Supplementary Information

Title
Quantitative Transcriptomics using Designed Primer-based Amplification.

Authors: Vipul Bhargava[1], Pang Ko[2], Erik Willems[3], Mark Mercola[2,3], Shankar Subramaniam[1,2,4*]

Affiliation:
[1]Bioinformatics and Systems Biology Graduate Program, University of California at San Diego, La Jolla, California, USA.

[2]Department of Bioengineering, University of California at San Diego, La Jolla, California, USA.

[3]Sanford-Burnham Medical Research Institute, La Jolla, California, USA.

[4]Departments of Cellular and Molecular Medicine and Chemistry and Biochemistry, University of California at San Diego, La Jolla, California, USA.

[*] Correspondence to shankar@ucsd.edu.

This file includes:
(1) Supplementary Methods
(2) Supplementary References
(3) Supplementary Table(s) S1-S6
(4) Supplementary Figures(s) S1-S12

Supplementary Methods

Targeted cDNA amplification with Heptamer Primers

cDNA preparation

Total RNA was extracted from harvested cells using Trizol (Invitrogen). About 1-5 ug of total RNA was later subjected to Oligo(dT) selection using Oligotex mRNA Mini Kit (Qiagen) according to the manufacturer's protocol. If the total RNA is less than 1 ug, we recommend using Dynabeads mRNA Purification Kit (Invitrogen) for extraction of poly-adenylated RNA. Next, first strand cDNA was synthesized with oligo dT (20-mer) primers using QuantiTect Reverse Transcription Kit (Qiagen) according to manufacturer's instructions. This kit allows synthesis of full-length cDNA (as long as 10 Kb). The reaction was later purified using Agencourt AMPure XP system (Beckman Coulter) according to manufacture's protocol and eluted in 20 ul of elution buffer (EB).

Primer hybridization and extension

Heptamer primer hybridization and extension was achieved by using Klenow (exo-) polymerase, a mesophilic polymerase with strand displacement capability. Exo-nuclease deficient version of Klenow polymerase was used to avoid degradation of heptamer primers. Since the 44 heptamer primers were split into three different primer sets, a master mix was prepared comprising of 1 – 5 ng of cDNA, Taq polymerase buffer (10X) supplemented with 2.5 mM $MgCl_2$, 4% DMSO and 0.2 mM dNTP (10 mM stock). DNase free water was added to make the total reaction volume of 24 µl. The master mix was split equally into three PCR reaction tubes. Later 1 µl of heptamer primer mixes were added to their respective tubes. The reaction mix was incubated at 95°C for 5 mins to denature the cDNA template. Mis-hybridization of the heptamer primers was minimized by ramping down the temperature of reaction mix to 37°C at the rate of -0.2°C/sec. At this point, 1 unit of Klenow polymerase (exo-) was added to each reaction tube and incubated for 30 mins at 37°C and then 5 mins at 42°C. Klenow polymerase retained most of its activity in Taq polymerase buffer and its extension rate was not affected at 2.5 mM $MgCl_2$ concentration, as reported earlier[1].

Taq polymerase amplification

Taq polymerase possesses optimal affinity for DNA (K_m ~ 2 nM) allowing efficient amplification of the PCR products while avoiding primer dimerization. Moreover, Taq polymerase allowed the addition of tail dATP at the 3' end of most of the amplicons thus eliminating this step from sequencing-library generation. A PCR master mix was prepared containing: 2 µl of Taq reaction buffer (10X), 1.25 mM of $MgCl_2$, Buffer Q (5X, Qiagen), 2 µl of primer mix (2 µM stock), 0.2 mM of dNTPs (10 mM stock) and 2 units of Taq polymerase. DNase free water was later added to top up the reaction mix to 20 µl. Similar reaction mixes were prepared for the other reaction tubes. Later, the reaction mix was added to the Klenow reaction (30 µl of total volume) and a 14-cycle amplification was performed consisting of denaturation (95°C for 30 s), annealing (46°C for 30 s) and elongation (72°C for 40 s). The amplified libraries obtained from the three

tubes were pooled together and purified using Agencount AMPure XP system. The amplicon library was eluted in 44 µl of EB.

Library Preparation

End Repair

The 5' ends of the PCR products were phosphorylated using T4 Polynucleotide Kinase (PNK) enzyme (NEB) in the presence of T4 DNA Ligase buffer containing ATP. The T4 PNK treatment was set up as follows:
Amplicon library: 44 µl
T4 DNA ligase buffer: 5 µl
T4 PNK: 1 µl (10 units)
The reaction was incubated at 37°C for 30 mins. Later, the reaction was purified using Agencourt AMPure XP system and eluted in 15 µl of EB.

Ligation

Custom adaptor oligos were ordered in 100 µM concentration (Valuegene Inc.) with following modifications:

a) Adaptor_A_F

5'- Biotin-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCT-S-T -3'

(-S- represents Phosphorothioate Modification; 5' end of the oligo is biotinylated)

b) Adaptor_A_R

5'- Phospho-AGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT -3'

(5' end of the oligo is phosphorylated)

c) Adaptor_B_F

5'- CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCT-S-T – 3'

d) Adaptor_B_R

5' – PhosphoAGAGCGGTTCAGCAGGAATGCCGAGACCGATCTCGTATGCCGTCTTCTGCTTG -3'

Adaptor oligos referring to adaptor A (a, b) and adaptor B (c, d) were mixed in equi-molar concentrations and diluted to 2 µM final concentration. Both adaptors were later denatured at 95°C for 5 mins and then brought back to room temperature gradually at -0.2°C/s. This allowed hybridization of the two oligos of the adaptor with 'T' overhang. The adaptor mix was further diluted 1:10 to get a stock concentration of 200 nM. The Ligation reaction was set up as follows:
T4 PNK treated PCR product: 6 µl
Adapter A: 1 µl
Adapter B: 1 µl
T4 DNA Ligase Buffer: 1 µl
T4 DNA Ligase (NEB): 1 µl (400 units)
The reaction was performed at room temperature for 1 hr or at 16 °C overnight.

Selection of adaptor orientation

Ligation reaction resulted in fragments with either two identical (A-A and B-B) or two distinct (A-B and B-A) adapter orientations. However, only distinct adapter orientation fragments can be sequenced in Illumina's platform. We enriched desired ligation products by utilizing the biotin (adaptor A) – streptavidin

chemistry. Streptavidin coated magnetic beads (Dynabeads MyOne Streptavidin C1, Invitrogen) were used to pull down A-A, A-B and B-A ligation products using manufacturer's protocol. The supernatant, containing B–B, was discarded. Later, 0.2 N NaOH was added to the beads. Incubation for 10 mins at room temperature denatured two strands of the ligation product. Only A'–B strand appeared in the supernatant while both strands of the A–A remained associated with the beads. The supernatant with distinct orientation was extracted and column purified using MinElute PCR Cleanup Kit (Qiagen). The pH of the supernatant was adjusted to allow maximal binding to the column. The single stranded DNA was eluted in 36 µl of EB.

Final PCR and size selection

The single stranded DNA obtained from previous step was amplified using adaptor specific primers. Following primers were ordered in 100 uM concentration:

a) Final_FP:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAATA -3'

b) Final_RP:

5'-
CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATA-3'

A 50 µl PCR reaction was set up with 18 µl of single stranded template, 5 µl of primers (2 µM stock), 4% DMSO, 5 µl Pfu Turbo reaction buffer (10X), 0.2 mM dNTP (10 mM stock), 2.5 units of Pfu Turbo Polymerase. The amplification consisted of 14 cycles of denaturation (95°C – 30s), annealing (62°C – 30s) and extension (72°C – 40 s).

The amplified product was run in 2% agarose gel at 80 – 100 volts for 1 hr. Using 50 bp ladder (NEB) a band corresponding to size range of 150 – 500 bp was cut out. The DNA was retrieved from the gel using MinElute Gel Extraction Kit (Qiagen) with 15 µl of elution.

Quantification of the sequencing library

Quantitative real time PCR was used to determine the concentration of the sequencing libraries prepared by our protocol. The standard curve for various dilutions of phiX control library was generated using the adapter specific primers recommended by Illumina. We later used the standard curve to determine the molarity of our sequencing libraries.

The concentration of sequencing library loaded into the flowcell was calibrated by the sequencing facility. We typically obtained good cluster density with 5 pM of library concentration on HiSeq v3 kit.

Oligonucleotides

All of our heptamer primers were flanked by universal adapter sequence (5'-CCGAATA-heptamer-3') and synthesized by Valuegene Inc. These primers were desalted and suspended in RNase/DNase free water to 100 µM concentration.

Later, the primers were pooled together into three different tubes as described in **Supplementary Table S4** at equi-molar concentration to prepare a stock solution containing 2 µM of each heptamer primer.

Mouse embryonic stem cell culture and differentiation

Mouse R1 or T-GFP embryonic stem cells were cultured on mouse embryonic fibroblast (MEF) on gelatin-coated dishes in high glucose DMEM (Hyclone, Logan, UT) supplemented with 10% fetal calf serum (FCS) (Hyclone, Logan, UT), 0.1 mM b-mercaptoethanol (GIBCO), 1% non-essential amino acids (GIBCO), 2 mM L-glutamine (Sigma, St. Louis, MO), sodium pyruvate (Sigma), antibiotics (Sigma), and 1,000 U/ml of LIF (Sigma) and passaged with 0.25% Trypsin (GIBCO).

For embryoid body (EB) differentiation, MEF were stripped from the cultures by 15 minutes incubations on gelatin-coated dishes. mESCs were collected and washed in PBS to remove traces of serum. mESCs were differentiated in serum free media containing N2 and B27 supplements as described elsewhere[2, 3]. mESCs were aggregated at 50,000 cells/ml in non-coated polystyrene plates. After 2 days, EBs were dissociated by trypsin treatment and re-aggregated in fresh media in presence of different growth factors and small molecules. Activin A and BMP4 were obtained from R&D Systems while SB-431542 was obtained from Sigma. IWR-1 was synthesized in house as described previously[4]. EBs were harvested at day 4 for RNA extraction and processing.

Library Generation for mRNA dilution series using DP-seq

Serial dilutions (10 ng, 1 ng, 100 pg, 50 pg, 10 pg, and 1 pg) were prepared for the mRNA derived from Activin A (3 ng/mL) sample. First strand cDNA synthesis was performed for all mRNA dilutions in duplicates to get the technical replicates. Later, the purified cDNA prepared from each dilution, was split into three tubes to perform amplification using our heptamer primers. The numbers of PCR cycles were increased for lower dilutions to get appropriate amounts of DNA for the library construction. The numbers of PCR cycles used are as follows:

10 ng and 1 ng – 13 cycles
100 pg and 50 pg – 16 cycles
10 pg – 19 cycles
1 pg – 23 cycles

The amplicon libraries thus constructed, were phosphorylated at the 5' end as mentioned above. Later, the libraries were ligated with Illumina's Y-adaptors and amplified using adaptor specific primers consisting of a different Illumina's Truseq barcode sequence for each library. The amplified libraries were run through the 2% agarose gel and size selected (150 – 500 bp) for sequencing. Similar methodology was used for the generation of sequencing libraries with ribosomal inhibition primers.

Library Generation using Std. RNA-seq protocol

Std. RNA-seq[5] libraries were constructed from about 10 ng of mRNA derived from Activin A (3 ng/mL) and Activin A (100 ng/mL) samples using Illumina's TruSeq RNA Sample Prep Kit v2.

Library Generation using Smart-seq

Smart-Seq cDNA generation and amplification was performed on 50 picograms of mRNA derived from Activin A (3 ng/mL and 100 ng/mL) treated samples using SMARTer Ultra Low RNA Kit for Illumina sequencing (Clontech). We performed 13 cycles of amplification to achieve about 1-10 ng of the amplified cDNA libraries. These libraries were later sheared using Covaris system to obtain 200-500 bp fragments. Later, standard Illumina library preparation protocol was followed to prepare the sequencing libraries using Illumina Paired-End DNA Sample Prep kit.

Reverse Transcription and Quantitative RT-PCR (qPCR)

Total RNA was extracted from cells using Trizol (Invitrogen) according to the manufacturer's instructions. About 1 µg of total RNA was treated for DNA removal and converted into first strand cDNA using Quantitect Reverse Transcription kit (Qiagen). SYBR Green qPCR was run on a LightCycler 480 (Roche) using the LightCycler 480 SYBR Green Master Kit (Roche). All primers were designed with a T_m of 60°C. Data was analyzed using the $\Delta\Delta C_t$ method, using β-actin as normalization control, which was determined as a valid reference in mouse ESC differentiation. The primer sequences are listed in **Supplementary Table S6**.

Flow Cytometry

Day 4 embryoid bodies from T-GFP mESC were dissociated with trypsin to single cell suspensions and analyzed on a FACSCanto (BD Biosciences). Prior to analysis, cells were stained with propidium iodide to label dead cells. Data analysis was performed using FlowJo (Treestar Inc.) where measured events were gated for PI negative populations (exclusion of dead cells) and forward/side scatter (exclusion of debris and aggregates) before generating dot plots.

Data Analysis

Mapping reads

Our libraries were sequenced on Illumina's GIIx Analyzer and HISEQ2000 platforms. We performed single end 36 sequencing cycles on version 5.0 of flowcell (FC-104-5001 | TruSeq SBS Kit v5 – GA (36-cycle)). The raw reads were truncated as 32-mer with the first and last 2 base pairs of the reads removed. The 32-mer reads were aligned to the RefSeq mRNA database allowing up to 2 mismatches using our in-house software which uses suffix array implementation. Reads that did not align to the mouse RefSeq mRNA database were later aligned to mouse genome using Bowtie[6].

Libraries constructed from serial dilutions of mRNA were sequenced in Illumina's HiSeq2000 systems (TruSeq SR Cluster Kit v3-cBot-HS and TruSeq SBS Kit v3-HS). The libraries were sequenced as 100 bp single-end reads. The first 14

sequences came from our heptamer primers including the universal tail sequence (5'-CCGAATA-3') as such the first 14 bps were truncated and next 32 bp sequence was aligned to the mouse transcriptome allowing ≤ 2 mismatches.

Mapping of Smart-seq reads

The number of reads obtained from Smart-seq was double the number of reads for DP-seq. Previous studies[7, 8] have demonstrated that the transcriptome coverage and the technical noise in expression measurements vary with the sequencing depth and global normalization of the reads across different samples is heavily affected by few highly expressing transcripts. In order to perform an objective comparison between Smart-seq and DP-seq, we downsized the Smart-seq libraries by generating multiple random sets, consisting of a similar number of reads obtained from DP-seq. The reads in these datasets were mapped to the mouse transcriptome allowing ≤ 2 mismatches. The analysis of these random sets showed similar transcriptome coverage and technical noise. The table below summarizes the results obtained for the three different random sets generated from Activin A 100 ng/mL dosage. The First two columns depict the entire dataset.

	AA100; TR1	AA100; TR2	AA100; TR1-1	AA100; TR2-1	AA100; TR1-2	AA100; TR2-2	AA100; TR1-3	AA100; TR2-3
Transcriptome Coverage	14052	14183	13400	13568	13389	13509	13392	13499
R^2 (Technical Noise)	0.8412		0.8478		0.8513		0.8509	

In this study, we present the comparison of DP-seq with one of the random sets (random set 1) generated from the Smart-seq library.

Differential gene expression analysis

We employed a local pooled variance test similar to LPE[9] to identify differentially regulated transcripts. For each transcript, unique reads coming from predicted and non-predicted primer-binding sites were combined in all samples. Prior to identifying the differentially expressed transcripts, the fold changes between control and treatment conditions (including technical replicates) were lowess normalized to eliminate average read dependent variations in the fold changes. The noise in the technical replicates reflected variability arising out of sample preparation and the sequencing platform. As such we used the expression measurements obtained from the technical replicates to determine the baseline (null) distribution where no differential expression of the transcripts was expected. The null distribution was determined by plotting M and A quantities for technical replicates, which are defined as:

$M_{i,j}=Log_2(\frac{Reads,i}{Reads,j})$

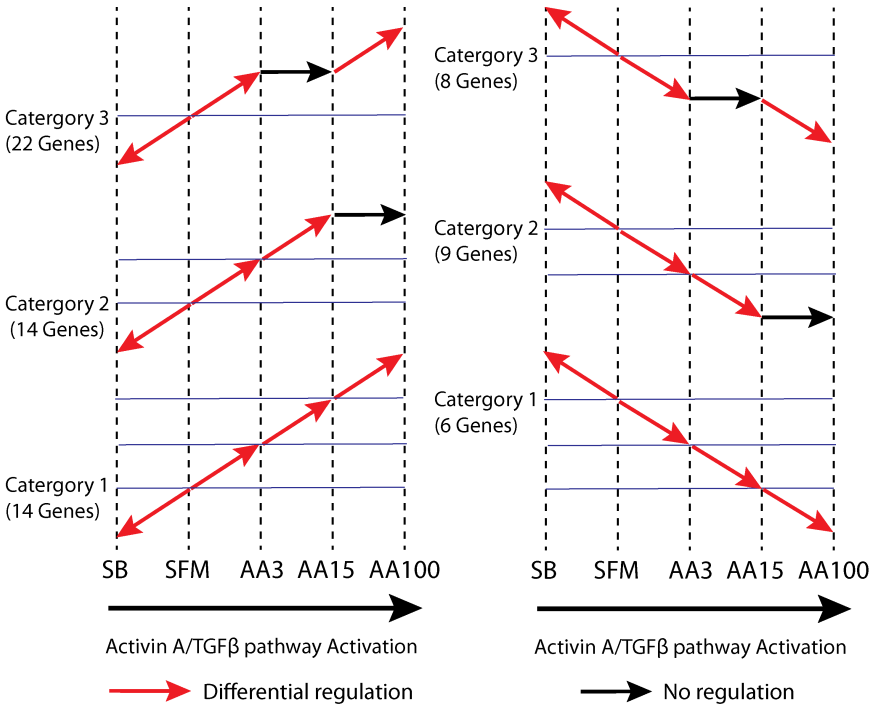
$A_{i,j}=0.5\times Log2(Reads,_i \times Reads,_j)$

where ‘i’ and ‘j’ represents any two samples. M corresponds to log ratio in unique reads for a transcript between samples ‘i’ and ‘j’ while A corresponds to average reads for the transcript in the two samples.

To quantify the technical noise, we pooled the expression of ~200 transcripts in the null distribution with similar reads and estimated the standard deviation in their fold change. We assumed that all transcripts with similar expression measurements possess similar noise. Also, the distribution of the fold changes was assumed to follow a Gaussian distribution. Next, a threshold for differentially expressed transcripts was determined as 1.96 times the standard deviation, corresponding to a less than 5% chance of the transcript being called differentially expressed by random chance. The experimental MA plot, which was defined as treatment/control, was overlaid on the technical replicate MA plot and any transcript representing a fold change above/below the threshold was designated as differentially expressed. Higher thresholds (blue curve) were used for the low expressing transcripts as demonstrated in **Supplementary Fig. S12.**

Identification of Activin A/TGFβ target genes

Putative Activin A/TGFβ target genes were determined as genes exhibiting opposite mode of regulation in AA3 and SB samples as compared to serum free media control. The target genes were further classified into three categories of expression as shown below:



A p-value cutoff of 0.05 was used to determine differentially expressed transcripts.

Supplementary References

1. Zhao, G. & Guan, Y. Polymerization behavior of Klenow fragment and Taq DNA polymerase in short primer extension reactions. *Acta Biochim Biophys Sin (Shanghai)* **42**, 722-728 (2010).
2. Gadue, P., Huber, T.L., Paddison, P.J. & Keller, G.M. Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc Natl Acad Sci U S A* **103**, 16806-16811 (2006).
3. Willems, E. & Leyns, L. Patterning of mouse embryonic stem cell-derived pan-mesoderm by Activin A/Nodal and Bmp4 signaling requires Fibroblast Growth Factor activity. *Differentiation* **76**, 745-759 (2008).
4. Chen, B. et al. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat Chem Biol* **5**, 100-107 (2009).
5. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628 (2008).
6. Langmead, B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics* **Chapter 11**, Unit 11 17 (2010).
7. Bullard, J.H., Purdom, E., Hansen, K.D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010).
8. McIntyre, L.M. et al. RNA-seq: technical variability and sampling. *BMC Genomics* **12**, 293 (2011).
9. Jain, N. et al. Local-pooled-error test for identifying differentially expressed genes with a small number of replicated microarrays. *Bioinformatics* **19**, 1945-1951 (2003).

Supplementary Table S1A: Comparison of sequencing libraries made from various dilutions of mRNA derived from Activin A (3ng/mL; AA3) sample using DP-seq.

Amount of mRNA	Total Reads	% Of reads aligned to Refseq Transcripts	% Of unmapped reads aligned to genomic locations	Number of Transcripts >=1 unique reads	R^2 for Technical Replicates	R^2 with 10ng Library
10 ng, TR1	6251585	67.79	18.46	13547	0.9508	
10 ng, TR2	20404270	68.70	18.01	15236		
1 ng, TR1	19807355	55.78	18.43	15151	0.9643	0.9615
1 ng,TR2	25119387	55.75	18.20	15306		
100 pg,TR1	13913778	59.35	17.92	12955	0.9016	0.8794
100 pg,TR2	13522446	61.24	17.99	12648		
50 pg, TR1	13378971	59.03	18.22	11986	0.8640	0.8565
50 pg, TR2	15297046	60.06	17.87	12002		
10 pg, TR1	14189544	27.46	16.88	9603	0.6102	0.7239
10 pg, TR2	13971891	31.03	12.62	9589		
1 pg, TR1	16038243	4.45	11.51	6531	0.1901	0.1002
1 pg, TR2	14281289	5.22	9.81	6465		

Supplementary Table S1B: Mapping summary of sequencing libraries made from different protocols using two different dosages of Activin A 3ng/mL (AA3) and 100ng/mL (AA100). Smart-seq mapping summary is given for one of the random sets obtained from all of the reads. Ribosome inhibition libraries were made from Activin A (3ng/mL; AA3) sample using DP-seq.

Amount of mRNA	Total Reads	% Of reads aligned to Refseq Transcripts	% Of unmapped reads aligned to genomic locations	Number of Transcripts >=1 unique reads	R^2 for Technical Replicates
Std. RNA-seq					
AA3; 10ng, TR1	18196250	81.21	15.82	17455	0.9755
AA3; 10ng, TR2	17638530	81.18	15.74	17380	
AA100; 10ng	17905346	79.50	16.77	17026	
DP-seq					
AA3; 50pg	24633672	58.59	17.65	13138	
AA100; 50pg, TR1	26108501	58.56	13.10	12910	0.8326
AA100; 50pg, TR2	27486701	65.27	14.53	13271	
Smart-seq					
AA3; 50pg, TR1	24272863	87.24	7.53	13798	0.8640
AA3; 50pg, TR2	26014738	86.89	7.30	13715	
AA100; 50pg, TR1	22298719	86.35	7.71	13400	0.8478
AA100; 50pg, TR2	24284435	87.25	7.43	13568	
Ribosome Inhibition (DP-seq)					
Primer Set 1; 500pg	21816975	67.93	21.90	14616	
Primer Set 2; 500pg	19668914	71.03	25.61	13246	
Primer Set 3; 500pg	10267103	68.49	27.05	11654	

Up-regulated in SB in comparison to AA15			
Term	PValue	Bonferroni	Benjamini
Neuron differentiation	1.79E-23	4.76E-20	4.76E-20
Neuron development	2.48E-17	6.59E-14	3.30E-14
Neuron projection development	5.72E-17	2.95E-13	9.81E-14
Forebrain development	4.04E-16	1.18E-12	2.95E-13
Axonogenesis	1.05E-13	2.79E-10	5.58E-11
Cell projection organization	4.54E-13	1.20E-09	2.01E-10
Neuron projection morphogenesis	1.53E-12	4.06E-09	5.80E-10
Axon guidance	1.71E-12	4.55E-09	5.68E-10
Cell motion	2.02E-12	5.35E-09	5.94E-10
Cell projection morphogenesis	2.36E-12	6.25E-09	6.25E-10
Neuron migration	4.17E-12	1.11E-08	1.01E-09
Cell morphogenesis involved in neuron differentiation	4.78E-12	1.27E-08	1.06E-09
Cell morphogenesis involved in differentiation	1.29E-11	3.42E-08	2.63E-09
Cell part morphogenesis	1.29E-11	3.42E-08	2.63E-09
Sensory organ development	6.41E-11	1.70E-07	1.21E-08
Cell morphogenesis	1.49E-10	3.96E-07	2.64E-08
Embryonic morphogenesis	5.74E-10	1.52E-06	9.52E-08
Pattern specification process	6.64E-10	1.76E-06	1.04E-07
Cell migration	2.70E-09	7.17E-06	3.98E-07
Up-regulated in AA15 in comparison to SB			
Tissue morphogenesis	5.66E-10	1.86E-06	1.86E-06
Tube morphogenesis	1.43E-08	4.68E-05	2.34E-05
Tube development	1.75E-08	5.74E-05	1.91E-05
Regulation of cell proliferation	4.47E-08	1.47E-04	3.67E-05
Muscle organ development	1.02E-07	3.34E-04	6.68E-05
Epithelium development	1.09E-07	3.59E-04	5.99E-05
Morphogenesis of a branching structure	7.34E-07	0.002407	3.44E-04
Embryonic development in birth or egg hatching	7.95E-07	0.002606	3.26E-04
Gastrulation	8.05E-07	0.002639	2.94E-04
Chordate embryonic development	1.27E-06	0.004150	4.16E-04
Muscle tissue morphogenesis	1.37E-06	0.004472	4.07E-04
Cardiac muscle tissue morphogenesis	1.37E-06	0.004472	4.07E-04
Cardiac muscle tissue development	1.64E-06	0.005377	4.49E-04
Blood vessel morphogenesis	1.79E-06	0.005852	4.51E-04
Epithelial cell differentiation	1.87E-06	0.006115	4.38E-04
Embryonic morphogenesis	2.26E-06	0.007406	4.95E-04
Formation of primary germ layer	2.54E-06	0.008290	5.20E-04
Endoderm development	2.68E-06	0.008762	5.18E-04
Striated muscle tissue development	3.44E-06	0.011238	6.28E-04
Heart morphogenesis	3.63E-06	0.011851	6.27E-04

Supplementary Table S2: GO (Biological Process) Enrichment for genes differentially regulated in SB/AA15. Genes up-regulated in SB are enriched for ectoderm related terms while genes up-regulated in AA15 are enriched for mesoderm and endoderm related terms. P-values were determined from background set of genes that showed expression in SB/AA15 samples.

Up-regulated in SB in comparison to AA15		
Term	PValue	Fold Enrichment
Axon guidance	1.57E-08	3.70
Pathways in cancer	1.51E-05	2.14
Focal adhesion	1.18E-04	2.35
Wnt signaling pathway	3.19E-04	2.50
Basal cell carcinoma	5.50E-04	3.73
Colorectal cancer	5.60E-04	3.04
Pancreatic cancer	0.001349	3.11
Notch signaling pathway	0.004505	3.36
TGF-beta signaling pathway	0.006142	2.57
ErbB signaling pathway	0.006142	2.57
Melanogenesis	0.006550	2.42
Adherens junction	0.006669	2.70
Chronic myeloid leukemia	0.006669	2.70
Hedgehog signaling pathway	0.007261	3.11
Non-small cell lung cancer	0.007261	3.11
Biosynthesis of unsaturated fatty acids	0.012782	4.15
Small cell lung cancer	0.014350	2.41
Endometrial cancer	0.019416	2.87
Prostate cancer	0.020781	2.28
Regulation of actin cytoskeleton	0.021865	1.72
Chondroitin sulfate biosynthesis	0.027081	4.24
ABC transporters	0.030854	2.90
Renal cell carcinoma	0.031684	2.40
MAPK signaling pathway	0.043071	1.55
VEGF signaling pathway	0.048220	2.21
Up-regulated in AA15 in comparison to SB		
Glioma	0.001299	2.74
Pathways in cancer	0.002808	1.59
Melanoma	0.003446	2.47
Alanine, aspartate and glutamate metabolism	0.007788	3.34
Arginine and proline metabolism	0.007922	2.60
Cysteine and methionine metabolism	0.013292	3.04
p53 signaling pathway	0.019104	2.18
Amino sugar and nucleotide sugar metabolism	0.020808	2.56
ABC transporters	0.023613	2.51
Fatty acid metabolism	0.023613	2.51
Non-small cell lung cancer	0.024969	2.32
MAPK signaling pathway	0.029124	1.46
Endocytosis	0.029288	1.55
Nitrogen metabolism	0.031517	3.27
Tight junction	0.037866	1.67
Focal adhesion	0.040549	1.52
Glycolysis / Gluconeogenesis	0.040867	2.03
Bladder cancer	0.045432	2.39

Supplementary Table S3: Kegg Pathways enriched in SB/AA15 samples. P-values were determined from background set of genes that showed expression in SB/AA15 samples.

1. cccagtg	1. caaagcc	1. cacacac
2. ccccaga	2. caacccc	2. cagcagc
3. cccccaa	3. cccagca	3. ccaccag
4. ctcccca	4. cccccaa	4. cccagca
5. cttcacg	5. ctcgtcc	5. cccccaa
6. gcaacag	6. cttcccc	6. ccttccc
7. tgacagc	7. gcctctc	7. cttcccc
8. tggctct	8. gcctctg	8. gcaacag
9. tggcttc	9. gcgaact	9. gcctcag
10. tccctcc	10. tcagccc	10. tccctcc
11. ccttccc	11. tctccga	11. tgaccca
12. cagaccc	12. tgccatc	12. tgagcct
13. gcaaacc	13. tgccttg	13. cagcact
14. ccaggac	14. tgagcct	14. gcgaact
15. cacacac	15. tcctcgt	15. ctcccag
16. tctccga	16. tctgcct	16. gccaaag
17. cctccca	17. ctgccct	17. ccccaga
18. tgaccca	18. tgccact	18. tcagcca
	19. cttcacg	19. gaagcca
	20. gcaacag	20. tgacagc
	21. cctctgc	
	22. gcaaacc	
	23. ccccaga	
	24. ctcagca	
	25. tgacagc	

Supplementary Table S4: List of heptamer primers used for our sequencing-library generation. 44 unique primers were split into three tubes with some primers repeated in different tubes to get coverage ≥80% mouse transcriptome.

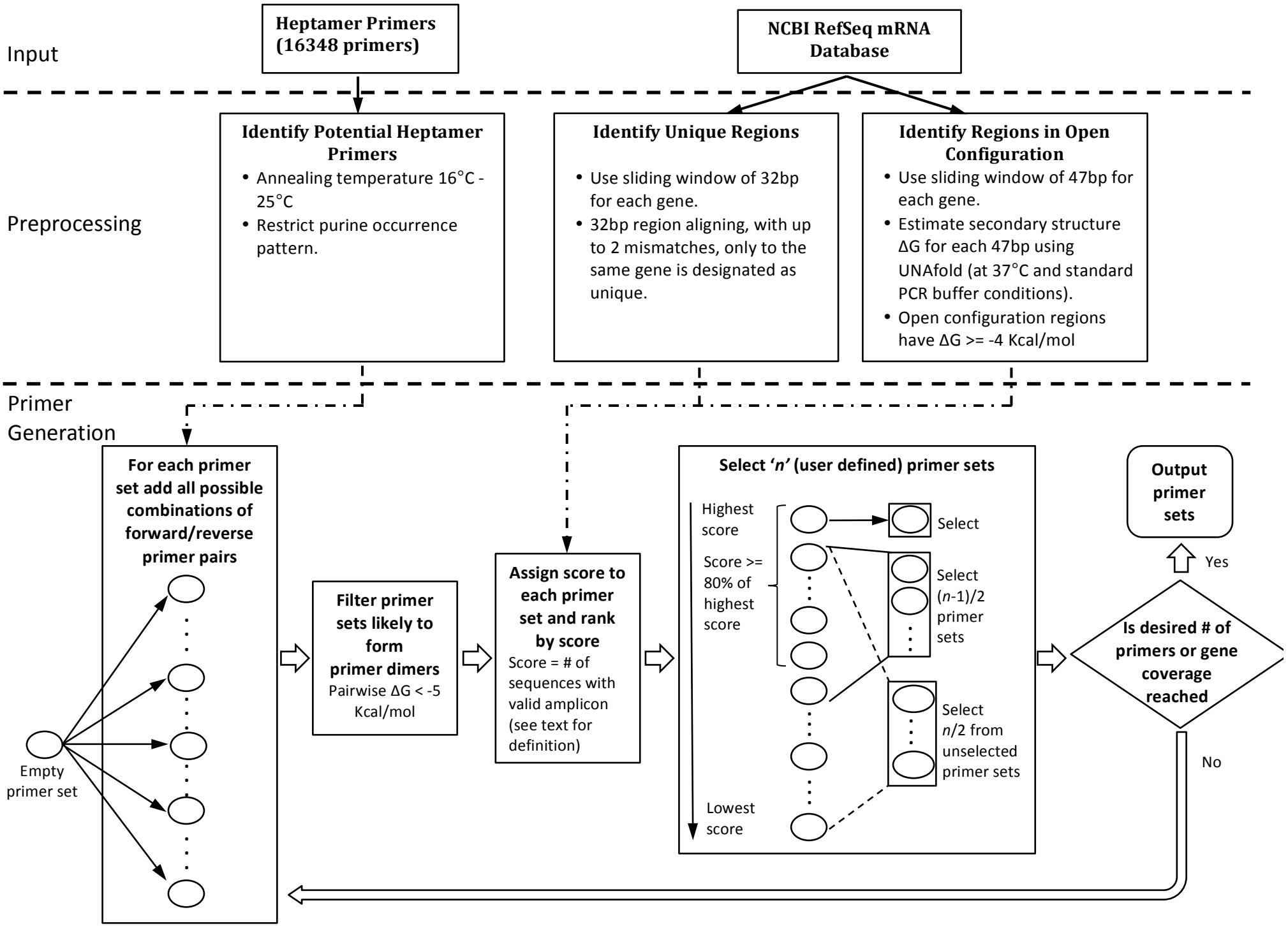
Supplementary Table S5: Primer Sets for Ribosomal Inhibition

Primer Set 1	Primer Set 2	Primer Set 3
CCTCCTG	GGACAGC	GAAAGCC
GCAGCCT	CACACAC	CACACAC
TCCCACA	GCAACAA	CCACACA
CACACTG	GCATGTG	TGCTGTG
CTTCCCC	GTGACCT	GACAACC
CCACCAC	CATCAGC	GTCACAC
CCTCCCC	CTTGAGC	GACACAC
CTTGCAG	TACAGCC	GCGTTTT
CCCACAC	GTTCTCG	GAGCCTC
CCTTCCC	CAAGCAC	GTGATGC
CACCCCA	TCAGCAC	CCGTCTT
CTCTCCC	TCGTTCC	TCCCTCA
CAGAGCC	GCGTCTG	GTTTCCG
CCCCAAA	CAAACCG	TCCAACC
CTCCCCA	CCGTGAC	CGAATGG
CAAGAGC	TGTCTCG	CCGTGTA
CCCTGGA	GCGTCAG	CAAACCG
CCCCCTC	CCCCTAC	CGAGTGT
CCCCTCA	CCGTGTA	GACTCCG
CTGAGCT	CCGTTGA	GCGAATT
CCCCCAG	GATCCCG	GGTGCCC
	CCGACTT	CGAGAGC
	GCGACAC	GATGCGT
	CTGAGCG	CGACTCA
		GCGTTAG
		CAGTACG
		GAATGCG
		CCGTGCT
		CAACCGA
		TGCTACG
		GTAACCG
		TGCCGAT
		CCCGTTA
		TAGAGCG
		CAAGCGT
		CGATAGC
		GACCGAC
		CGATCCC
		CGAGTGC
		CGATTGC

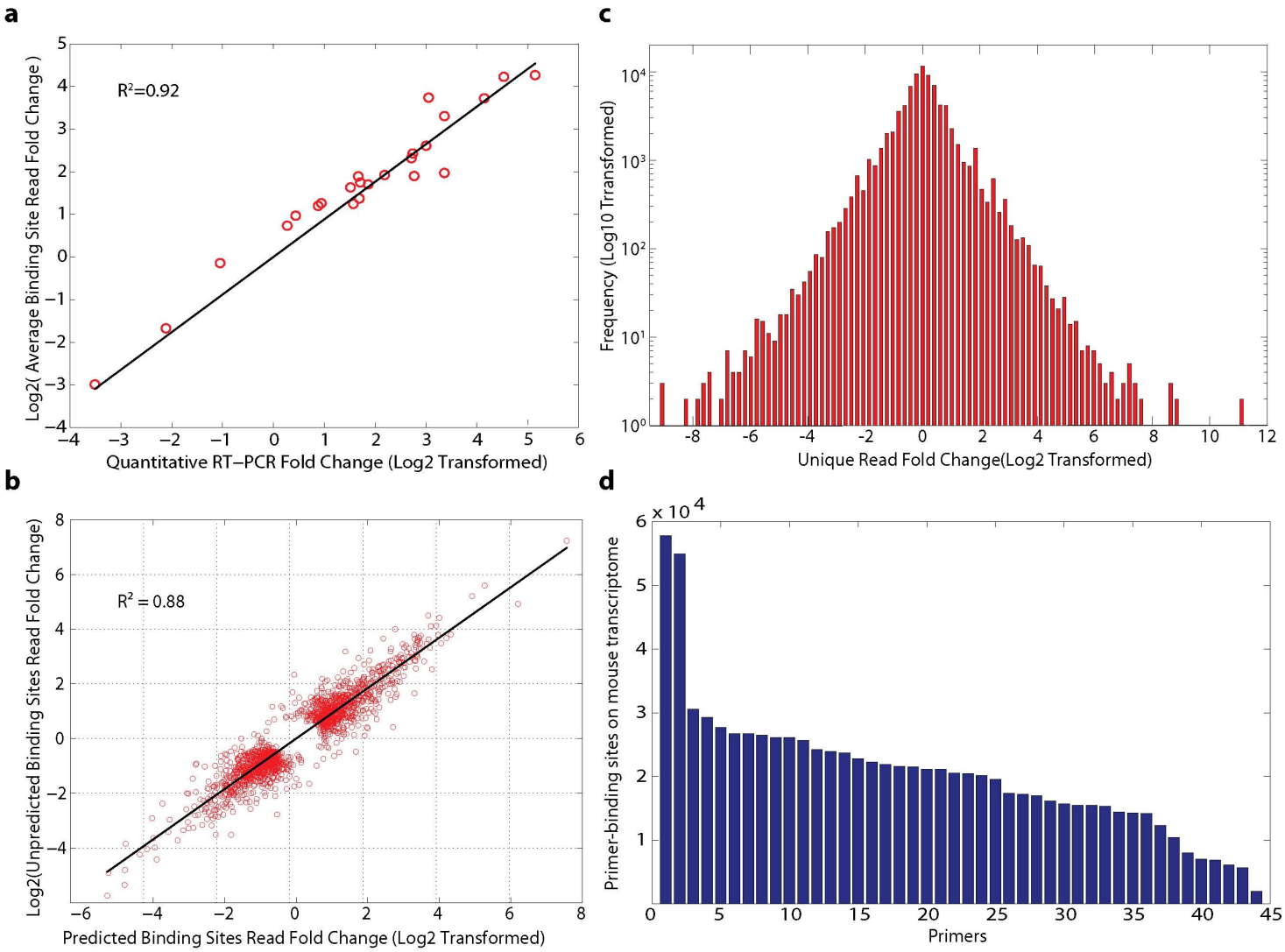
Gene	Forward Primer	Reverse Primer
Lefty1	CGCTGAATCTGGGCTGAGTCCC	GCCTAGGTTGGACATGTTTGCCCA
Lefty2	TGCAAGTAGCCGACTTCGGAGC	CCTATTCCCAGGCCTCTGGCCA
Gsc	GGGGGTCGAGAAAGCAACGAGG	ACGAGGCTCACGCAGGCAGC
Flk-1	AGAGGAAGTGTGCGACCCCAA	CACTGGCCGGCTCTTTCGCTT
Oct4	TGAAGTGCCCGAAGCCCTCCCTA	GCCCTTCTGGCGCCGGTTACA
Mesp1	TCTAGAAACCTGGACGCCGCC	TCCGTTGCATTGTCCCCTCCAC
T	CTCCGATGTATGAAGGGGCTGCT	GCTATGAGGAGGCTTTGGGCCG
Foxa2	CCCCATGCCAGGCAGCTTGG	AAGTGTCTGCAGCCAGGGGC
Sox1	TTCCCCAGGACTCCGAGGCG	GTTCAGTCTAAGAGGCCAGTCTGGT
Arx	AAGCATAGCCGCGCTGAGGC	TTCGGGGAACGCCCTAGGGG
Lnsm1	TACAGCTCCCCGGGCCTGAC	ACTCTAGCAGGCCGGACGCA
Pax6	ACCTCCTCATACTCGTG	ACTGATACCGTGCCTT
Dbx1	GACGTGCAGCGGAAAGCCCT	CGCTAGACAGGAGCTCGCGC
Dmrt3	AACCGGCCACCCCTGGAAGT	GTCGCCCCCGCAACCTTTCA
Hes5	TCCGACCCCGTGGGGTTGTT	TCTACGGGCTGGGGTGAGCC
Neurog2	ACACGAGACTCGGGCGAGCT	CCGGAACCGAGCACGGTGTC
Lhx2	TGGGCTCAGCCGGGGCTAAT	ACAGCTAAGCGCGGCGTTGT
Pax5	ACACTGTGCCCAGCGTCAGC	GCACTGGGGGACGTGATGCC
Lhx5	GAGCTCAACGAAGCGGCCGT	CCGAGAAATTGCGCAGGCGC
Sox2	GCACATGAAGGAGCACCCGGA	GGTTCACGCCCGCACCCAG
Asb5	GGGACACGCCACTGCATGCT	GCCAAGTCGACAGGCCGCAA
Lmx1a	TGACGTCATGCCCGGGACCA	GCCCCCTACACCCGCCTCAT
Pax3	CCCCCACCTATAGCACCGCAGG	ACATGCCTCCAGTTCCCCGTTCT
Hoxa5	AGGGAACCGAGTACATGTCCCAGT	TGCAACTGGTAGTCCGGGCCA
Triml2	TGCGCAGCCTCCAGACGATG	TCTGGAGCAGTGCAACGGCA
Afp	TTCCTCCCAGTGCGTGACGGA	TCCTCGGTGGCTTCCGGAACA
Dppa3	CCGGCGCAGTCTACGGAACC	ACCGACAACAAAGTGCGGACCC
Fgf8	GCGAAGCTCATTGTGGAGAC	CACGATCTCTGTGAATACGCA
Nodal	ACCAACCATGCCTACATCCAGAG	CCCTGCCATTGTCCACATAAAGC
Epha1	TACGCCTGCCCAGCCTGAGT	GGTGTCCAGCCCAGCCGAAC
Rab25	TCAGCCAGGCCCGAGAGGTC	GATGGCACTGGTCCGGGTGC
Evx1	GAGTGGCGTCACCAGCGGTACT	TCACCTTGTGATGCGAGCGC
Lrrc6	GGGAAATCCTGCCTGCCGGTC	CTGTGATTCGGCCCATGGTGCTT
Pou6f1	CGCCTTTCCTGCCTGGTGGG	GCTAGCAGTGGGCAGTGGCC
Pgr	CGCCATCTACCAGCCGCTCG	ACTGTGGGCTCTGGCTGGCT
Foxa3	TTTGGGGGCTACGGGGCTGA	TGCAGCCCACGCCCATCATG
Ell2	TGCAGGCCTCCTACCACCCC	TCCCCAGGCCTTCTGGAGTGC
Lbh	ACGTTGGGGCAAGAGCGTGG	GAGACGGGGGAGGGGGTGAC
Etv4	GAAGGTGGCTGGCGAACGCT	GCGGGGCCAGTGAGTTCTGG
Klf9	CCGCGTACTCGGCTGATGCC	CACACGTGGCGGTCGCAAGT
Wnt3a	ACCAAGACCTAACAAACCC	CATGGACATCACGGACC
Prdm1	GCCGAGGTGCGCGTCAGTAC	GGGGCAGCCAAGGTCGTACC
Ankrd1	ACGCAGACGGGAACGGAAGC	TGCGGCACTCCTGACGTTGC
Per2	GGTGGCCTCTGCAAGCCAGG	CCTCCGTGCTCAGTGGCTGC
Hes1	CCCTGCAAGTTGGGCAGCCA	CGAAGGCCCCGTTGGGGATG
Bnc1	GCTGGAGCACCTGGGTGAGC	CCTCCACTGTGCACGCGTGT
Foxc2	AGGGACTTTGCTTCTTTTTCCGGGC	CCCGCAGCGTCAGCGAGCTA
Prdm6	CCGGCCTTTCAAGTGCGGCT	GGCATGCGCTGGTGTCGACT

Armc4	GCATCCCCTTGCTGGCTCGG	GGCCATGGCACAGTGCTCCT
Cxcr4	TACCCCGATAGCCTGT	GCACGATGCTCTCGAA
Tbx3	CCAAGCGATCACGCAACGTGG	CTCTGACGATGTGGAACCGCGG
Arg1	GCGAGACGTAGACCCTGGGG	GGTCGCCGGGGTGAATGCTG
Foxq1	GGAGCCGCCGCAGGGTTATATTG	TGGCGCACCCGCTACTTTTGAG
Asb4	TCACCTCCGTGCGTCCTGCT	TTCGGGCAAGAGTGGCAAGCC
Six2	ACTCGTCGTCCAGTCCCGCTC	CAAGGTTGGCCGACATGGGGT
Lhx1	ACTAGGGACCGAGGGACGCG	CAGTTTGGCGCGGATTGCCG
Sox17	GAGCCAAAGCGGAGTCTC	TGCCAAGGTCAACGCCTTC
Cer1	AGAGGTTCTGGCATCGGTTCA	TCTCCCAGTGTACTTCGTGGC
Creb3l1	ACAGGACGGACACCCTGGCA	GGTCAGCCCAGGGGAGCAGT
Bcl6	AAGCACGGCGCCATCACCAA	TTTGGGGAGCTCCGGAGGCA
Hey1	AATGGCCACGGGAACGCTGG	CACCACGGGAAGCACCGGTC
Basp1	AGGGGGCGGGGAGAATCCAAA	GGAGCCTAGGGGACAGCGGTT
β-Actin	GCTGTATTCCCCTCCATCGTG	CACGGTTGGCCTTAGGGTTCAG

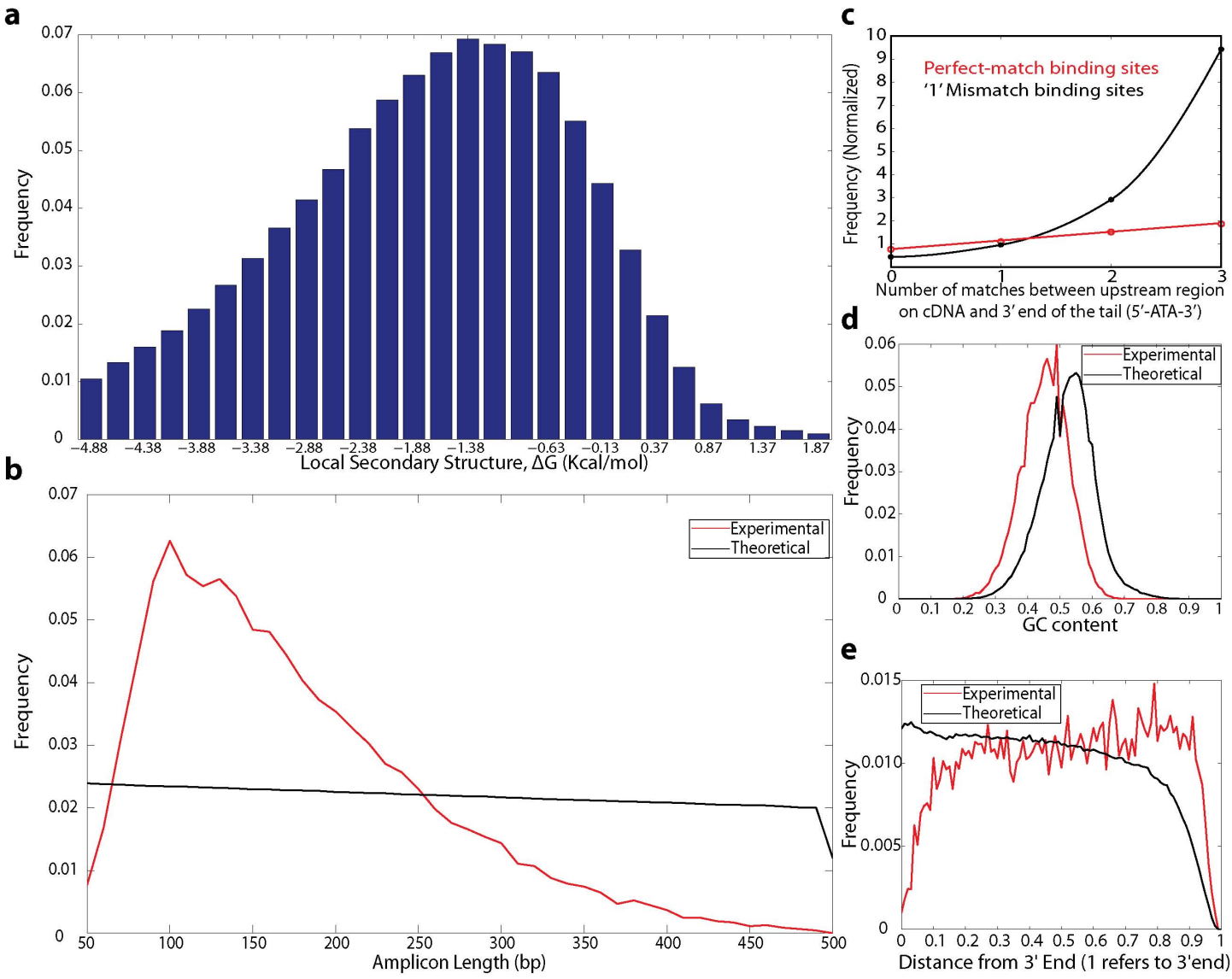
Supplementary Table S6: List of quantitative RT-PCR primers used in the study.



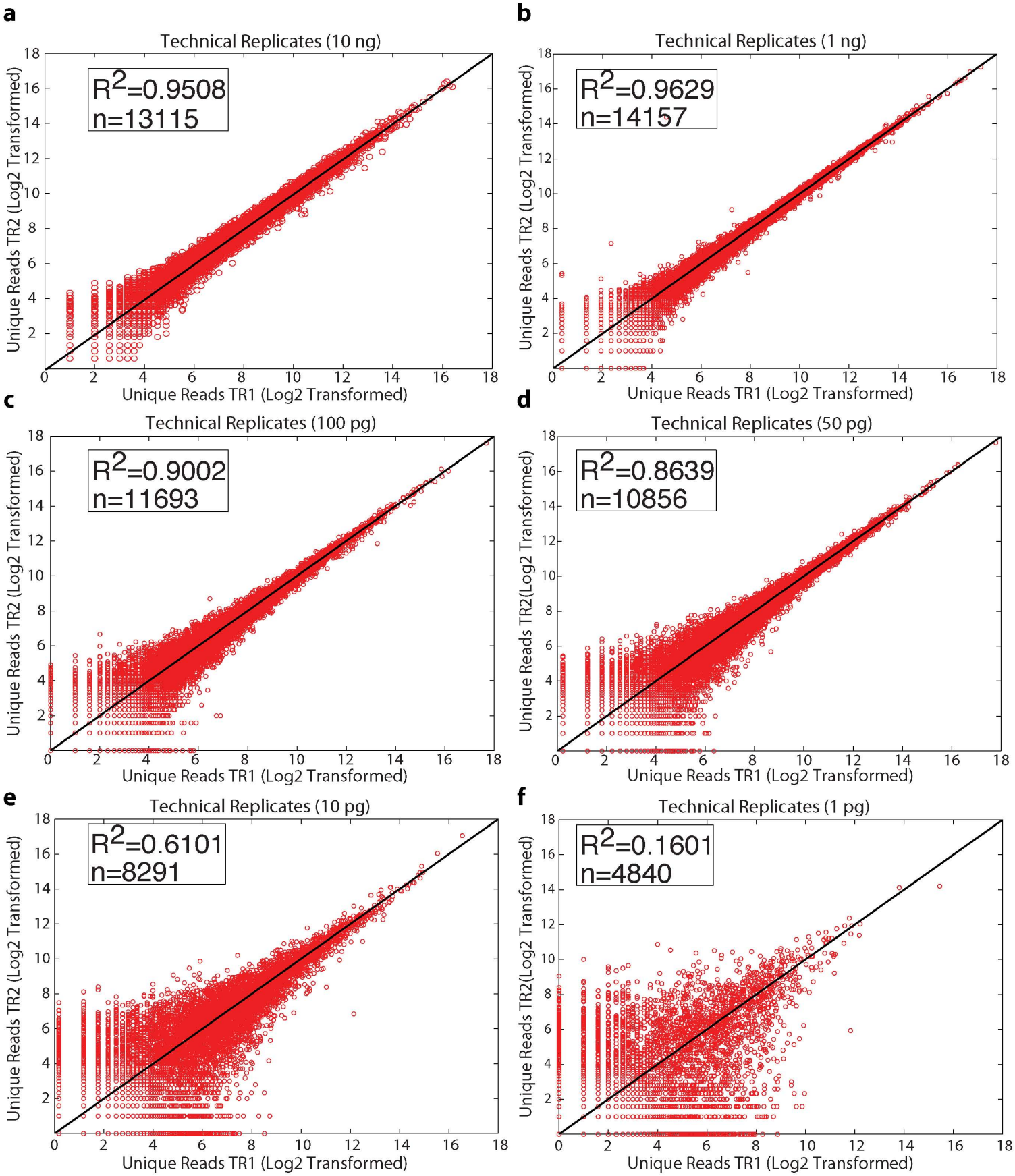
Supplementary Figure S1: Flowchart of heptamer primer generation using an iterative randomized algorithm.



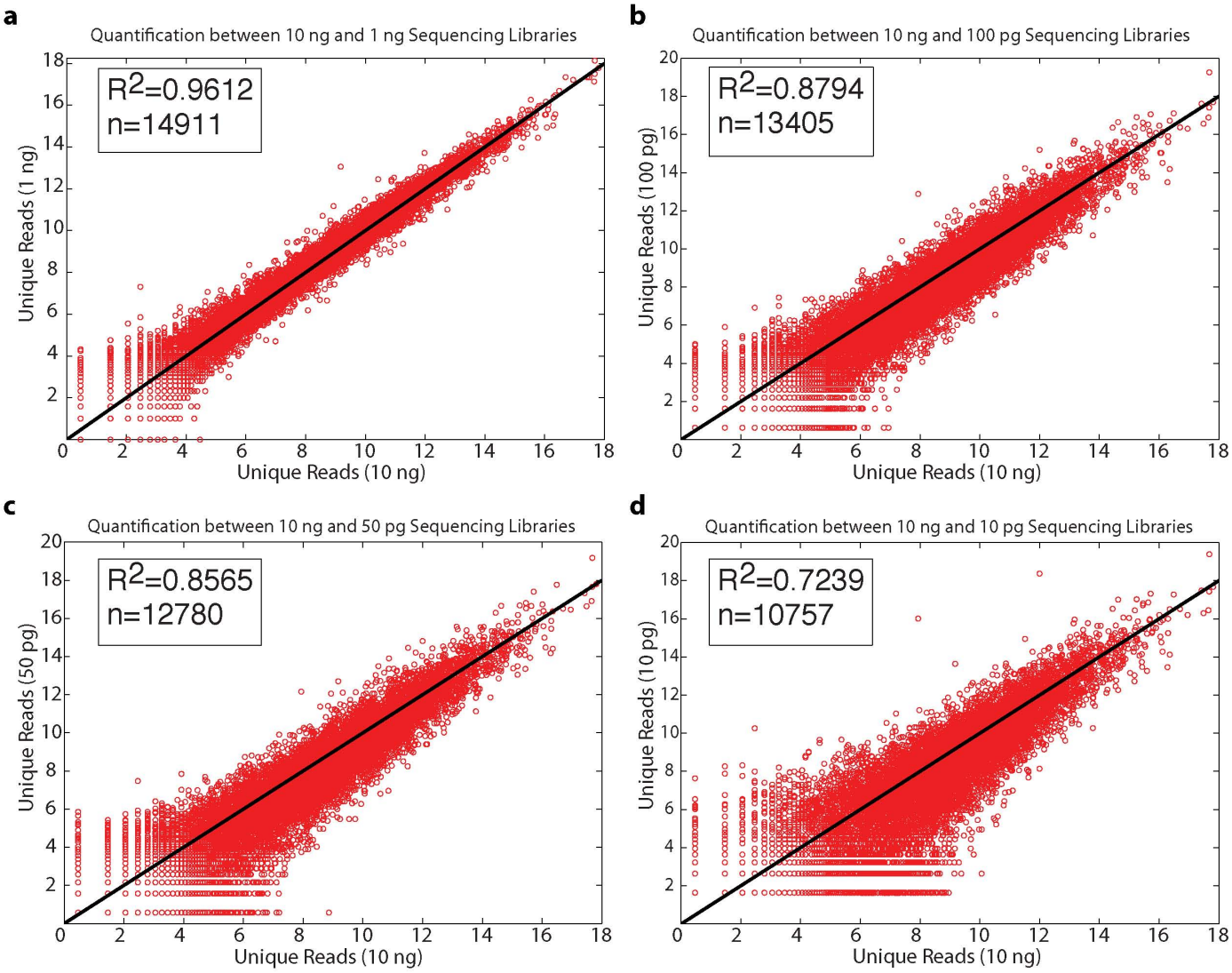
Supplementary Figure S2: Performance of heptamer primers based amplification strategy. (a) Expression measurements obtained from multiple heptamer primer-binding sites on a transcript provided independent measurements of the relative abundance of the transcript. The average fold change obtained from multiple primer-binding sites for a transcript was in concordance with quantitative RT-PCR (n=24). (b) Mis-primed PCR products maintained relative abundance of the gene expression. Fold changes observed in predicted vs. mis-primed binding sites for differentially expressed transcripts (in SB vs. AA100) showed strong correlation. (c) Distribution of fold changes observed across all of the samples. The majority of the trancripts were not differentially regulated. Our methodology captured fold changes in range of 2^{-8} – 2^{10} demonstrating broad dynamic range. (d) Distribution of heptamer primer-binding sites on the mouse transcriptome.



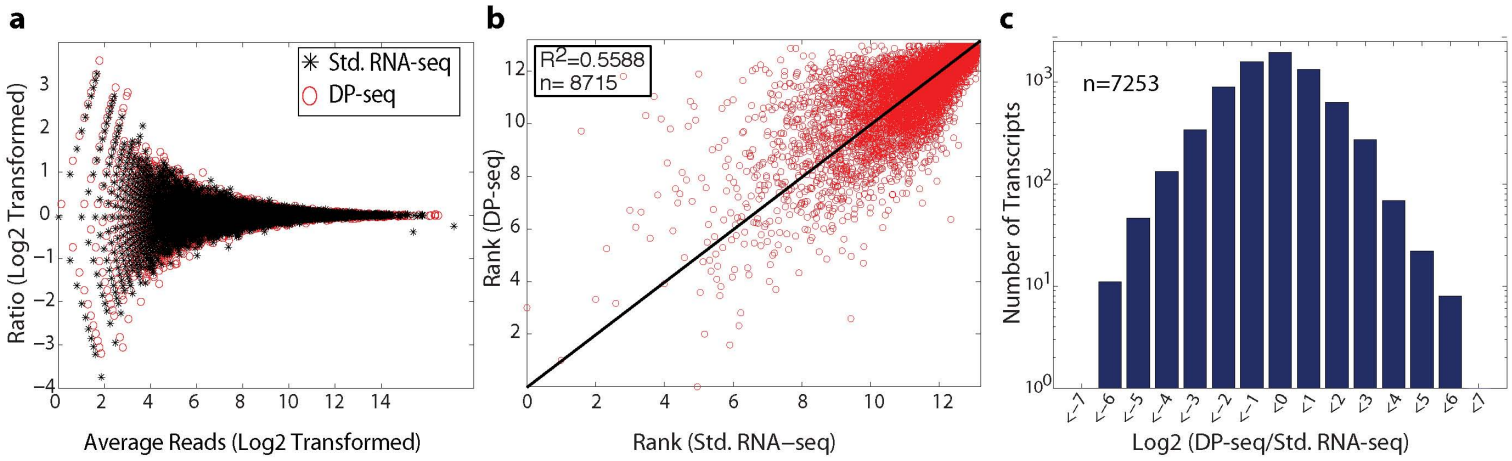
Supplementary Figure S3: PCR biases observed in our methodology. (a) PCR bias caused by the secondary structure of the cDNA. The distribution is shifted towards high Gibbs free energy (ΔG) implying that the primer-binding sites forming stable secondary structure shielded heptamer primers from annealing to their target sequences. (b) Bias towards shorter PCR fragments. The black curve represents the distribution estimated for all theoretically possible amplicons from the 44 heptamer primers in the mouse transcriptome. The experimental curve dropped sharply around 100bp because of the size selection step performed at the last stage of the sequencing library generation. (c) Tail Interaction. Heptamer primer binding sites with '1' mismatch had significantly higher tail interaction as compared to perfectly matched primer-binding sites. (d) GC bias. The amplicons with lower GC content are preferentially amplified. (e) PCR bias caused by reverse transcriptase. Majority of the primer-binding sites came from 3' end of the genes mainly because of the inability of the reverse transcriptase to produce full-length first strand cDNA.



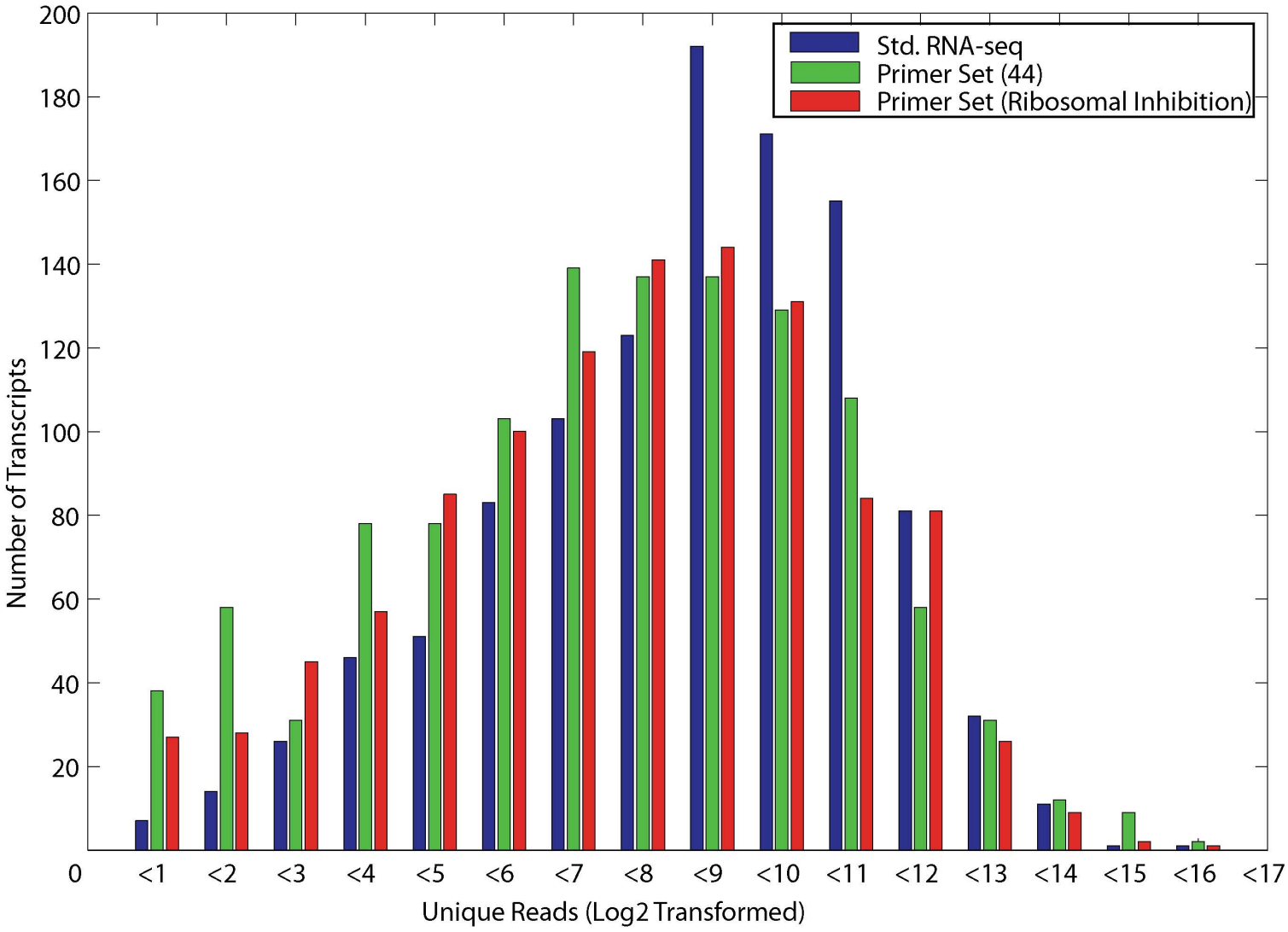
Supplementary Figure S4: Techincal Replicates for sequencing libraries prepared from various amounts of starting material (mRNA). The transcriptome coverage dropped with lower amounts of mRNA. Significant technical noise was observed for the sequencing libraries prepared from 1 pg of mRNA.



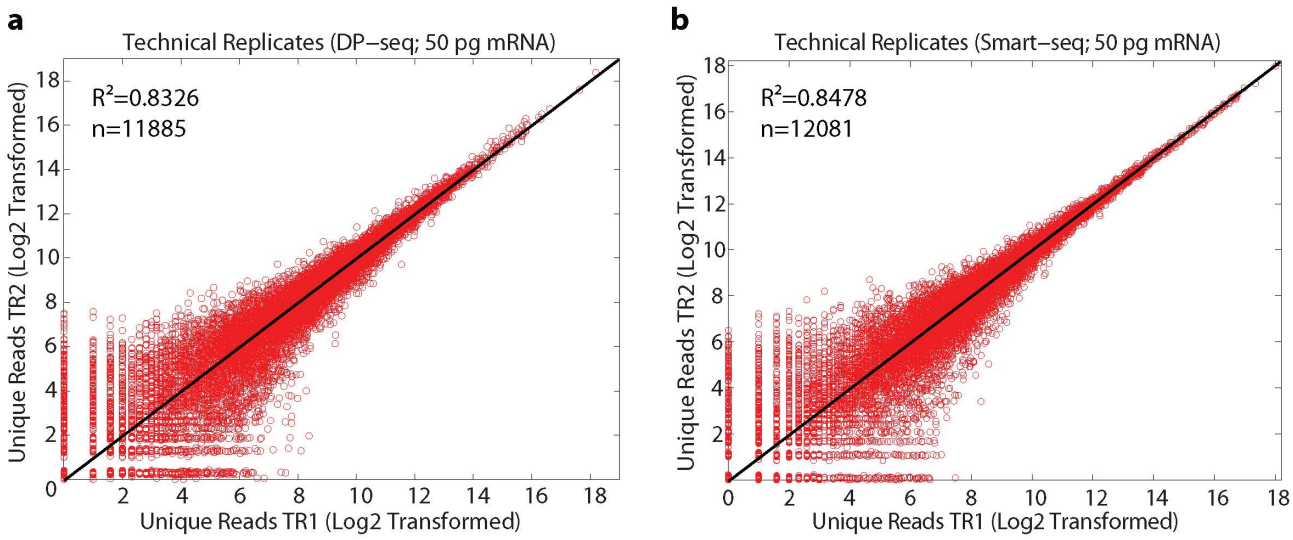
Supplementary Figure S5: Transcript representation is conserved with serial dilutions of the starting material (mRNA). Transcripts abundance obtained from dilutions (1 ng, 100 pg, 50 pg, 10 pg) were compared with respect to highest concentration of 10 ng.



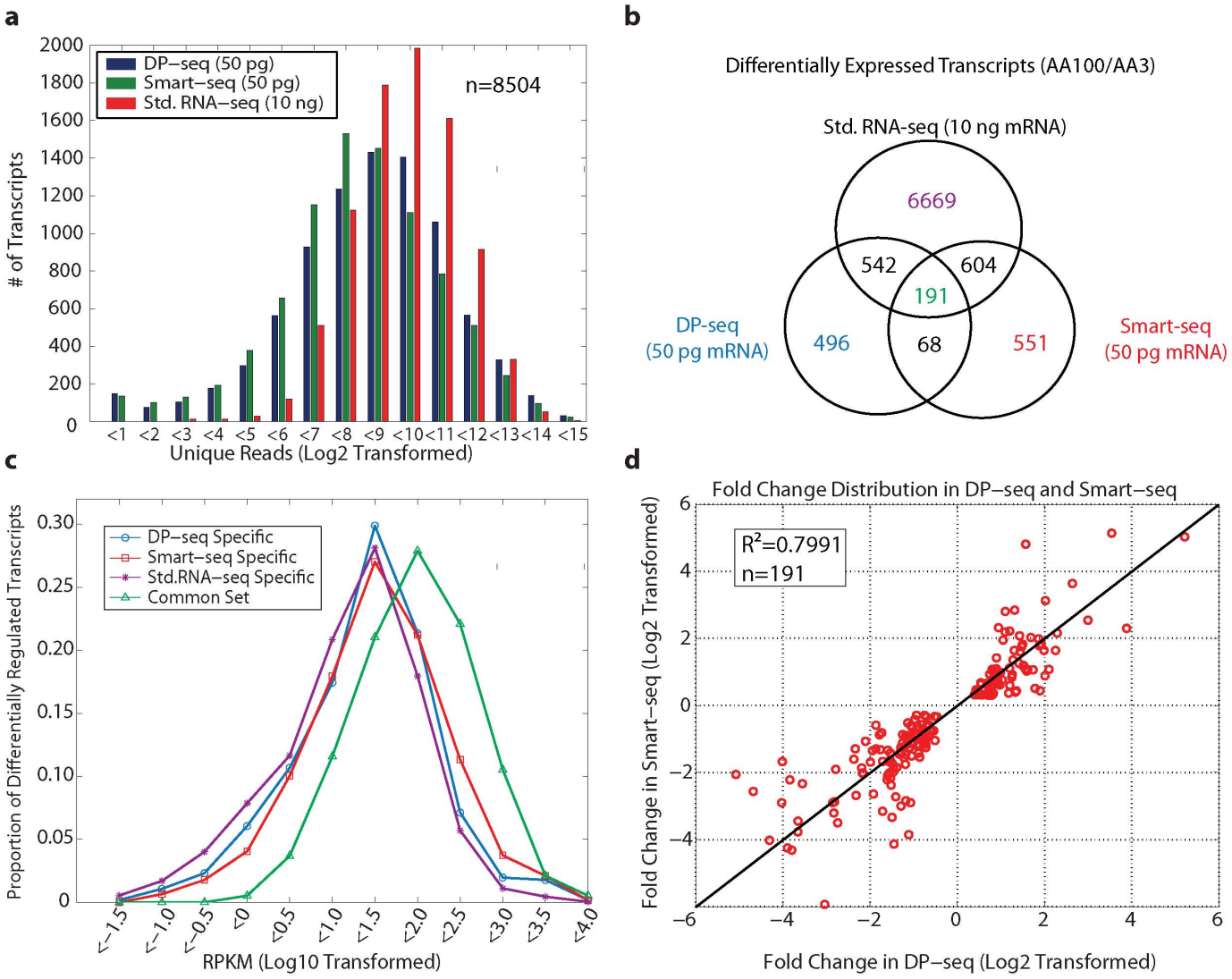
Supplementary Figure S6: (a) Std. RNA-seq exhibited similar technical noise in the technical replicates as DP-seq. (b) PCR biases observed in our protocol distorted the order of transcript expression resulting in poor Rank Correlation with respect to the Std. RNA-seq. (c) Distribution of the ratio of unique reads obtained for the low expressed transcripts (RPKM<=10) in DP-seq and Std. RNA-seq.



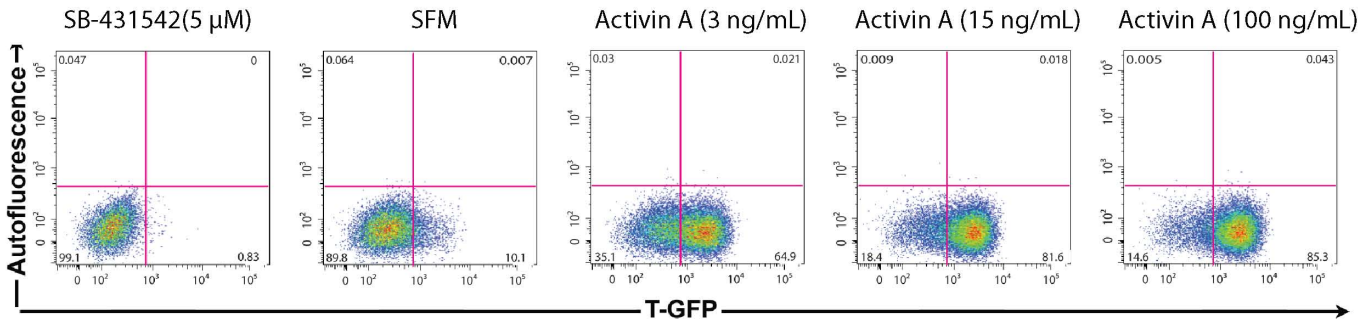
Supplementary Figure S7: Sequencing libraries prepared from Std. RNA-seq and DP-seq (44 primer set and a primer set used for suppression of the ribosomal transcripts) displayed overlapping distributions of reads mapping to the mouse transcription factors (n=1148; AA3 sample).



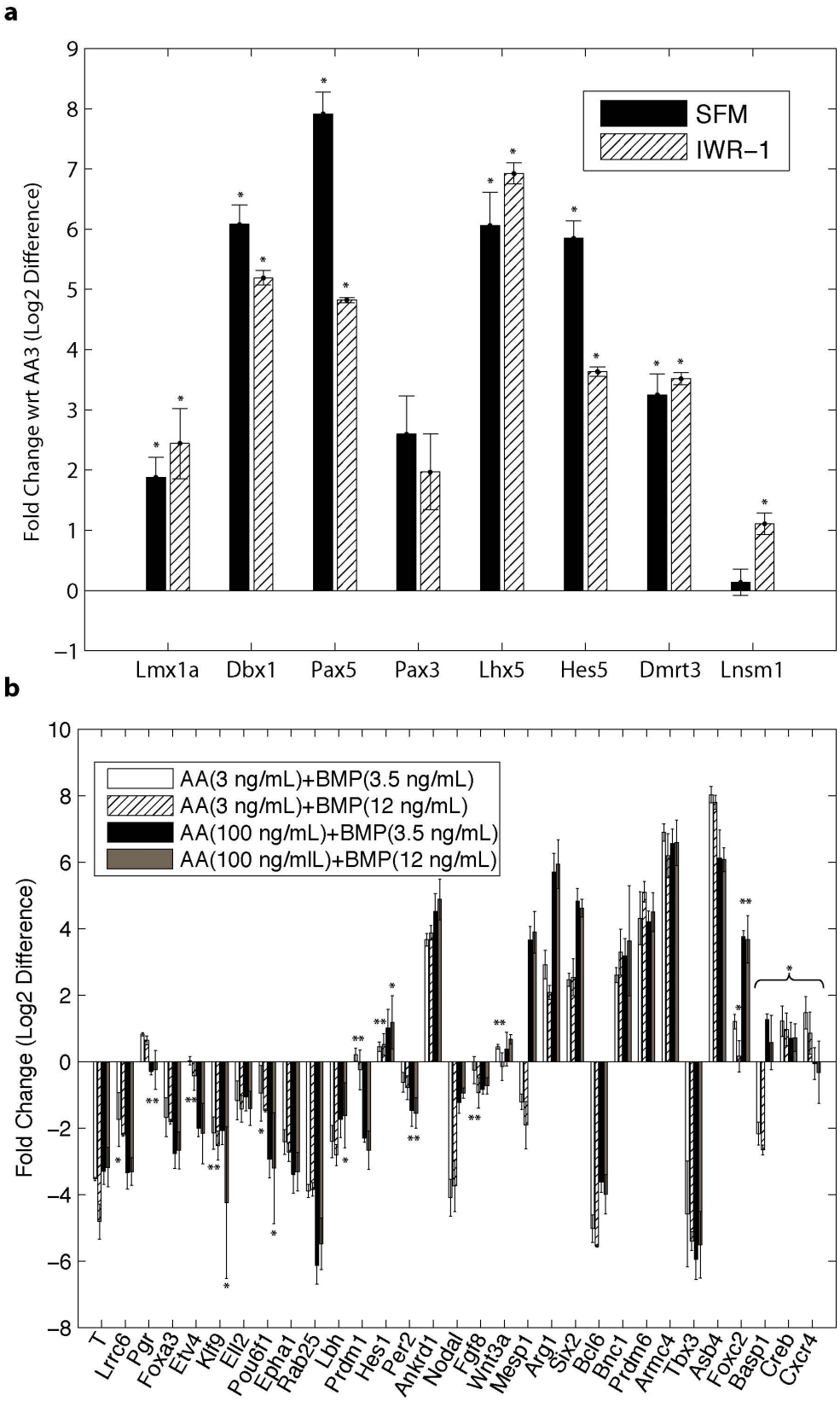
Supplementary Figure S8: Technical replicates prepared from 50 picograms of mRNA derived from Activin A 100ng/mL dosage exhibited high correlation in expression measurements for DP-seq and Smart-seq.



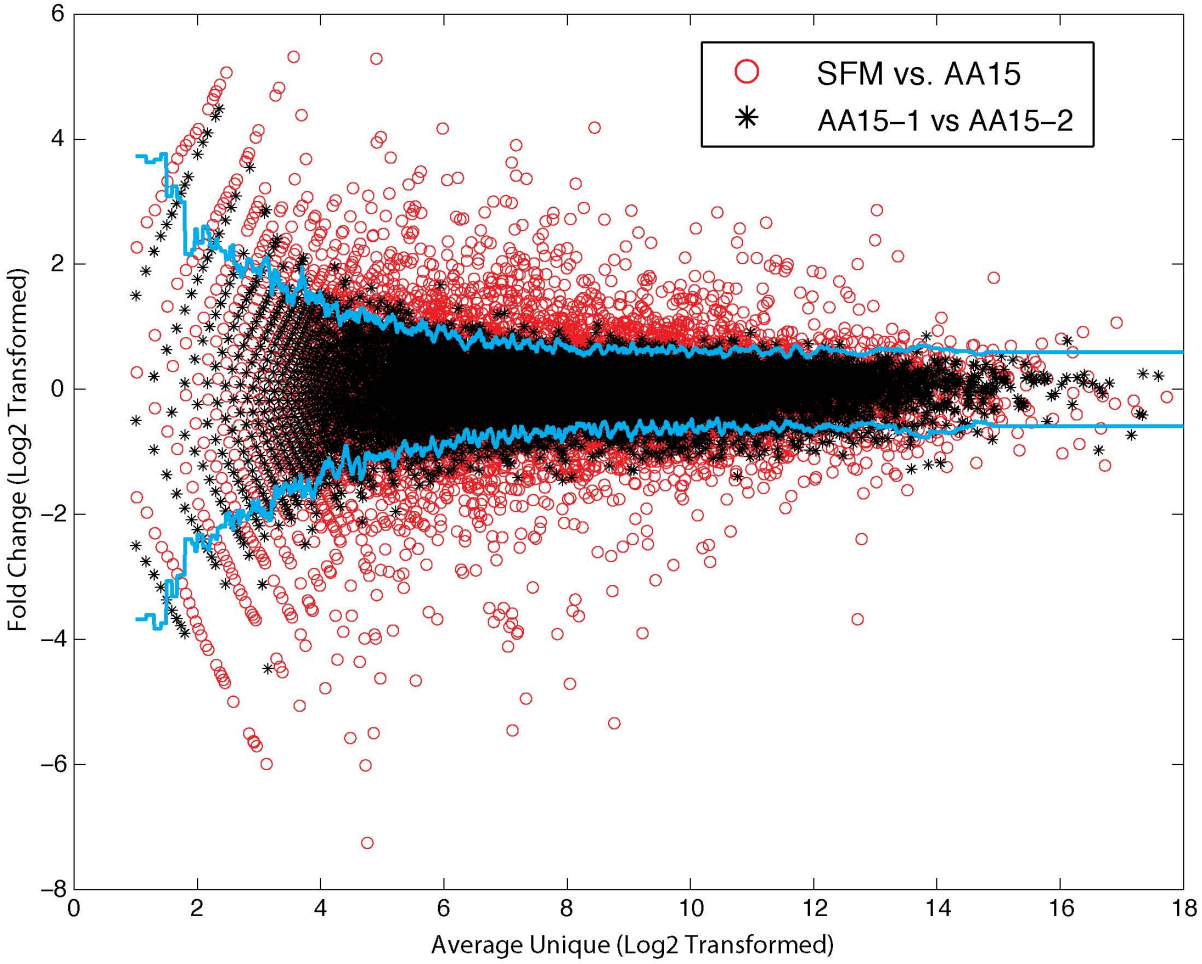
Supplementary Figure S9: Comparison of the sequencing libraries prepared from DP-seq, Smart-seq and Std. RNA-seq methods. (a) Histogram of unique reads obtained for the moderately expressed transcripts (3<RPKM<300) in the three methods. The amounts of mRNA used for the sequencing library generation are mentioned in the parentheses. (b) Venn diagram depicting the overlap of the differentially expressed transcripts between Activin A 100ng/mL and 3ng/mL dosages identified in the three methods. (c) The expression profile of the common set (green) is shifted towards higher RPKM as compared to the method specific differentially expressed transcripts. (d) Correlation in fold changes for the common set between DP-seq and Smart-seq. The RPKM measurements were made from Std. RNA-seq experiment performed on AA100 sample.



Supplementary Figure S10: Flow cytometry on T-GFP mESCs at day 4 of differentiation upon treatment with SB and Activin A. Graded activation of Activin A/TGFβ signaling pathway led to increased expression of mesoderm marker, T.



Supplementary Figure S11: (a) Validation of neuro-ectoderm specific genes by using small molecule inhibitor of Wnt Signaling pathway, IWR-1 to efficiently induce neuro-ectoderm in an in-vitro differentiation model. The quantitative RT-PCR fold changes were normalized to Activin A (3 ng/mL) dosage. Error bars represent standard deviation in biological replicates (n=3). Asterisks indicates p>0.05 (Student's t-test) compared with controls. (b) Expression profiles of Primitive Streak markers in response to BMP4 signaling. Quantitative RT-PCR fold changes for two BMP4 dosages (3.5 and 12 ng/mL) were normalized with respect to Activin A alone induction. Error bars represent standard deviation in biological replicates (n=3). Asterisks indicate p>0.05 (Student's t-test) compared with controls.



Supplementary Figure S12: Identification of the differentially expressed transcripts. Baseline distribution was determined from MA plot of the technical replicates. Experimental MA plot of untreated control vs. Activin A (15 ng/mL) was overlaid on top of the baseline distribution. The blue curve represents p-value threshold of 0.05 and experimental ratios above/below the curve were designated as differentially regulated.