin, R. *Bioinformatics* **2009**, *25*, 1754.

(4) Langmead, B.; Salzberg, S. L. *Nature methods* **2012**, *9*, 357.

(5) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. *Nature methods* **2008**, *5*, 621.

(6) Fu, G. K.; Xu, W.; Wilhelmy, J.; Mindrinos, M. N.; Davis, R. W.; Xiao, W.; Fodor, S. P. A. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, DOI:10.1073/pnas.1323732111.