Supplementary Information

Dynamic Transcriptomes during Neural Differentiation of Human Embryonic Stem Cells Revealed by Short, Long, and Paired-end Sequencing

SUPPLEMETARY FIGURES:

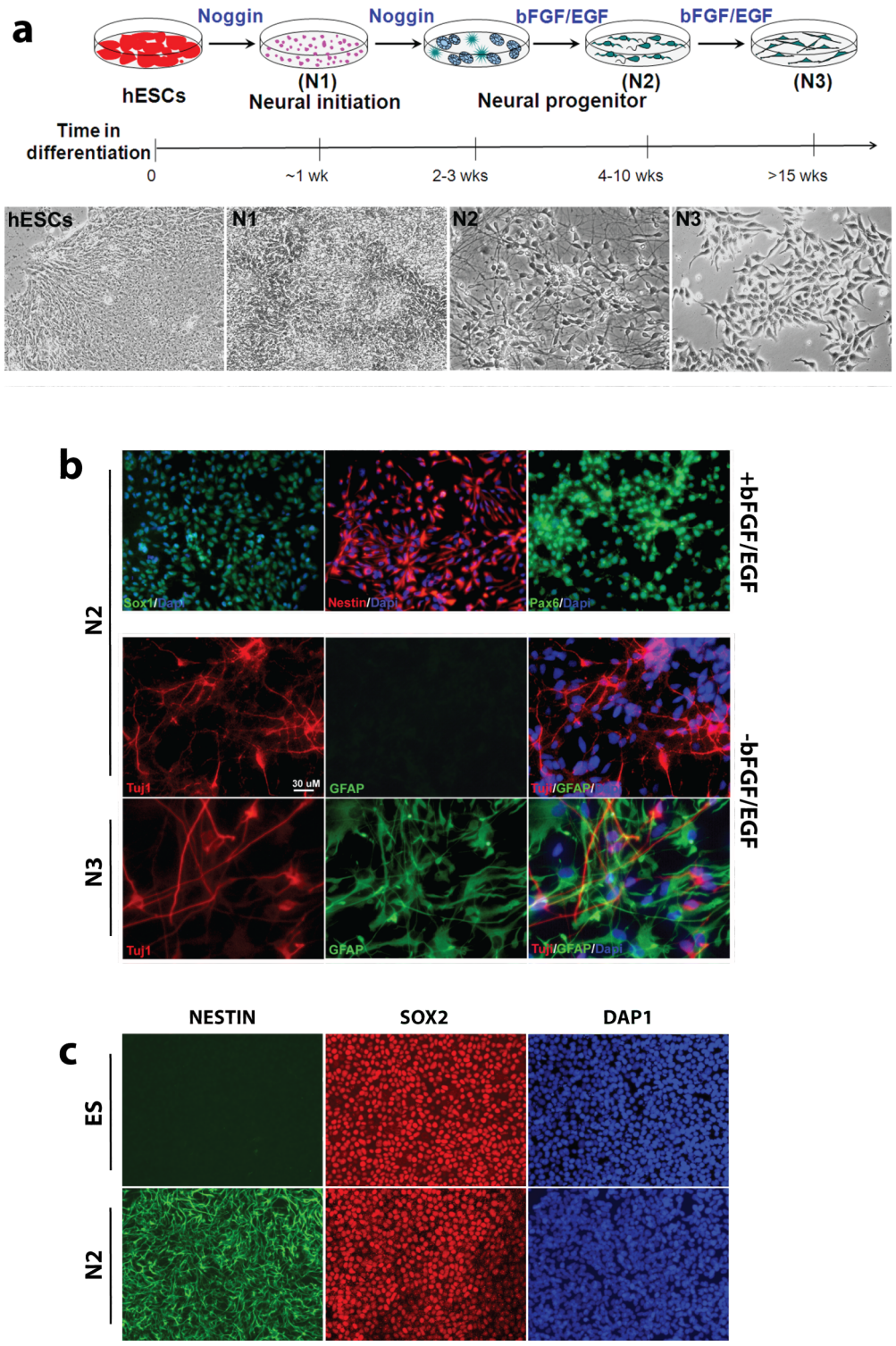

Figure S1. High-resolution photos of Fig. 1. Please see Fig. 1 legend.

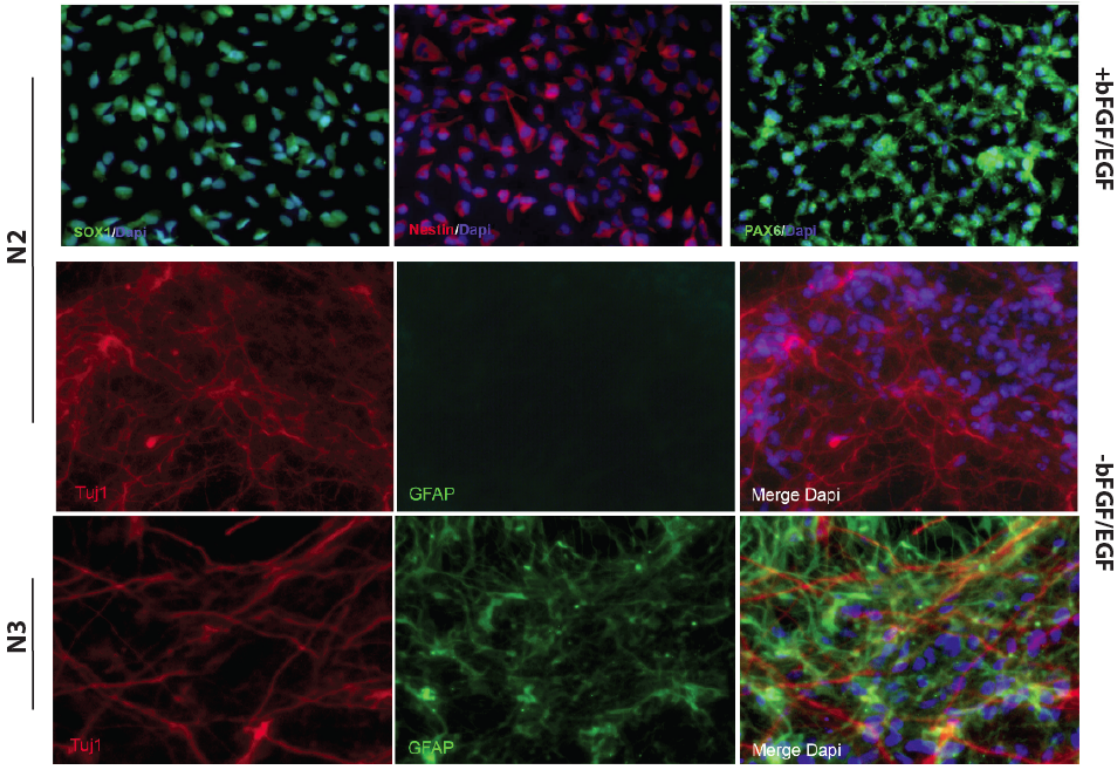

Figure S2. H7 hESC cells have similar differentiation potential to that of H1 cells. Immunofluorescence labeling of genes expressed in N2 and N3 stages of H7 cells prepared by approach A with or without growth factors showed similar patterns to H1 cells. SOX1, NESTIN and PAX6 are expressed in N2 with growth factors bFGF/EGF. TUJ1 is expressed in N2 without growth factors, but not GFAP. GFAP becomes expressed in N3 after growth factor withdrawal, whereas TUJ1 expression is still visible. Nuclei staining by DAPI is shown in blue. The color code of each gene is indicated in each figure.

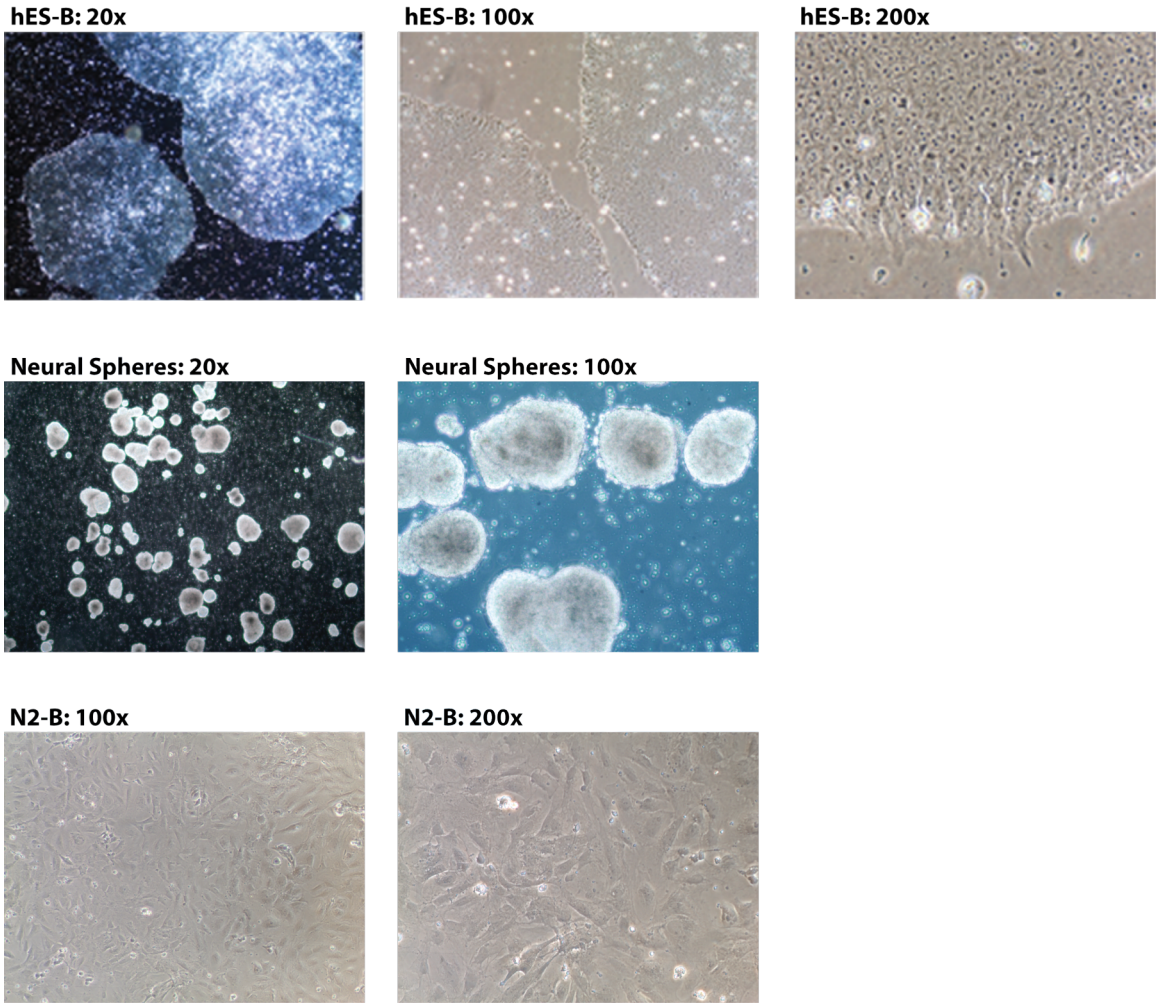

Figure S3. Differentiation of hESCs-B into N2-B. Brightfield images of undifferentiated hESCs (upper panel), the neural spheres (middle panel) and the derived NPCs (lower panel) are shown.

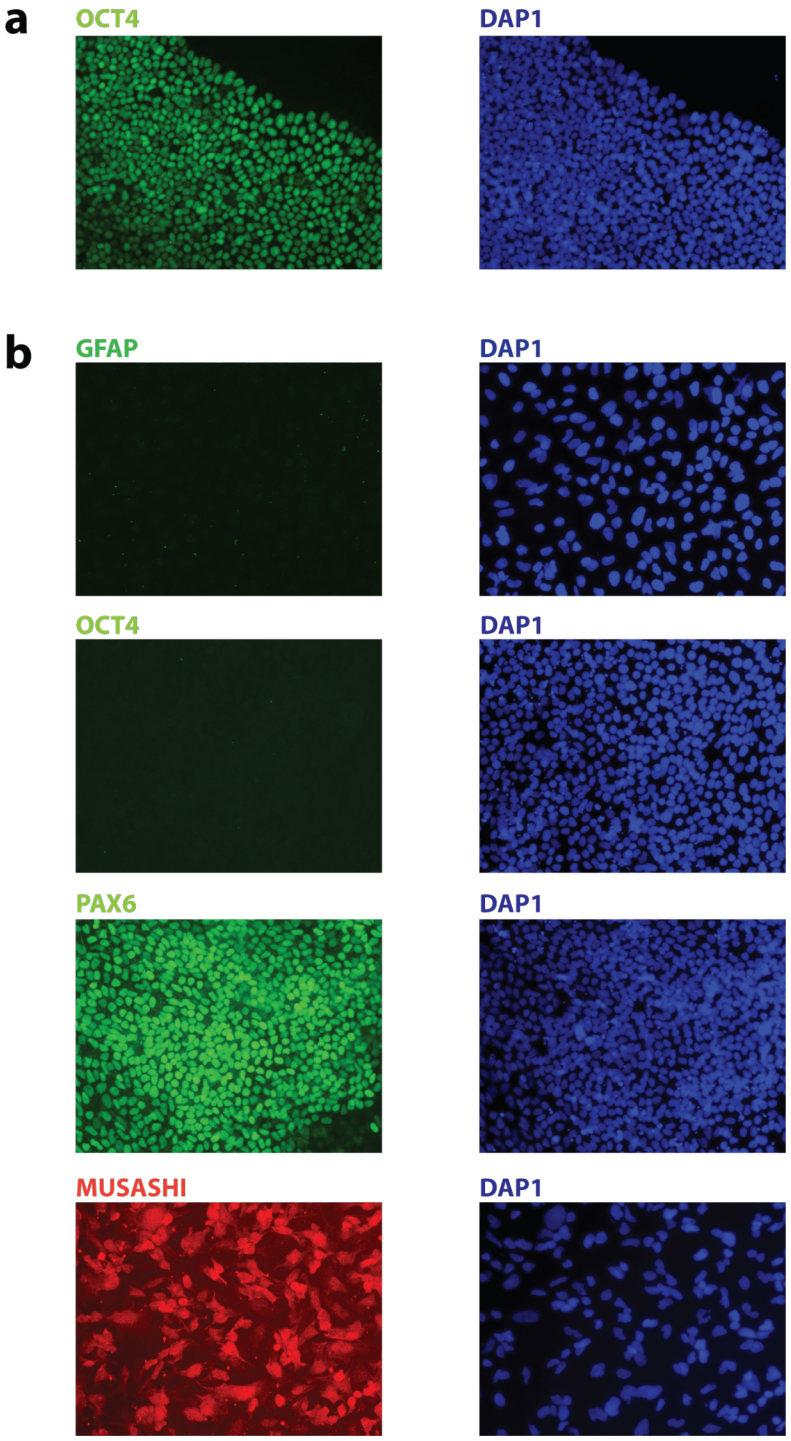

Figure S4. Immunostaining characterization of the differentiation of hESCs-B into N2-B. (a) Immunostaining of the pluripotency marker OCT4 using hESCs-B. Undifferentiated hESCs (passages 50-65) cultured on Matrigel in feeder-free and serum free, component-defined conditions before neural lineage differentiation. Nuclei are labeled by DAPI. **(b)** Immunoflourescence characterization of N2-B using various markers. Nuclei were counterstained by DAPI.

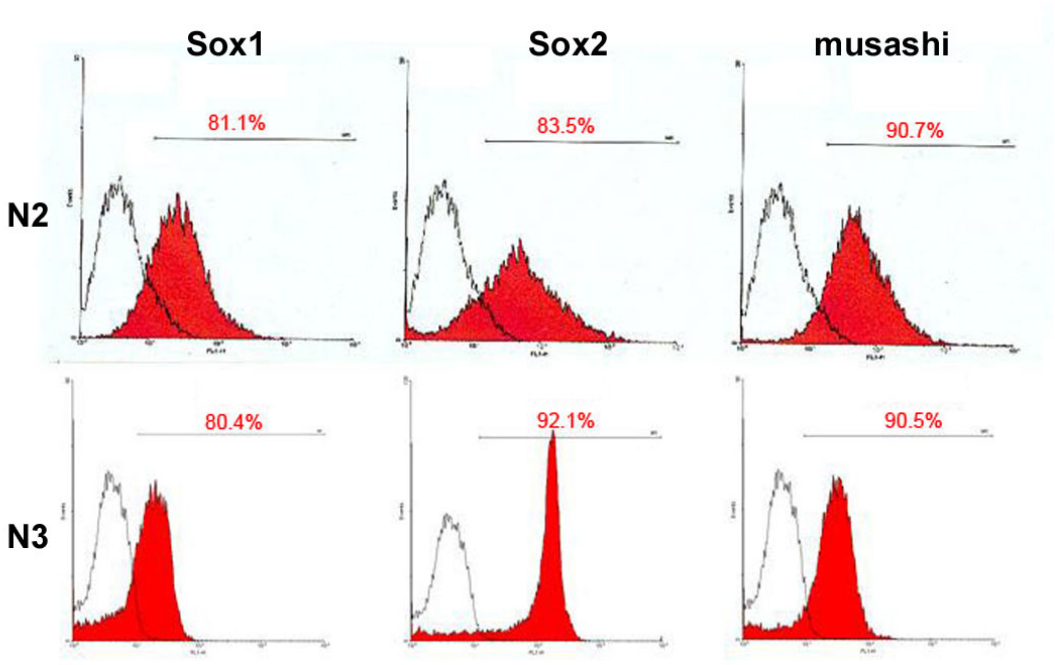

Figure S5. FACS (Fluorescence activated cell sorting) analysis showed that the differentiation cell cultures were highly homogeneous. The positive cells are shown as percentage. Cells stained with corresponding IgM or IgG are used as negative controls. FACS analyses were performed using neural derivatives at N2 and N3 stages from H1 hESCs using neural stem cell markers (SOX1, SOX2 & MUSASHI1). It showed over 80% cells positive for Sox1 and more than 90% for SOX2 and MUSASHI1.

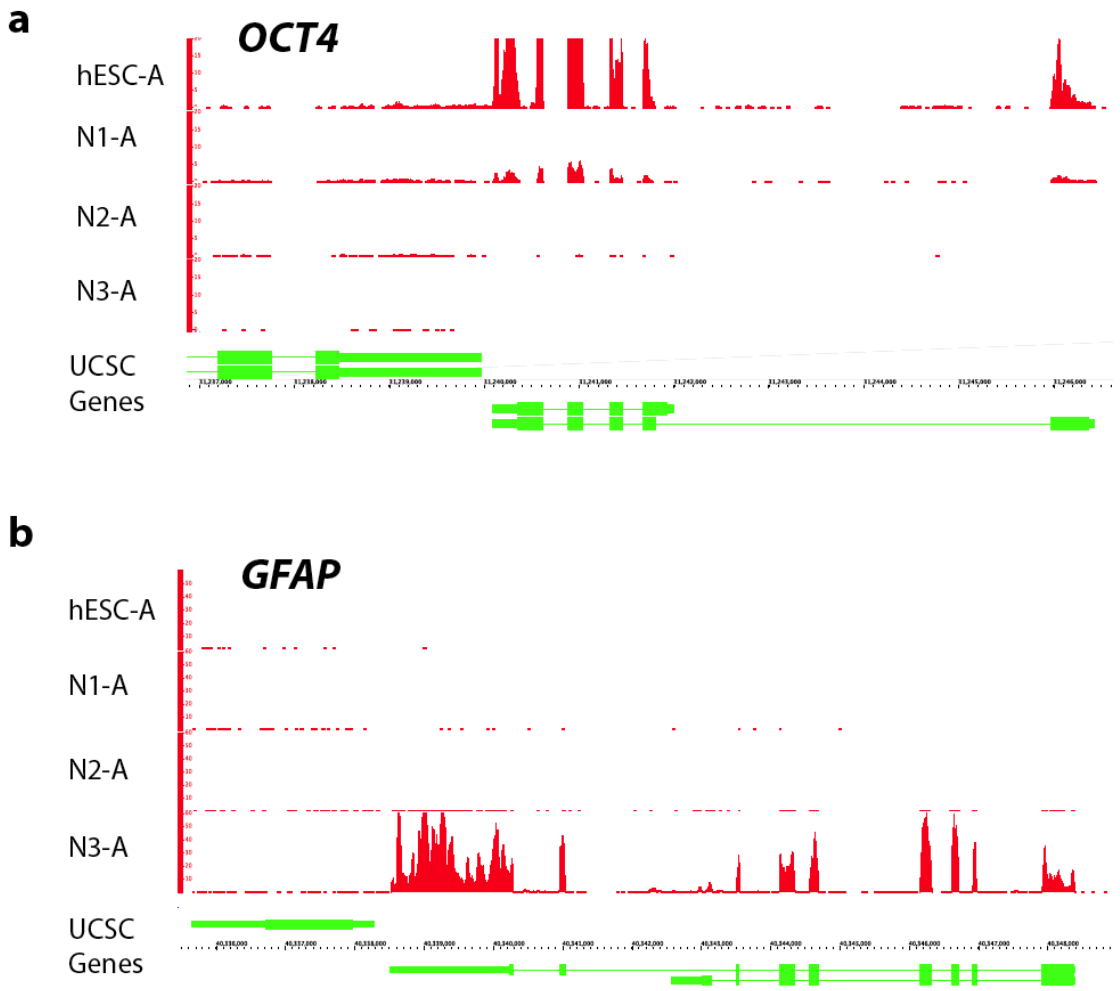

Figure S6. RNA-Seq signal tracks of neural differentiation stage-specific genes. (a) *OCT4 (POU5F1)* is expressed in hESCs, and decreases to very low levels during the neural initiation stage (N1). The y-axis represents the read density normalized by the number of mapped reads per million for each cell type. **(b)** Glial fibrillary acidic protein (*GFAP*) is a marker for N3 stage. All four stages of neural differentiation by approach A are shown in the figure.

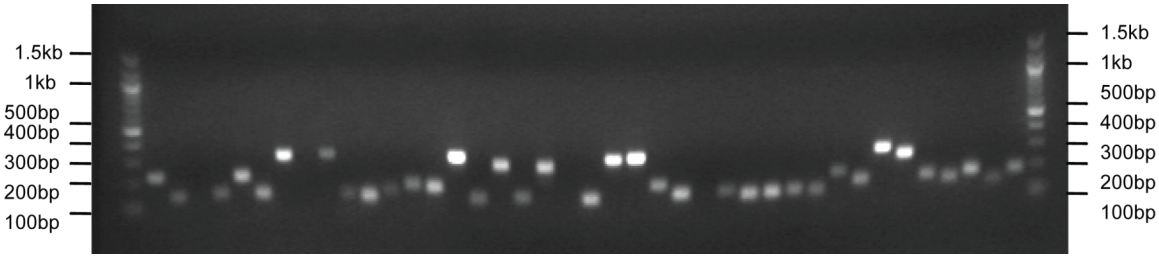

Figure S7. RT-PCR validation of unannotated TARs. From a random sampling of 10 unannotated TARs from each of the four differentiation stages, 90% were validated by RT-PCR. Please see Supplementary Methods for primer information.

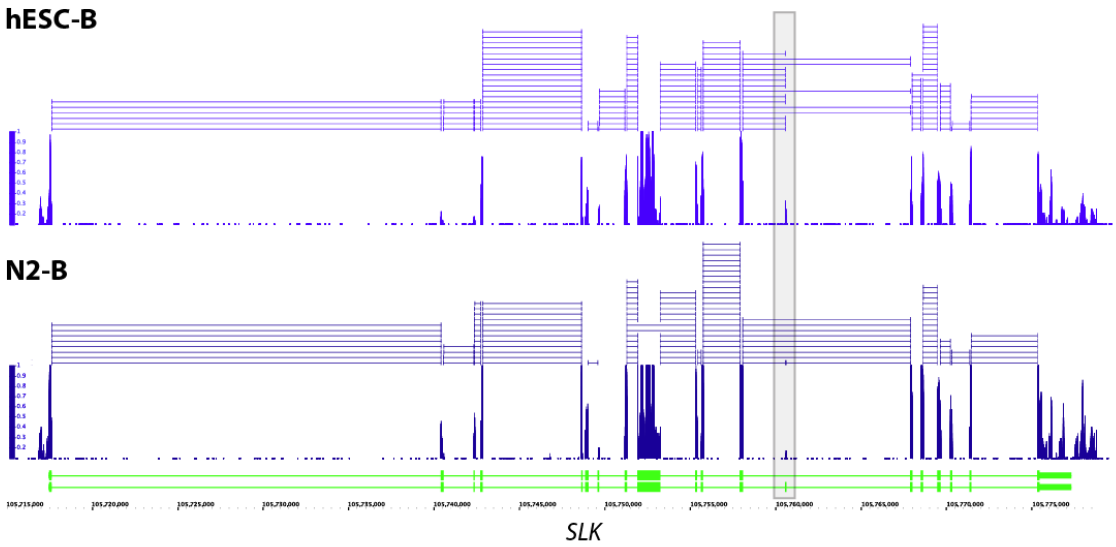

Figure S8. RNA-Seq paired-end reads and signal tracks show that an isoform of serine/threonine kinase 2 (*SLK*) is specifically expressed in hESCs.

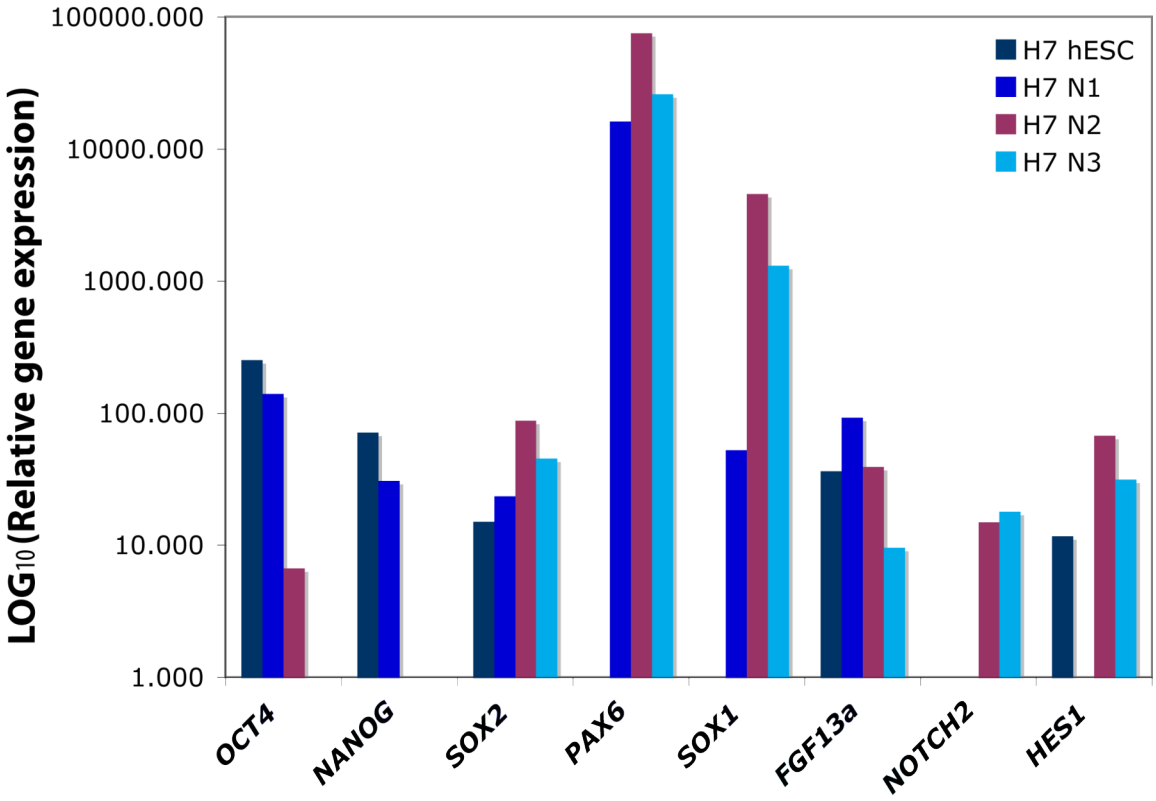

Figure S9. The quantitative changes of gene expression of H7 cells during neural differentiation showed similar patterns to that of H1 cells by qPCR. Y-axis: the relative gene expression level for each stage was normalized using housekeeping gene *HPRT*.

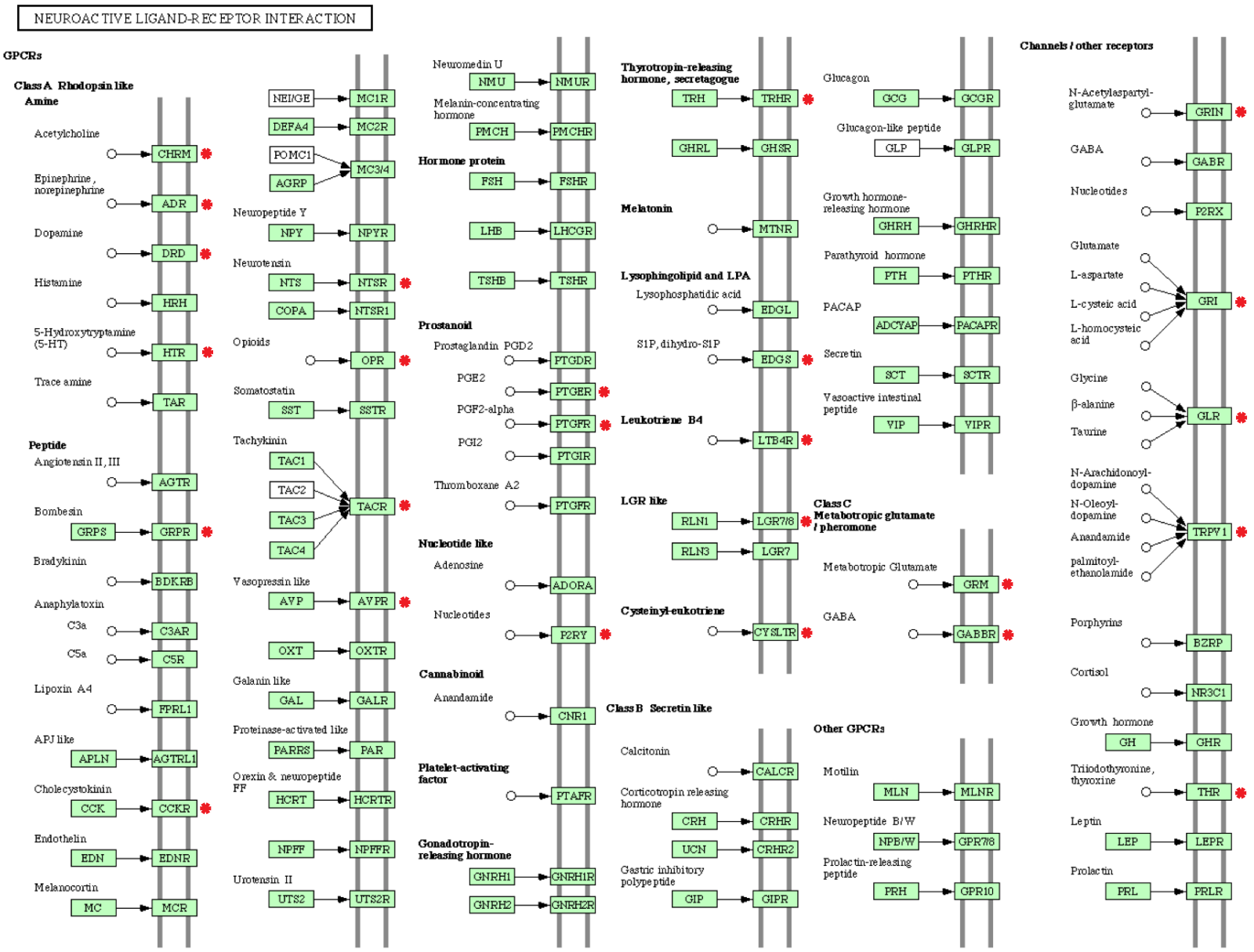

Figure S10. Neuroactive ligand-receptor interaction pathway is enriched among genes that are up-regulated at N1 and/or N2, but down-regulated at N3 (gene expression changes as: "Up(ES→N1)Flat(N1→N2)Down(N2→N3)", "Flat_Up_Down" or "Up_Up_Down" are labeled with *.

SUPPLEMENTARY TABLES:

Approach A				
Number of uniquely mapped reads	hESC-A	N1-A	N2-A	N3-A
35bp Single-End Reads	6,477,731	11,904,041	17,073,257	10,879,155
35bp Paired-End Reads	724,250	594,053	1,285,140	1,044,559
Grand Total	7,201,981	12,498,094	18,358,397	11,923,714
Approach B				
Number of uniquely mapped reads	hESC-B		N2-B	
250-450bp Reads	1,263,516		258,107	
35bp Single-End Reads	60,967,216		33,668,106	
35bp Paired-End Reads	2,046,128		1,655,169	
Grand Total	64,276,860		35,581,382	

Supplementary Table 1. **Summary of sequencing reads for each cell type.**

Stage	Known TAR intersection	Unannotated TAR intersection	Unique known TAR intersection	Unique unannotated TAR intersection
hESC-A	2,320	298	431	35
N1-A	1,904	301	103	31
N2-A	2,147	263	287	24
N3-A	2,093	378	211	90

Supplementary Table 2. **The number of novel, known, and unique TARs that intersected with Bertone TARs**. If a TAR overlaps with the UCSC genes annotation it is categorized as "known". If there is no overlap with the UCSC genes annotation it is classified as "unannotated". If a TAR does not overlap with any TAR in other differentiation stages it is categorized as "unique" to that particular stage.

SUPPLEMENTARY METHODS:

Flow Cytometry. Cells were detached by trypsin, fixed with 4% paraformaldehyde for 15 minutes, permeated with 100% ethanol for 2 minutes and incubated with 10% goat serum (Sigma) for 15 minutes. Cell were then stained with primary antibodies (SOX1 and SOX2 from Abcam, Musashi1 from Chemicon, all 1:100) for 30 minutes on ice followed by secondary antibody (Goat-anti rabbit conjugated fluorescein, Santa Cruz) for 30 minutes. Ten to twenty thousand cells were acquired for each sample using a FACScan (BD Biosciences) and analyzed with CELLQUEST software (BD Biosciences).

Neural differentiation using Approach B. The hESC neural spheres were generated by detaching undifferentiated H1 colonies from the plates with dispase (StemCell Technology) and culturing these hESC clusters for a total of four weeks in the presence of 250 ng/ml Noggin and 20ng/ml bFGF with utilizing extra-low attachment plates (Corning). The medium, based on DMEM/F12 supplemented with 2 mM L-glutamine,100 u/mL penicillin and 100 µg/mL streptomycin, 1X B27 , noggin and bFGF, was changed every 2 days. During the fourth week of culture Noggin was omitted. The neural progenitor cells were then dissociated by Accutase (Millipore) into single cell suspension and plated onto glass slides precoated with polyornithine and fibronectin and grown for a minimum of 24 hours for immunocytochemical characterization. For larger scale studies, the single cell suspension were plated on polyornithine/fibronectin – coated plastic dishes and cultured for 24 hrs, then RNA was harvested. For neuronal differentiation experiments, the neural progenitors were cultured using polyornitine/laminin or polyornithine/fibronectin-coated plastic dishes and passaged by Accutase. To convert the neural progenitors into neurons, the cells were cultured in a neuronal differentiation medium lacking bFGF (DMEM/F12 with B27 and streptomycin/penicillin and supplemented with 10 ng/ml BDNF and 10 ng/ml NT3 for 2-3 weeks.

Immunofluorecent Staining. The cells were either fixed with 4% formaldelhyde or 3% paraformaldehyde for 10 min, followed by standard fluorescent immunocytochemical techniques using the following primary antibodies: monoclonal OCT4 (1:20) and polyclonal SOX2 (1:50) from Santa Cruz, monoclonal NESTIN (1:200), polyclonal MUSASHI (1:100), monoclonal PAX6 (1:50) from Dev.Studies Hybridoma Bank, Iowa, monoclonal TUJ1 (1:1000) and polyclonal GFAP (1: 250) from Chemicon, and polyclonal TUJ1 (1:1000) from Covance. The images were acquired with a Nikon Eclipse E800 fluorescent microscope or LEICA confocal fluorescent microscope.

Construction of Solexa sequencing library

mRNA preparation

mRNA samples were extracted and double polyA purified from cell cultures using Oligotex® Direct mRNA Kits followed by Oligotex® mRNA Kits according to manufacturer's instructions (QIAGEN). 500ng of mRNA was used in each sequencing library**.** mRNA was fragmented using 10X Fragmentation Buffer (Ambion) at 70°C for exactly 5 minutes. 1µl of Stop Buffer was added to terminate the reaction. The fragmented mRNA was purified using RNAclean beads according to the manufacturer's protocol (Agencourt).

cDNA preparation

First-strand cDNA was synthesized using SuperScript II (invitrogen) reverse transcriptase and random primers. The strand of mRNA was removed using RNaseH and double-stranded cDNA was generated with SuperScript II double-stranded cDNA sythesis kit (invitrogen). The reaction was purified using the QIAquick PCR Purification Kit (QIAGEN) and eluted in 34 µl H_2O.

Adapter ligation and size selection

Double-stranded cDNA was end-repaired by T4 DNA polymerase and phosphorylated by T4 polynucleotide kinase. 'A' bases were added to the 3' end of the blunt cDNA fragments by Klenow DNA polymerase, and the cDNA was ligated to the sequencing adaptors (single end, or paired-end adaptors) following the instructions of mRNA-Sequencing Sample Prep Kit (Illumina). The products of the ligation reaction were size-selected on a 2% agarose for a size range of 200bp ±25 bp in the case of single end, or 300bp ±25 bp, 300-600bp and 600-1000bp in the case of paired-end for downstream PCR enrichment. QIAquick Gel Extraction Kit (QIAGEN) was used to purify the gel slices and DNA was eluted in 25 µl of H_2O.

PCR amplification

cDNA was PCR amplified with Phusion DNA Polymerase and a primer set (single end or paired-end primers) that anneals to the ends of the adapters (Illumina). The PCR program was as follows: 98°C for 30 seconds followed by 15 cycles of: 10 seconds at 98°C, 30 seconds at 65°C, 30 seconds at 72°C, and is concluded by 5 minutes at 72°C. The PCR products were size-selected again on a 2% agarose for a size range of 200 bp±25 bp in the case of single end, or 300bp ±25 bp, 300-600bp and 600-1000bp in the case of paired-end. The gel slices were purified with QIAquick Gel Extraction Kit and DNA was eluted in 30 µl of H_2O (QIAGEN). The resulting DNA was quantified using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc.) and sequenced on the Illumina Genome Analyzer, as described (18, 42).

454 sequencing library preparation

mRNA preparation

mRNA samples were extracted and double polyA purified same as described above. 200ng-500ng of mRNA samples were heat fragmented for 2 min at 82°C in 40 mM Tris-acetate, 100 mM potassium acetate, and 31.5 mM magnesium acetate, pH 8.1. The fragmented mRNA was then purified using RNAclean beads according to the manufacturer's protocol (Agencourt). After purification, 1 µl of mRNA was evaluated on an RNA 6000 Pico Lab Chip on the 2100 Bioanalyzer (Agilent) and compared to a non-fragmented sample to confirm fragmentation.

Preparation of single-stranded cDNA

First-strand cDNA was synthesized using 200 units SuperScript II (invitrogen) reverse transcriptase and random primers (N7, 5'-Phosphate). cDNA/mRNA hybrids were melted by incubation in a cDNA Denaturizing Solution (0.5 M NaOH, 0.25 M EDTA) at 65°C for 20 minutes. The reaction is neutralized to a pH between 7.0 and 8.5 using cDNA Neutralization Solution (0.5 M HCl, 0.5 M Tris-HCl). The reaction is then cleaned up using RNAclean beads (Agencourt).

Adapter preparation

Double-stranded adapters were prepared: cDNA Adapter Set A (cDNA Adapter Oligo A: 5'-GCCTCCCTCGCGCCATCAG-3'; cDNA Adapter Oligo A prime: 5'- N*N*N*NNC TGA TGG CGC GAG GGA*G*G*/3ddC) and cDNA Adapter Set B (cDNA Adapter Oligo B: 5'-Bio-GCCTTGCCAGCCCGCTCAGNNNN*N*N*-3'; cDNA Adapter Oligo B prime: 5'-CTGAGCGGGCTGGCAAGG/3ddC (MWG Biotech)). Note: * are phosphorothioated bases, 5'-Bio is a 5' Biotin, and 3ddC is a dideoxynucleotide in 3' of the last C nucleotide.

cDNA Adapter Set A

Combine reagents according to the table below. Mix well by vortexing and then pulse spinning to 3,000g. Place in thin-walled tube on thermocycler and run protocol 80°C - 25°C

Ingredient	Stock Conc.	Qty Req'd	Final Conc.
cDNA Adapter Oligo A	200 µM	25 µl	50 µM
cDNA Adapter Oligo A-prime	200 µM	30 µl	60 µM
1 M Tris-Cl pH 7.5		1 µl	10 µM
Molecular Biology Grade Water		44 µl	
Total		100 µl	

*Program 80°C -25°C = 80 °C 5 min, 65 °C 7 min, 60 °C 7 min, 55 °C 7 min, 50 °C 7 min, 45 °C 7 min, 40 °C 7 min, 35 °C 7 min, 30 °C 7 min, 25 °C 7 min, 4 °C forever.

cDNA Adapter Set B

Combine reagents according to the table below, and proceed in similar fashion as for adapter A.

Ingredient	Stock Conc.	Qty Req'd	Final Conc.
cDNA Adapter Oligo B	500 µM	48 µl	240 µM
cDNA Adapter Oligo B-prime	500 µM	40 µl	100 µM
1 M Tris-Cl pH 7.5		1 µl	10 µM
Molecular Biology Grade Water		11 µl	
Total		100 µl	

Adapter sets were ligated to the single-stranded cDNA using 1 µl of 2000 unit/µl of the Quick T4 DNA Ligase enzyme (New England BioLabs) at 37°C for 2 hours.

Isolation of single-stranded cDNA adapted library

Adapters-ligated single-stranded cDNA library was isolated using Streptavidin Magnetic Beads Sera-Mag 30 beads (Thermo Scientific) and released, then purified using the RNAclean beads (Agencourt). Finally single-stranded cDNA library was quality assessed on an RNA 6000 Pico Lab Chip on the 2100 Bioanalyzer and quantified with the Quant-iT RiboGreen RNA Assay kit according to the manufacturer's guidelines (Invitrogen).

Sequencing

cDNA library was sequenced using the emPCR II Kit (Amplicon A) and on the 454 Genome Sequencer FLX instrument following the manufacturer's instructions. GS FLX Titanium cDNA Libraries were prepared and sequenced at the 454 Life Sciences Sequencing Centre (Branford, CT). Standard 454 software procedures were used to generate nucleotide sequences and quality scores for all reads.

RT-PCR validation experiments. 1μg each of polyA RNAs from cell of ES, N1, N2 and N3 stages was separately set up in 200μl Reverse Transcription (RT) reactions (5ng/μl). RT reactions were performed using SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen, CA, USA) that contains both oligo$(dT)_{20}$ and random hexamers. In parallel, reactions without reverse transcriptase (RTase minus) were also performed as the negative control for genomic contamination.

PCR primers were designed using Primer3 or BatchPrimer3. RT was followed by PCR amplification using Advantage™ 2 PCR Enzyme System (Clontech, CA, USA). 1μl RT reaction and 1μl RTase minus negative control from the above were used in 25μl PCR reactions. The PCR program for unannotated singleton TARs was 95°C for 30 seconds, followed by 35 cycles of 95°C for 15 seconds, 68°C 30 seconds, and concluded by an extension cycle of 72°C for 1 minute. The PCR program for multi-exonic unannotated transcripts connected via a group of paired-end reads was: 95°C for 1 minute, followed by either 28 or 35 cycles of 95°C for 15 seconds, 68°C 3 minutes, and concluded by an extension cycle of 72°C for 5 minutes. The PCR products were visualized on a 1% agarose gel.

Real-time quantitative RT-PCR. Total RNA was purified with TRI reagent (Sigma) and trace contaminated DNA was removed by DNase treatment (Invitrogen). First-strand cDNA was synthesized from 2 μg total RNA in a 20-μl volume using oligo-dT_{15} primer and SuperScriptII (Invitrogen). The PCR reaction consisted of 2μl of 1:10-diluted cDNA, 15 μl of SYBR green-Taq mixed solution (Sigma) and 9 pmol each of 5' and 3' primers (see table below) in a total volume of 30 μl and was performed in a Opticon thermal cycler (Biorad) for 40 cycle with denaturation at 95°C for 15 second, annealing at 60°C for 30 second and extension at 72°C for 30 second. RNA without reverse transcriptase treatment was used as negative control.

Mapping sequence reads to the human genome. The 454 250-450bp long reads were mapped to human genome (hg18) using BLAT(3) with default parameters. Reads were removed in a subsequent post-processing step if less than eighty percent of the read mapped to the genome. A three-step approach was adopted to map the short single-end reads to the genome. First, reads were aligned to the human genome (hg18) with Bowtie allowing up to two mismatches(4). Only reads that mapped to a unique location in the genome were retained. In a second step, the remaining reads were aligned to a splice junction library consisting of all possible unique pair-wise splice junctions within each transcript of the AceView annotation set(5). This alignment step was also performed with Bowtie allowing only unique alignments with up to two mismatches. Lastly, the reads that did not align in the previous two steps were aligned to the genome using less stringent parameters. In this step reads were allowed to map up to five genomic locations. One of these locations was selected according to the read density of the uniquely mapped reads contained within non-overlapping bins (50bp) across the genome. The short paired-end reads were mapped to the genome using ELAND, which is a component of the Illumina software pipeline, operating in the eland_pair mode. Each end was aligned separately and then the best-matched pair was selected and reported.

Annotation sets and composite gene models. Initially, the AceView annotation set, consisting of 258,618 transcripts, was used to create a splice junction library in order to map the short single-end reads. However, for many subsequent analyses the UCSC Genes annotation set, comprised of 66,803 transcripts, was utilized because it contains information about the various splice isoforms (6, 7). The various transcript isoforms of a particular gene were merged into a composite gene model by taking the union of all the exons from the various transcript isoforms.

Number of genes detected as a function of read coverage. To assess the number of genes detected as a function of read depth the mapped reads were intersected with the composite gene models of the UCSC Genes annotation set (7). Reads were sampled randomly at various intervals of five millions and the fraction of genes detected was calculated. The fraction of genes detected was determined at two-fold and five-fold coverage. The coverage is defined as the number of nucleotides obtained form all the reads that overlap with a composite gene model divided by the length of the composite gene model.

Quantification of gene transcription. The level of gene transcription was quantified by intersecting the mapped reads with the composite gene models of the UCSC Genes annotation set. The transcription values were determined by summing the nucleotide overlaps from all the reads that intersect with a composite gene model divided by the length of the composite gene model and the number of mapped reads in millions.

Differential gene expression. Gene expression values from various differentiation stages were compared to assess differential gene expression. Genes with at least a two-fold change in their expression values were referred to as differentially expressed. In order to capture the global changes in gene expression across the four differentiation stages each gene was assigned to one of 27 gene expression patterns. Between any two

differentiation stages the change in gene expression was assigned to one of three categories: up, down, or flat. The 'up' or 'down' categories represent at least a two-fold change in gene expression between the two differentiation stages while the 'flat' category indicates an unchanged gene expression or a change less than two fold. After assigning each gene to one of the 27 gene expression patterns, the logarithm of the expression values across the four stages were plotted for each pattern type.

Splice junction coverage. The splice junction coverage was determined by counting the number of known and unannotated splice junctions obtained from random samples of short single-end reads spanning two exons. Since the splice junction library consisted of all possible unique pair-wise splice junctions within a transcript each splice junction can be categorized as either known or unannotated. Known splice junctions are defined as the junctions that are consistent with annotated transcripts while unannotated splice junctions refer to skipped exons.

TAR analyses and connecting TARs using paired-end reads. In order to discover unannotated transcriptionally active regions (TARs) the signal track of the mapped reads was segmented using the common maxGap/minRun algorithm (maxGap = 10, minRun = 50, threshold = 2). The set of unannotated TARs identified for each differentiation stage were intersected with the set of TARs reported by Bertone et al. If two TARs from the two different sets overlapped by at least one nucleotide they were counted as overlapping.

Splice junction diversity. A subset of genes was selected to analyze the splice junction diversity across the four differentiation stages. Because the sequence coverage for low abundant transcript is lower and it would be difficult to ascertain splicing diversity, the 500 most highly transcribed genes were selected based on the sum of their transcription values in the four stages. For this subset of genes the junction diversity per composite gene model was calculated. The junction diversity is defined as the number of unique splice junctions detected in the composite gene model given all the mapped splice junction reads. In order to facilitate a comparison between the various developmental stages the junction diversity values were normalized for the number of mapped reads per million. In addition, the splice junction diversity values were normalized for the number of annotated splice junctions in the composite gene model. In the next step normalized junction diversity values were clustered using k-means clustering (k = 3). Lastly, the normalized junction diversity values and the associated gene transcription values were plotted.

Enriched Gene Ontology (GO) categories and pathway analysis. An internal software tool was employed to determine statistically significant over-represented GO categories within lists of genes. The hypergeometric distribution was utilized to calculate p-values. These p-values were then corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure. Enriched KEGG pathways of differentially expressed genes were identified using **D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery (**DAVID**) 2008 (8).

References:

1. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5: 621-628.
2. Morin, R. D., O'Connor, M. D., Griffith, M., Kuchenbauer, F., Delaney, A., et al. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res* 18: 610-621.
3. Kent, W. J. (2002) BLAT--the BLAST-like alignment tool. *Genome Res* 12: 656-664.
4. Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10: R25.
5. Thierry-Mieg, D., Thierry-Mieg, J. (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* 7 Suppl 1: S12 11-14.
6. Karolchik, D., Kuhn, R. M., Baertsch, R., Barber, G. P., Clawson, H., et al. (2008) The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res* 36: D773-779.
7. Hsu, F., Kent, W. J., Clawson, H., Kuhn, R. M., Diekhans, M., et al. (2006) The UCSC Known Genes. *Bioinformatics* 22: 1036-1046.
8. Huang da, W., Sherman, B. T., Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57.

Real-time PCR primers:

Gene name	Forward	Reverse
HPRT	TCCTTGGTCAGGCAGTATAATCC	GTCAAGGGCATATCCTACAACAAA
PAX6	TCCGTTGGAACTGATGGAGT	GTTGGTATCCGGGGACTTC
SOX1	AACACTTGAAGCCCAGATGGA	GCAGGCTGAATTCGGTTCTC
DCX	AGACCGGGGTTGTCAAAAA	TCAGGACCACAGGCAATAAA
ZIC1	AAGATCCACAAAAGGACGCA	CACGTGCATGTGCTTCTTG
MAP2	TTCGTTGTGTCGTGTTCTCA	AACCGAGGAAGCATTGATTG
Olig2	CCTGAGGCTTTTCGGAGC	GATAGTCGTCGCAGCTTTCG
Notch2	GCCTTCGCCTCCTGTACTC	GCCCATTTAGGGGGTTGGT
Hes1	CGGACATTCTGGAAATGACA	TACTTCCCCAGCACACTTGG
FGF2	AGCCAGGTAACGGTTAGCAC	GGAGAAGAGCGACCCTCAC
FGFR1	GGAAGGACTCCACTTCCACA	GTCACAGCCACACTCTGCAC
FGF11	AGTTTGGTGACGATGCCTTT	AGAAGCAGCTCCTCATCCTG
FGF12	TGTAGTGGCCATCCAAGGAG	TTCCTTGAATTTGCATTCTGG
FGF13a	AGCTCGCTCATCCGTCAGA	TTGGAGCCGAAGAGTTTGACC
FGF13b	CTATACAGCCGACAAGGCTACC	ACCCACAGGGATGAGGTTAAA

RT PCR primers:

NCAM1 ForwardPrimer1	GAGAACCAGCAAGGAAAATCCAAGG
NCAM1 ReversePrimer1	CTCTTCCACACAGAATGGTTGCTCA
NCAM1 ReversePrimer2	TGTGAGTGGTGACATGTCTGGAGGT

Singleton TAR primers

ES_S1F	CCTCCTTGCCAAATGTGCTAAA
ES_S1R	ACCACCACAGCACCCAAAGA
ES_S2F	TGCCCGGCCTCATTTATATT
ES_S2R	AGACATGGTACACTGCACAGCA
ES_S3F	TCACAGTGGAGGCAAGGAACA
ES_S3R	GCAACCCTGAGACACTTTACAGTCC
ES_S4F	CCAGCAACAGATCCCCATCC
ES_S4R	TGCTTCAGATTTTATGCCCTGAGA
ES_S5F	GCACTGCCTAGGTGTTGGAAA
ES_S5R	AAGATTCCCTCTTGGCCTCCA
ES_S6F	TGTGCAGCTTTGGTGCCTGT
ES_S6R	GCAATGAGGAGGGATTCCTGT
ES_S7F	AACCCTGGGGGAAAATGGTC
ES_S7R	TCCCCTGCAGAGATGGAGAAA
ES_S8F	CCCCAAGTGCACTAACTCCTCAA
ES_S8R	TCACCACCTACCAGAACAGAGCA
ES_S9F	GCTTCTGCACAGCAAAGAAAGC
ES_S9R	TGTGGCTTTGACATGCGTTTT
ES_S10F	CTGCAGCATTCCGAGGACAG
ES_S10R	TCCTCAGTTAAGCCTTCCCACCT
N1_S16F	TTACAGTTCCAGAAGTGCGTCGAA
N1_S16R	TGGGGTAGTGAAATTAAAGCCCATA
N1_S17F	GAGAGCTTAGCACCCATCTTGC
N1_S17R	GGCTTTGGGGAAGAGTTTATATGAGG
N1_S18F	CTCCCACTGCTCGGCACCTA
N1_S18R	GCTGTGAATGAGGAATGCTAACTCC
N1_S19F	CGCCACAGAAATTGCCGAAT
N1_S19R	CCCCCTGCTGATTTTAATGAATC
N1_S20F	TGGACAGGTATCGGAATCTAGGACA
N1_S20R	GCAGAGAAAGGGTGCTGGATA
N1_S21F	GCGCACCCCACGTTCTCTTA
N1_S21R	TGAGGGAGTGGTAAGGGGAAAA
N1_S22F	CAGCCGCCTTAATCCCTCCT
N1_S22R	CCCAGGGTCATAGTGCCTTCC
N1_S26F	AAAGGAAGTGACAACTGAAACTAATGG
N1_S26R	TTGCTGGTTAACCTGTACAACATTCC
N1_S28F	GGGGCTGGAAGAGAGGGAGT
N1_S28R	GGAAAGCATGCCATGGGGTA

N1_S29F	TCCCATTTTCTCTTATGATTTTATCCA
N1_S29R	TGCCTTTGTTGAAATGAAAAATGT
N2_S31F	CACGGGCCTCAGGCAAAAAT
N2_S31R	TGCCTACCTTTGGGGGTGCT
N2_S32F	GAGAACAGGGGCAGGGGATT
N2_S32R	CCGGTCCCAGGAAGGTGAAT
N2_S33F	ATCCGAACTGCCTGGGGACT
N2_S33R	TTGACCAAGCCTGGATCACAT
N2_S34F	GAACGTGTGTGATGTGCTGGTG
N2_S34R	GAGGGTTTGATACCAATAGGGGAAA
N2_S35F	TTGCTACCTGATGGCCATTCTT
N2_S35R	TGCTTCCTTCTGGGCTGTTTG
N2_S36F	TTTTTGAGCAAACCCAACCA
N2_S36R	TTCAGGTAGTGGCCCTGTTCA
N2_S37F	CCCCGCGAAGAACATGTAGA
N2_S37R	TTGCATGATTCGAATTGCAGGT
N2_S38F	GCCAAAAATGTGCAAGGCATC
N2_S38R	GCACAGGGTGGTAGCTGACAGA
N2_S39F	CAGGGAGAGAGCCGGAAGAA
N2_S39R	CCCCCGCAGTTTCATTCTCA
N2_S40F	TGCTGGTTTTCAAACAAGAAAAGACC
N2_S40R	TCCTGCTGTCAAGGAGGGTACA
N3_S46F	TCCCAAAAGGCAAAGCTGGT
N3_S46R	AGGGAGGAAGGGAGGGAAGG
N3_S47F	CCCCTGGTCGTGTAGCAGTG
N3_S47R	GCGACCTGCGACTCCTTGAC
N3_S48F	TGCCAGAGACAATCTTTGGTGTTC
N3_S48R	GGCCTGTGCCATCAATATCTTA
N3_S49F	TGCGCAGGGGTTTTAGTTAAATG
N3_S49R	GGGTCGAAGGTGGATTTAGCA
N3_S50F	TGCCGGTTTCTATTGACTTGG
N3_S50R	ACACCTCCACGGGAGACAGC
N3_S51F	AAATGTTGGGGTGGGTGCTG
N3_S51R	TCCAATTCCACATCTCTCCCTCTC
N3_S52F	TGCTTGCTTTTTCCATCTGAACC
N3_S52R	GCCCGGAGAAAGGGTACGAA
N3_S53F	CGCCGGTTTCTATTGACTTGG
N3_S53R	CAGCAACGCAGCTCAAAACG
N3_S54F	ACTGCCCTGGGTTTCCTCCT
N3_S54R	TGGCATCACAAGCGTCACAA
N3_S55F	TCCATGTCTTCTTGCCTTTCCTTC
N3_S55R	TCTGCCTGGAATGACTTCTCTGA

Novel transcript primers

Left primer	
transcript1A	AAGACTGTGGAACAGGTGGAGTGTC
transcript2A	TGAGGACAAGTCAACGAGAGATTCC
transcript3A	CAAATGCAAAAACTCAGGGCAGTTA
transcript4A	GTAATTAACCGGTGCTGCTTGGAC
transcript5A	TCCAATGGCAAGTTTGTGACATCTA
transcript6A	TCTCATACCAGGAAACGAGCTTGAC
Right primer	
transcript1B	GCATCTGCTAGCAATTCACTTTGTG
transcript2B	AGTAACCTGGCCCATTGTCTGTTTT
transcript3B	ACATCAGGACATTGCCTTTGTCTGT
transcript4B	GTCTGGTTCCTGTACTCGGTGAAGA
transcript5B	CTCGAAGAAGAAGTCCACAGGAGTG
transcript6B	GTCCAAGGTCATCCATGACAACTCT

SUPPLEMETARY DATA:

File containing the expression values for all genes across the four differentiation stages (hESC-A, N1-A, N2-A, and N3-A): ExpressionValues.xls