IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

FILED VIA ELECTRONIC FILING SYSTEM
COMMISSIONER FOR PATENTS

PROVISIONAL APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 C.F.R. § 1.53(c).

TITLE: **GENE EXPRESSION ANALYSIS IN SINGLE CELLS**

Inventor:

Last	First	MI	City, State or Foreign Country and City
Linnarsson	Sten		Stockholm, Sweden

☒ Submitted herewith are 22 pages of specification, 4 pages of claims, and 7 sheets of drawings.

FEE CALCULATION

								Rate		Fee
Basic Filing Fee										$110.00
Total number of pages (specification & drawings)	33	-	No. pages included in basic fee	100	=	No. of extra pages, divided by 50, rounded up to nearest whole number	0	x $135.00	=	$ 0.00
TOTAL FILING FEE										**$110.00**

☒ Small entity status is claimed for this application.

☒ $110 is being submitted herewith via EFS. The Director is hereby authorized to charge any additional fees which may be required in connection with the filing of this provisional application or credit over-payment, to Account No. 02-4550.

☒ Address all telephone calls to Tanya M. Harding, Ph.D. at telephone number (503) 595-5300. Address all correspondence to the address associated with
CUSTOMER NUMBER 24197

Respectfully submitted,

KLARQUIST SPARKMAN, LLP

One World Trade Center, Suite 1600
121 S.W. Salmon Street
Portland, Oregon 97204
Telephone: (503) 595-5300
Facsimile: (503) 595-5301

By /Tanya M. Harding/
Tanya M. Harding, Ph.D.
Registration No. 42,630

PARSE EX1015.001

FILED VIA EFS
DATE OF SUBMISSION: March 30, 2009
Document Number: 6877121

GENE EXPRESSION ANALYSIS IN SINGLE CELLS

The present invention relates to the analysis of gene expression in single cells. In particular, the invention relates to a method for preparing a cDNA library from a plurality of single cells, and to a cDNA library produced by this method. The cDNA libraries prepared by the method of the invention are suitable for analysis of gene expression by sequencing.

The determination of the mRNA content of a cell or tissue (i.e. "gene expression profiling") provides a method for the functional analysis of normal and diseased tissues and organs. For example, gene expression profiling can be used in the study of embryogenesis; for the characterisation of primary tumour samples; for the analysis of biopsies from diseased and normal tissue in, for example, psoriasis; for the comparative analysis of cell types from different species to delineate the evolution of development; as an assay system for diagnostics; as a quality control system in cell replacement therapy (i.e. to ensure that a culture of cells is sufficiently pure, and the cells are correctly differentiated); and as an *in vitro* tool to measure the effect of a transfected gene or siRNA on downstream targets in spite of less than 100% transfection efficiency.

Gene expression profiling is usually performed by isolating mRNA from tissue samples and subjecting this mRNA to microarray hybridisation. However, such methods only allow previously known genes to be analysed, and cannot be used to analyse alternative splicing, promoters and polyadenylation signals.

Therefore, direct sequencing of the all, or parts, of the mRNA content of a tissue is being increasingly used (Cloonan, Forrest et al. 2008). However, current methods of analysing the mRNA content of cells by direct sequencing rely on analysing bulk mRNA obtained from tissue samples typically containing millions of cells. This means that much of the functional information present in single cells is lost or blurred when gene expression is analysed in bulk

PARSE EX1015.002

mRNA. In addition, dynamic processes, such as the cell cycle, cannot be observed in population averages. Similarly, distinct cell types in a complex tissue (e.g. the brain) can only be studied if cells are analysed individually.

Gene expression in single cells has previously been analysed using a variety of methods (see, for example, Brail, Jang et al. 1999; Levsky, Shenoy et al. 2002; Bengtsson, Stahlberg et al. 2005; Esumi, Kakazu et al. 2005). In particular, single cell gene expression in neural cells has been studied by microarray analysis (see (Esumi, Wu et al. 2008)). However, these methods require that each single cell is analysed individually and treated separately during the entire procedure, which is time-consuming and expensive. In addition, the preparation and amplification of samples from single cells independently potentially introduces cell-to-cell variation. Finally, as the cDNA of each cell must be amplified to an amount that can be reasonably handled for the subsequent analysis, there is potential amplification bias. For example, a single cell contains about 0.3 pg of mRNA, and at least a 300 ng is commonly needed for subsequent analysis by sequencing. Therefore, an amplification of at least a million-fold is required.

The present invention aims to overcome, or reduce, these problems by providing a method of preparing cDNA libraries which can be used to analyse gene expression in single cells.

Therefore, according to the invention there is provided a method of preparing a cDNA library from a plurality of single cells, the method comprising:

(i) releasing mRNA from each single cell to provide a plurality of individual samples, wherein the mRNA in each individual mRNA sample is from a single cell;

(ii) synthesising a first strand of cDNA from the mRNA in each individual mRNA sample and incorporating a tag into the cDNA to provide a plurality of tagged cDNA samples, wherein the cDNA in each tagged cDNA sample is complementary to mRNA from a single cell;

PARSE EX1015.003

(iii) pooling the tagged cDNA samples; and
(iv) amplifying the pooled cDNA samples to generate a cDNA library comprising double-stranded cDNA.

As each cDNA sample obtained from a single cell is tagged, gene expression can be analysed at the level of a single cell. This allows dynamic processes, such as the cell cycle, to be studied and distinct cell types in a complex tissue (e.g. the brain) to be analysed. However, pooling the samples simplifies handling of the samples from each single cell and reduces the time required to analyse gene expression in the single cells, thus allowing high throughput analysis of gene expression. As the samples are pooled before amplification, technical variation between samples is virtually eliminated. In addition, as the cDNA samples are pooled before amplification, less amplification is required to generate sufficient amounts of cDNA for subsequent analysis compared to if cDNA samples from each single cell were amplified and treated separately. This reduces amplification bias, and also means that any bias will be similar across all the cells used to provide pooled cDNA samples. In addition, RNA purification, storage and handling are not required, which helps to eliminates problems caused by the unstable nature of RNA.

As the cDNA libraries produced by the method of the invention are suitable for the analysis the gene expression profiles of single cells by direct sequencing, it is possible to use these libraries to study the expression of genes which were not previously known, and also to analyse alternative splicing, promoters and polyadenylation signals.

Single cells may, for example, be obtained from a tissue of interest, or from a biopsy, blood sample, or cell culture, and the method of preparing the cDNA library may include the step of obtaining single cells. A single cell suspension may be obtained using standard methods, e.g. enzymatically using trypsin or papain, or mechanically. Single cells may be placed in any suitable reaction vessel in which single cells may be treated individually. For

PARSE EX1015.004

example a 96-well plate, such that each single cell is placed in a single well.

Methods for manipulating single cells are known in the art and include fluorescence activated cell sorting (FACS), micromanipulation and the use of semi-automated cell pickers (e.g. the Quixell™ cell transfer system from Stoelting Co.). Individual cells may, for example, be individually selected based on features detectable by microscopic observation, such as location, morphology, or reporter gene expression.

mRNA may be released from the cells by lysing the cells. Lysis may be achieved by, for example, heating the cells, or by the use of detergents or other chemical methods, or by a combination of these. However, any suitable lysis method known in the art may be used. A mild lysis procedure may advantageously be used to prevent the release of nuclear chromatin, thereby avoiding genomic contamination of the cDNA library, and to minimise degradation of mRNA. For example, heating the cells at 72°C for 2 minutes in the presence of Tween-20 is sufficient to lyse the cells while resulting in no detectable genomic contamination from nuclear chromatin. Altrnatively, cells may be heated to 65°C for 10 minutes in water (Esumi *et al.* 2008); or 70°C for 90 seconds in PCR buffer II (Applied Biosystems) supplemented with 0.5% NP-40 (Kurimoto *et al.* 2006); or lysis may be achieved with a protease such as Proteinase K or by the use of chaotropic salts such as guanidine isothiocyanate (US patent application no. 11/783,575).

Synthesis of cDNA from mRNA in step (ii) of the above method may be performed directly on cell lysates, such that a reaction mix for reverse transcription is added directly to cell lysates. Alternatively, mRNA may be purified after its release from cells. This may help to reduce mitochondrial and ribosomal contamination. mRNA purification may be achieved by any method known in the art, for example by binding the mRNA to a solid phase. Commonly used purification methods include paramagnetic beads (e.g. Dynabeads).

PARSE EX1015.005

Alternatively, specific contaminants, such as ribosomal RNA can be selectively removed using affinity purification.

cDNA is typically synthesised from mRNA by reverse transcription. Methods for synthesising cDNA from small amounts of mRNA, including from single cells, have previously been described (Kurimoto, Yabuta et al. 2006; Kurimoto, Yabuta et al. 2007; Esumi, Wu et al. 2008). In order to generate an amplifiable cDNA, these methods introduce a primer annealing sequence at both ends of each cDNA molecule in such a way that the cDNA library can be amplified using a single primer. The Kurimoto method uses a polymerase to add a 3′ poly-A tail to the cDNA strand, which can then be amplified using a universal oligo-T primer. In contrast, the Esumi method uses a template switching method to introduce an arbitrary sequence at the 3′ end of the cDNA, which is designed to be reverse complementary to the 3′ tail of the cDNA synthesis primer. Again, the cDNA library can be amplified by a single PCR primer. Single-primer PCR exploits the PCR suppression effect to reduce the amplification of short contaminating amplicons and primer-dimers (Dai, Zhu et al. 2007). As the two ends of each amplicon are complementary, short amplicons will form stable hairpins, which are poor templates for PCR. This reduces the amount of truncated cDNA and improves the yield of longer cDNA molecules.

Thus, the synthesis of the first strand of the cDNA in step (ii) may be directed by a cDNA synthesis primer (CDS) that includes an RNA complementary sequence (RCS) that is at least partially complementary to some mRNA in an individual mRNA sample. This allows the primer, which is typically an oligonucleotide, to hybridise to at least some mRNA in an individual mRNA sample to direct cDNA synthesis using the mRNA as a template. The RCS may comprise oligo(dT), or be gene family-specific, or may be composed of a random sequence, such as random hexamers. To avoid the CDS priming on itself and thus generating undesired side products, a non-self-complementary semi-random sequence may be used. For example, one letter of the genetic code may be excluded, or a more complex design may be used while restricting the CDS to be non-self-complementary.

PARSE EX1015.006

The RCS may also be at least partially complementary to a portion of the first strand of cDNA, such that it is able to direct the synthesis of a second strand of cDNA using the first strand of the cDNA as a template. Thus, following first strand synthesis, an RNase enzyme (e.g. an enzyme having RNaseH activity) can be added after synthesis of the first strand of cDNA to degrade the RNA strand and to permit the CDS to anneal again on the first strand to direct the synthesis of a second strand of cDNA. For example, the RCS could comprise random hexamers, or a non-self complementary semi-random sequence (which minimises self-annealing of the CDS).

A template-switching oligonucleotide (TSO) that includes a portion which is at least partially complementary to a portion of the 3' end of the first strand of cDNA may be added to each individual mRNA sample in step (ii) of the above method. Such a template switching method is described in (Esumi, Wu et al. 2008) and allows full length cDNA comprising the complete 5' end of the mRNA to be synthesised. As the terminal transferase activity of reverse transcriptase typically causes 2-5 cytosines to be incorporated at the 3' end of the first strand of cDNA synthesised from mRNA, the first strand of cDNA may include a plurality of cytosines, or cytosine analogues that base pair with guanosine, at its 3' end (see US 5,962,272). The first strand of cDNA may include at least 2, at least 3, at least 4, at least 5 or 2, 3, 4, or 5 cytosines, or cytosine analogues that base pair with guanosine, at its 3' end. The first strand of cDNA may include 2-5 cytosines, or cytosine analogues that base pair with guanosine, at its 3' end, such as 5-aminoallyl-2'-deoxycytidine. Therefore, the TSO may include a 3' portion comprising a plurality of guanosines or guanosine analogues that base pair with cytosine, such as deoxyriboguanosine, riboguanosine, locked nucleic acid-guanosine, and peptide nucleic acid-guanosine. The guanosines may be ribonucleosides or locked nucleic acid monomers. The TSO may include a 3' portion including at least 2, at least 3, at least 4, at least 5, or 2, 3, 4, or 5, or 2-5 guanosines, or guanosine analogues that base pair with cytosine. The presence of a plurality of guanosines (or guanosine analogues

PARSE EX1015.007

that base pair with cytosine) allows the TSO to anneal transiently to the exposed cytosines at the 3' end of the first strand of cDNA. This causes the reverse transcriptase to switch template and continue to synthesis a strand complementary to the TSO.

The 3' end of the TSO may be blocked, for example by a 3' phosphate group, to prevent the TSO from functioning as a primer during cDNA synthesis.

If the mRNA is released from the cells by cell lysis, and if this is achieved partially by heating, then the CDS and/or the TSO may be added to each individual mRNA sample during cell lysis (i.e. during step (i)), as this will aid hybridisation of the oligonucleotides. Reverse transcriptase, however, may be added after cell lysis to avoid denaturation of the enzyme.

The tag may be incorporated into the cDNA during its synthesis. For example, the CDS and/or the TSO may include a tag, such as a particular nucleotide sequence, which may be at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15 or at least 20 nucleotides in length. For example, the tag may be a nucleotide sequence of 4-20 nucleotides in length, e.g. 4, 5, 6, 7, 8, 9, 10, 15 or 20 nucleotides in length. As the tag is present in the CDS and/or the TSO it will be incorporated into the cDNA during its synthesis and can therefore act as a "barcode" to identify the cDNA. Both the CDS and the TSO may include a tag. The CDS and the TSO may each include a different tag, such that the tagged cDNA sample comprises a combination of tags. Each cDNA sample generated by the above method may have a distinct tag, or a distinct combination of tags, such that once the tagged cDNA samples have been pooled, the tag can be used to identify which single cell from each cDNA sample originated. Thus, each cDNA sample can be linked to a single cell, even after the tagged cDNA samples have been pooled in step (iii) of the above method.

Before the tagged cDNA samples are pooled, synthesis of cDNA may be stopped, for example by removing or inactivating the reverse

PARSE EX1015.008

transcriptase. This prevents cDNA synthesis by reverse transcription from continuing in the pooled samples.

The tagged cDNA samples may optionally be purified before amplification, ether before or after they are pooled.

The pooled cDNA samples may be amplified by PCR in step (iv) of the above method. For example, the cDNA samples may be amplified by single primer PCR. The CDS may comprise a 5′ amplification primer sequence (APS), which subsequently allows the first strand of cDNA to be amplified by PCR using a primer that is complementary to the 5′ APS. The TSO may also comprise a 5′ APS, which may be at least 70% identical, at least 80% identical, at least 90% identical, at least 95% identical, or 70%, 80%, 90% or 100% identical to the 5′ APS in the CDS. This means that the pooled cDNA samples may be amplified by PCR using a single primer (i.e. by single primer PCR), which exploits the PCR suppression effect to reduce the amplification of short contaminating amplicons and primer-dimers (Dai, Zhu et al. 2007). As the two ends of each amplicon are complementary, short amplicons will form stable hairpins, which are poor templates for PCR. This reduces the amount of truncated cDNA and improves the yield of longer cDNA molecules. The 5′ APS may be designed to facilitate downstream processing of the cDNA library. For example, if the cDNA library is to be analysed by a particular sequencing method, e.g. Applied Biosystems' SOLiD sequencing technology, or Illumina's Genome Analyser, the 5′ APS may be designed to identical to the primers used in these sequencing methods. For example, the 5′ APS may be identical to the SOLiD P1 primer, and/or a SOLiD P2 sequence could be inserted in the CDS, so that the P1 and P2 sequences required for SOLiD sequencing are integral to the amplified library.

The method of preparing a cDNA library described above may further comprise processing the cDNA library to obtain a library suitable for sequencing. In particular, the cDNA library may be suitable for any high-throughout screening method, such as Applied Biosystems' SOLiD sequencing technology, or Illumina's Genome Analyser. As

PARSE EX1015.009

such, the cDNA library may be processed by fragmenting the cDNA library (e.g. with DNase) to obtain a short-fragment 5'-end library. Adapters may be added to the cDNA, e.g. at one or both ends to facilitate sequencing of the library. The cDNA library may be further amplified, e.g. by PCR, to obtain a sufficient quantity of cDNA for sequencing.

According to the invention, there is also provided a cDNA library produced by the method described above. This cDNA library may be sequenced to provide an analysis of gene expression in single cells.

Also, according to the invention there is provided a method for analysing gene expression in a plurality of single cells, the method comprising the steps of:

(i) preparing a cDNA library using the method described above; and

(ii) sequencing the cDNA library.

The cDNA library may be sequenced by any suitable screening method. In particular, the cDNA library may be sequenced using a high-throughout screening method, such as Applied Biosystems' SOLiD sequencing technology, or Illumina's Genome Analyser. The cDNA library may be shotgun sequenced. The number of reads may be least 10,000, at least 1 million, at least 10 million, at least 100 million, or at least 1000 million. A read is a length of continuous nucleic acid sequence obtained by a sequencing reaction.

The invention will now be illustrated by reference to the following non-limiting example, with reference to the accompanying drawings. All documents cited anywhere herein are incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows an overview of a method of analysing gene expression in a plurality of single cells. (A) The tissue of interest is dissected; (B) a plurality of single cells are selected; (C) single

PARSE EX1015.010

cells are placed in separate wells of a 96-well plate and lysed; tagged reverse transcription is performed on each sample to produce cDNA; (D) cDNA samples are pooled and amplified; (E) sequencing is performed to obtain 100 million reads; and (F) identification of expressed genes and identification of cells from which they originated.

Figure 2 shows the synthesis of cDNA by template switching.

Figure 3 shows an example of a template switching oligonucleotide comprising a 5′ amplification primer sequence (APS), a cell tag and a 3′ sequence for template switching.

Figure 4 shows an example of a cDNA synthesis primer (CDS) comprising a 5′ amplification primer sequence (APS), a cell tag and a 3′ RNA complementary sequence (RCS).

Figure 5 shows visualisation of cDNA samples L001 and L002 following full-length cDNA amplification by PCR. Lane 1: 100 bp marker ladder; lanes 2-3: 25 cycles; lanes 4-5: 30 cycles; lanes 6-7: 35 cycles. Even lanes contain sample L001 and odd lanes contain sample L002.

Figure 6 shows a dilution series from a test PCR using sample L001 (lanes 3-10). Lanes 2 and 11 contain a 100 bp ladder as a size marker.

Figure 7 shows a cDNA library following final amplification by PCR (16 cycles) (lanes 5 and 6). In (B), the 125-200 bp region has been excised. Lanes 3 and 8 contain 100 bp ladder as a size marker.

Figure 8 shows an example of a sequenced cDNA molecule from a tagged cDNA library. The primer sequences for SOLiD sequencing (P1 and P2) are underlined. The cell-specific tag is boxed. The 2-5 Gs from the template-switching mechanism are shaded by a grey box. Sequence from the TOPO cloning vector is shown in italics. The insert in this case is tubulin beta 2c.

PARSE EX1015.011

Figure 9 shows results from a quantitative real-time PCR comparing bulk cDNA (horizontal axis) *versus* 96-cell tagged cDNA (vertical axis). Each circle represents a PCR primer pair directed against the indicated genes. The units are arbitrary and are derived from the cycles-to-threshold value, C_T.

An embodiment of the method of the invention is summarised in Figure 1. Cells are obtained from a tissue of interest and a single-cell suspension is obtained. A single cell is placed in one well of a 96-well plate in Cell Capture Mix. The cells are lysed and reverse transcription reaction mix is added directly to the lysates without additional purification. This results in the synthesis of cDNA from cellular mRNA and incorporation of a tag into the cDNA. The tagged cDNA samples are pooled and amplified and then sequenced to produce 100 million reads. This allows identification of genes that are expressed in each single cell.

This embodiment of the method of the invention may be called "single-cell tagged reverse transcription" (STRT) and is described in detail below.

Cell collection and lysis

A 96-well plate containing Cell Capture Mix was made by aliquoting 5 µl/well from the Cell Capture Master Plate (see Table 1 below) into an AbGene Thermo-Fast plate.

Reagent	**For one well**	**For one plate**	**Final concentration**
STRT-T30-BIO (100 µM)	0.25 µL	27.5 µL	400 nM
STRT buffer (5x)	12.5 µL	1375 µL	1x
STRT-FW-n (5 µM)	5 µL	(5 µL/well)	400 nM
Water	44.75	4.9 mL	
Total	**62.5 µL**		

Table 1 Making a STRT Cell Capture Master Plate. 27.5 µL STRT-T30-BIO (100 µM) was mixed with 1375 µL STRT 5x buffer and 4.9 mL Rnase/Dnase-free water. 57.5 µL of this solution was aliquoted to

PARSE EX1015.012

each well of a 96-well plate and 5 µL/well of STRT-FW-n (from 5 µM stock plate) was added, i.e. a different oligo in each well.

The sequence of STRT-T30-BIO (which is a CDS) is:

5'-BIO-AAGCAGTGGTATCAACGCAGAGT$_{30}$VN-3',

and the sequence of STRT-FW-n (which is a TSO) is:

5'-AAGCAGTGGTATCAACGCAGAGTGGATGCTXXXXXrGrGrG-3'(X=cell tag)
BtsCI->2/0

n is 1-96 and each oligonucleotide has a distinct cell tag, such that a different oligonucleotide is added to each well containing a single cell.

Mouse embryonic stem cells (R1) were grown without feeder cells, trypsinized, cleared through cell strainer and resuspended in 1x PBS. Cells were then picked by FACS into the Capture Plate, with a single cell being placed in each well. The Capture Plate was transferred to a PCR thermocycler and incubated at 72°C for 2 minutes, and then cooled to 4°C for 5 minutes to allow annealing to occur. The detergent in STRT buffer helps reduce adsorption of mRNA and cDNA to the walls of the reaction tube during subsequent steps, and also improves lysis of the cells. The heating step causes the cell to lyse completely and release its RNA. When the temperature is reduced, the oligo(dT) primer anneals.

Reverse transcription

5 µL/well of RT mix (see Table 2 below) was added and the plate was incubated at 42°C for 45 minutes, without heated lid.

Reagent	**For 96 reactions**	**For one reaction**	**Final conc.**
STRT buffer (5x)	110 µL	1 µL	1x
DTT (20 mM)	110 µL	1 µL	2 mM
dNTP (10 mM)	110 µL	1 µL	1 mM
Dnase/Rnase-free water	209 µL	1 µL	
Superscript II RT (200 U/µL)	11 µL	1 µL of 20 U/µL (dilute in 1x RT buffer)	2.5 U/µL
Total volume	**550 µL**	**5 µL**	

PARSE EX1015.013

Table 2 Composition of RT mix

When the RT mix is added, the reverse transcriptase enzyme (Superscript II RT) synthesizes a first strand and the tagged template-switching oligo introduces an upstream primer sequence.

Figure 2 shows the synthesis of cDNA by template switching. The 5′ end of the cDNA (which corresponds to the 3′ end of the mRNA) can be controlled by adding a tail (which is oligo-dT in this case) to the cDNA synthesis primer (CDS). The 3′ end of the cDNA can be controlled using the template-switching oligo (TSO). When reverse transcriptase reaches the 5′ end of the mRNA template, it preferentially adds 2-5 cytosines. The template-switching oligo, which has 2-5 guanosines, anneals transiently, and reverse transcriptase then switches template and synthesizes the complementary strand. By this mechanism, both ends of the cDNA can be arbitrarily controlled.

The structure of a typical TSO is shown in Figure 3. In this particular TSO, the 3′ template switching sequence includes three riboguanines (rG). The cell-tag is shown as "XXXXX" and may in general have any length or nucleotide composition. Arbitrary sequence can be inserted at the 5′ end of the TSO, after the 5′ APS, or after the cell-tag, but not at the 3′ end.

The structure of a typical CDS is shown in Figure 4. The RCS is oligo-dT with an anchor nucleotide (V = A,C,G degenerate). The cell tag "XXXXX" may have any length or nucleotide composition. Additional arbitrary sequences can be inserted at the 5′ end, after the 5′ APS, or after the cell-tag.

Purification of cDNA

50 µl PBI (Qiaquick PCR Purification Kit) was added to each well to inactivate reverse transcriptase. The PBI inactivates reverse transcriptase and cDNA from all the wells was then pooled. Adding PBI before pooling prevents cDNA synthesis from proceeding once the cDNA samples have been pooled. The pooled cDNA was loaded on a

PARSE EX1015.014

single Qiaquick column and the purified cDNA was eluted in 30 µl EB buffer into a Beckman Polyallomer tube. The purification step removes primers (<40 bp) as well as proteins and other debris.

Full length cDNA amplification

The cDNA was amplified by PCR by adding the reagents shown in Table 3.

Reagent	**For one tube:**	**Final conc.**
Rnase/Dnase-free water	54 µL	
Advantage2 PCR buffer (10x)	10 µL	1x
dNTP (10 mM)	2 µL	200 µM
STRT-PCR (10 µM)	2 µL	200 µM
Advantage2 DNA Polymerase mix (50x)	2 µL	1x
Total volume	**100 µL**	

Table 3. Reagents used for full-length cDNA amplification.

The sequence of STRT-PCR is:
5'-BIO-AAGCAGTGGTATCAACGCAGAGT-3'

PCR was performed using a heated lid as follows: 1 min @ 95°C, 25 cycles of [5 s @ 95°C, 5 s @ 65°C, 6 min @ 68°C], 4°C forever.

30 µL of the reaction was transferred to a new PCR tube, labelled "Optimization". The remaining 70 µL was stored at 4°C until later. 10 µL from the Optimization tube was removed and the rest of the sample was run for three more cycles. This was repeated to obtain aliquots from 25, 28 and 31 cycles. A diagnostic 2% agarose gel was used to determine the optimal cycle number (which is the cycle just before saturation of the PCR), as well as to visualize the size range of the product (see Figure 5). Typically, the optimal number of cycles was around 28. The remaining 70 µL of reaction was run to reach the optimal number of cycles (in addition to the 25 cycles already run).

The PCR product was purified using a Qiaquick column (PCR purification kit) and eluted in 50 µL EB into a Beckman polyallomer

PARSE EX1015.015

tube. The expected concentration at this stage was about 20-40 ng/µL (1-2 µg total yield).

DNase treatment

The sample was treated with DNaseI in the presence of Mn^{2+} to generate double-strand breaks and reduce the size. First, the following components were mixed in the order shown in Table 4.

Reagent	**Volume**	**Final concentration**
cDNA template	50 µl	8 – 16 ng/µL
Water	42.8 µl	
10 x DNase I buffer	11.6 µl	
100 mM $MnCl_2$ (*)	11.6 µl	10 mM
Total volume	**116 µl**	

Table 4. Composition of reaction mix for DNase treatment. (*) It is crucial to add MnCl2 last to the reaction, as the BSA present in the buffer will precipitate otherwise.

Diluted DNaseI (0.01 units/µl) was prepared just before use as follows: 40 µL 10x DNase I buffer, 318 µL water, 40 µL 100 mM $MnCl_2$ and 2 µL DNaseI (2 U/µL).

4 µl of this diluted DNase I was added to the reaction mix described in Table 4, and was incubated at RT for exactly 10 minutes. The reaction was then stopped by adding 600 µL PBI. The sample was purified on a Qiaquick column and eluted in 30 µL EB.

Bead capture and end repair/nick translation

The fragments were next bound to beads to capture 5′ and 3′ ends, and then treated with TaqExpress to repair frayed ends and nicks. 30 µL Dynabeads MyOne C1 Streptavidin were washed twice in 2x B&W (Dynal), then added to the Dnase-treated sample, incubated for 10 minutes, and then washed 3x in 1x B&W. About 10% of the sample was bound to the beads (i.e. about 30 – 60 ng), since internal fragments were not biotinylated.

The beads were washed once in 1x TaqExpress buffer and resuspended in the reaction mix shown in Table 5:

PARSE EX1015.016

Reagent	**Volume**	**Final concentration**
10x TaqExpress buffer (blue, with $MgCl_2$)	4 µL	1x
dNTP (10 mM)	0.8 µL	200 µM
Water	33 µl	
TaqExpress (5 U/µL)	2 µL	0.25 U/µL
Total volume	**40 µl**	

Table 5. Composition of the reaction mix used for end repair/nick translation.

The reaction was incubated at 37°C for 30 minutes, and then washed three times in 1x NEB4 buffer.

Fragment release and RDV/FDV adapter ligation

The fragments were released by BtsCI digestion, and simultaneously ligated to the FDV and RDV adapters. The beads were then resuspended in the reaction mix shown in Table 6.

Reagent	**Volume**	**Final concentration**
10x NEB4 buffer	4 µL	1x
ATP (10 mM)	4 µL	1 mM
Adapter STRT-RDV-A (10 µM)	4 µL	1 µM
Adapter STRT-FDV (10 µM)	4 µL	1 µM
Water	26 µL	
T4 DNA Ligase (5 U/µL; Invitrogen)	2 µL	0.25 U/µL
BtsCI (20 U/µL)	2 µL	1 U/µL
Total volume	**40 µl**	

Table 6. Reaction mix for resuspension of the beads.

The sequence of STRT-FDV, made by annealing STRT-ADP1U and STRT-ADP1L, was:

5′-----CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATCT-3′

3′-PHO-GGTGATGCGGAGGCGAAAGGAGAGATACCCGTCAGCCACTA-PHO-5′

The sequence of STRT-RDV-A, made by annealing STRT-ADP2U-T and STRT-ADP2L was:

5′-----AACTGCCCCGGGTTCCTCATTCTCTT-3′

3′-PHO-TTGACGGGGCCCAAGGAGTAAGAGA-PHO-5′

PARSE EX1015.017

The beads were incubated for 30 minutes at 37°C. The reaction was stopped by adding 200 µL PBI, while the beads were held on the magnet. The supernatant was loaded on a Qiaquick column, purified and eluted in 30 µl EB in a Beckman polyallomer tube. The concentration of the cDNA was about 1 – 2 ng/µL.

Library PCR amplification

Eight reactions were set up using 4, 2, 1, 1/2, 1/4, 1/8, 1/16 and 1/32 µL aliquots of the adapted library, each in 4 µL. Each library was amplified using the PCR reaction mix shown in Table 7.

Reagent	**For one tube:**	**Final conc.**
cDNA library (dilution series)	4 µL	(20 pg/µL)
TaqExpress blue buffer (10x)	5 µL	1x
dNTP (10 mM)	1 µL	200 µM
SOLID-P1 (10 µM)	2 µL	400 nM
SOLID-P2 (10 µM)	2 µL	400 nM
Rnase/Dnase-free water	35.7 µL	
TaqExpress polymerase (25 U/µL)	0.3 µL	0.15 U/µL
Total volume	**50 µL**	

Table 7. PCR reaction mix for cDNA library amplification.

The sequence of SOLID-P1 was:
5′-CCACTACGCCTCCGCTTTCCTCTCTATG-3′

The sequence of SOLID-P2 was:
5'-CTGCCCCGGGTTCCTCATTCTCT-3′

PCR was run with heated lid: 5 min @ 94°C, 18 cycles of [15 s @ 94°C, 15 s @ 68°C], 5 min @ 70°C.

All eight reactions were loaded on a 2% E-gel, 10 µL + 10 µL water to determine which reaction was just shy of saturated (see Figure 6).

PARSE EX1015.018

A fresh PCR reaction was then performed using the optimal number of cycles and starting material. For example, if 1/4 µL was optimal at 18 cycles, then 14 cycles were performed.

The PCR product was loaded on a 2% E-gel, 125 – 200 bp region was excised from the gel and purified by Qiagen Gel Extraction Kit (see Figure 7). The purified cDNA was eluted in 50 µL EB.

The cDNA library was now prepared for SOLiD sequencing, and could go directly into emulsion PCR.

Quality control by Sanger sequencing

To verify cDNA library quality, an aliquot was cloned using Invitrogen TOPO TA cloning kit, and sequenced by Sanger sequencing. Figure 8 shows a typical result demonstrating the presence of primer sequences for SOLiD (P1 and P2; underlined), the cell-specific tag (boxed), and the 2-5 Gs (shaded in a gray box) from the template-switching mechanism. From 22 Sanger sequences, 7 were not mappable to anything in GenBank. All except one of these were misligations of the SOLiD adapters, which can be re-designed to stop this happening. In separate experiments, we found no misligated adapters after blocking their blunt ends with 3′ phosphate. Alternatively, the non-ligating 3′ ends could be blocked using dideoxy nucleotides or by designing a protruding strand incompatible with the ligating ends of the adapters.

Of the remaining 15 sequences, one was a ribosomal RNA (45S), which is not polyadenylated. It probably occurred due to internal mispriming during first-strand synthesis. The remaining 14 reads were all from polyadenylated mRNA, in the correct orientation and with correct cell tags.

To summarize this dataset, 15 of 22 reads were mappable and 14 of these 15 were correct transcript tags.

All the transcripts seen in the Sanger sequence dataset are listed below:

PARSE EX1015.019

Gene	Length (of mRNA)
Ribosomal protein L35	452
B2_Mm2	~200
Tubulin beta 2c	1 561
B2_Mm1	~195
RIKEN 1110008L16 gene	3 127
Sod2	661
Chchd2	910
mt-Cox2	947
Hnrnpab	2 545
Ribosomal protein L24	558
Ribosomal protein S18	524
RIKEN 2700060E02	941
B2_Mm1	~195
Ribosomal protein S28	356

As expected, the list was dominated by highly expressed genes like ribosomal proteins. Several long transcripts were present in this sample, indicating that there was no strong bias (if any) towards short mRNAs.

Interestingly, three copies of B2 repeats (of subfamilies Mm1 and Mm2) were observed. These are SINE-family repeats expressed from a pol III promoter (not pol II as most mRNAs), but with strong polyadenylation signals. They have been shown to be expressed at extremely high levels in ES cells, together comprising more than 10% of all mRNA. Even more interestingly, they peak just before S-phase in dividing cells, and are thus an early indication that it will be possible to characterize the cell cycle in unsynchronized primary cells using this method.

Quality control by quantitative real-time PCR

To verify that the libraries were representative of the mRNA content of the original ES cell population, quantitative real-time PCR was performed against a set of markers for pluripotency, as well as markers for differentiated tissues. A cDNA library prepared according to classical methods from 1 µg total RNA

PARSE EX1015.020

(~100,000 cells) was compared with the library prepared from 96 single cells using the STRT protocol.

Well-known markers of pluripotency, such as Sox2, Oct4 and Nanog were detected at similar levels in both samples, whereas markers of germ layer differentiation such as Brachyury, Gata4 and Eomes were detected only at very low levels in both samples (see Figure 9). The quantitative correlation was good (Pearson correlation coefficient 0.84), with the exception of Plk1, which was not detected in single cells in this experiment.

Reagents used

Reagent	**Source**
2x B&W	10 mM Tris HCl pH 7.5, 1 mM EDTA, 2 M NaCl
10% Tween-20 stock	Make 10% Tween-20, filter at 0.45 µm
Dynabeads MyOne Streptavidin C1	Invitrogen
5x STRT buffer	100 mM Tris-HCl pH 8, 375 mM KCl, 30 mM $MgCl_2$, 0.1% Tween-20
DNase/RNase-free water	Ambion
20 mM DTT	
10 mM and 1 mM dNTP	NEB / In Vitro
Superscript II RT	Invitrogen
10x Advantage2 PCR buffer	Clontech
50x Advantage2 polymerase mix	Clontech
10x DNase I buffer	0.5 M Tris pH 7.5, 0.5 mg/ml BSA
100 mM $MnCl_2$	Sigma 1M stock
Thermo-Fast 96-well plate	AbGene
Thermo-Strip tube strips	AbGene
1.5 mL polyallomer tubes	Beckman-Coulter
2% E-gel with SYBR Safe	Invitrogen
Qiaquick PCR Purification Kit	Qiagen
Qiaquick Gel Extraction kit	Qiagen
BtsCI restriction endonuclease	NEB
T4 DNA Ligase	Invitrogen
10 mM ATP	
10x NEB4 buffer	NEB
5 U/µLTaqExpress polymerase	Genetix, UK

PARSE EX1015.021

10x TaqExpress buffer (blue)	Genetix, UK

Table 8. List of reagents used in the method described above.

PARSE EX1015.022

REFERENCES

Bengtsson, M., A. Stahlberg, et al. (2005). "Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels." Genome Res **15**(10): 1388-92.

Brail, L. H., A. Jang, et al. (1999). "Gene expression in individual cells: analysis using global single cell reverse transcription polymerase chain reaction (GSC RT-PCR) [In Process Citation]." Mutat Res **406**(2-4): 45-54.

Cloonan, N., A. R. Forrest, et al. (2008). "Stem cell transcriptome profiling via massive-scale mRNA sequencing." Nat Methods **5**(7): 613-9.

Dai, Z. M., X. J. Zhu, et al. (2007). "PCR-suppression effect: kinetic analysis and application to representative or long-molecule biased PCR-based amplification of complex samples." J Biotechnol **128**(3): 435-43.

Esumi, S., N. Kakazu, et al. (2005). "Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons." Nat Genet **37**(2): 171-6.

Esumi, S., S. X. Wu, et al. (2008). "Method for single-cell microarray analysis and application to gene-expression profiling of GABAergic neuron progenitors." Neurosci Res **60**(4): 439-51.

Kurimoto, K., Y. Yabuta, et al. (2006). "An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis." Nucleic Acids Res **34**(5): e42.

Kurimoto, K., Y. Yabuta, et al. (2007). "Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis." Nat Protoc **2**(3): 739-52.

Levsky, J. M., S. M. Shenoy, et al. (2002). "Single-cell gene expression profiling." Science **297**(5582): 836-40.

PARSE EX1015.023

CLAIMS

1. A method of preparing a cDNA library from a plurality of single cells, the method comprising the steps of:

(i) releasing mRNA from each single cell to provide a plurality of individual mRNA samples, wherein the mRNA in each individual mRNA sample is from a single cell;

(ii) synthesising a first strand of cDNA from the mRNA in each individual mRNA sample and incorporating a tag into the cDNA to provide a plurality of tagged cDNA samples, wherein the cDNA in each tagged cDNA sample is complementary to mRNA from a single cell;

(iii) pooling the tagged cDNA samples; and

(iv) amplifying the pooled cDNA samples to generate a cDNA library comprising double-stranded cDNA.

2. A method according to claim 1, wherein in step (ii) the tag is incorporated into the cDNA during its synthesis.

3. A method according to claim 1 or 2, wherein synthesis of the first strand of cDNA in step (ii) is directed by a cDNA synthesis primer (CDS) that includes an RNA complementary sequence (RCS) that is at least partially complementary to some mRNA in an individual mRNA sample.

4. A method according to claim 3, wherein the RCS is at least partially complementary to a portion of the first strand of cDNA, such that it is able to direct the synthesis of a second strand of cDNA using the first strand of cDNA as a template.

5. A method according to any one of claims 2 to 4, wherein a template-switching oligonucleotide (TSO) that includes a portion which is at least partially complementary to a portion at the 3′ end of the first strand of cDNA is added to each individual mRNA sample.

PARSE EX1015.024

6. A method according to any one of claims 2 to 5, wherein the first strand of cDNA includes a plurality of cytosines, or cytosine analogues that base pair with guanosine, at its 3′ end.

7. A method according to claim 6, wherein the TSO includes a 3′ portion comprising a plurality of guanosines, or guanosine analogues that base pair with cytosine.

8. A method according to claim 7, wherein the guanosines are ribonucleosides or locked nucleic acid monomers.

9. A method according to any one of the preceding claims, wherein the CDS or the TSO includes a tag.

10. A method according to any one of claims 2 to 8, wherein both the CDS and the TSO include a tag.

11. A method according to claim 10, wherein the CDS and the TSO each include a different tag, such that the tagged cDNA sample comprises a combination of tags.

12. A method according to any one of claims 9 to 11, wherein the tag is a nucleotide sequence of 4-20 nucleotides in length.

13. A method according to any one of the preceding claims, wherein each cDNA sample has a distinct tag or combination of tags.

14. A method according to any one of claims 2 to 13, wherein the CDS comprises a 5′ amplification primer sequence (APS) and a 3′ RCS.

15. A method according to claim 13, wherein the 3′ RCS comprises oligo(dT), a gene family-specific sequence, a random sequence or a non-self-complementary semi-random sequence.

16. A method according to any one of claims 5 to 15, wherein the TSO includes a 5′ APS.

PARSE EX1015.025

17. A method according to claim 15 or 16, wherein the CDS and the 5′ APS of the TSO is at least 80% identical to the 5′ APS of the CDS.

18. A method according to claim 17, wherein the CDS and the 5′ APS of the TSO is 100% identical to the 5′ APS of the CDS.

19. A method according to any one of the preceding claims, wherein the cells are lysed to release mRNA.

20. A method according to any one of the preceding claims, wherein the mRNA is purified following step (i).

21. A method according to any one of the preceding claims, wherein the synthesis of cDNA from mRNA is stopped before the tagged cDNA samples are pooled.

22. A method according to any one of the preceding claims, wherein the tagged cDNA samples are purified before amplification of the cDNA.

23. A method according to any one of the preceding claims, wherein in step (iv) the pooled cDNA samples are amplified by PCR.

24. A method according to claim 23, wherein the pooled cDNA samples are amplified by emulsion PCR.

25. A method according to claim 23 or 24, wherein the pooled cDNA samples are amplified by single primer PCR.

24. A method according to any one of the preceding claims, wherein the method further comprises processing the cDNA library to obtain a library suitable for sequencing.

PARSE EX1015.026

25. A method according to claim 24, wherein the processing comprises fragmenting the cDNA library.

26. A method according to claim 24 or 25, wherein the processing includes the step of adding an adapter to the cDNA.

27. A method according to any one of claims 24 to 26, wherein the cDNA is amplified.

28. A cDNA library produced by the method of any one of the preceding claims.

29. A method for analysing gene expression in a plurality of single cells, the method comprising the steps of:

(i) preparing a cDNA library according to the method of any one of claims 1 to 26; and

(ii) sequencing the cDNA library.

30. A method according to claim 29, wherein sequencing is by shotgun sequencing.

31. A method according to claim 30, wherein the cDNA library is sequenced to obtain at least 10,000, at least 1 million, at least 10 million, at least 100 million, or at least 1 billion reads, wherein a read is a length of continuous nucleic acid obtained by a sequencing reaction.

PARSE EX1015.027

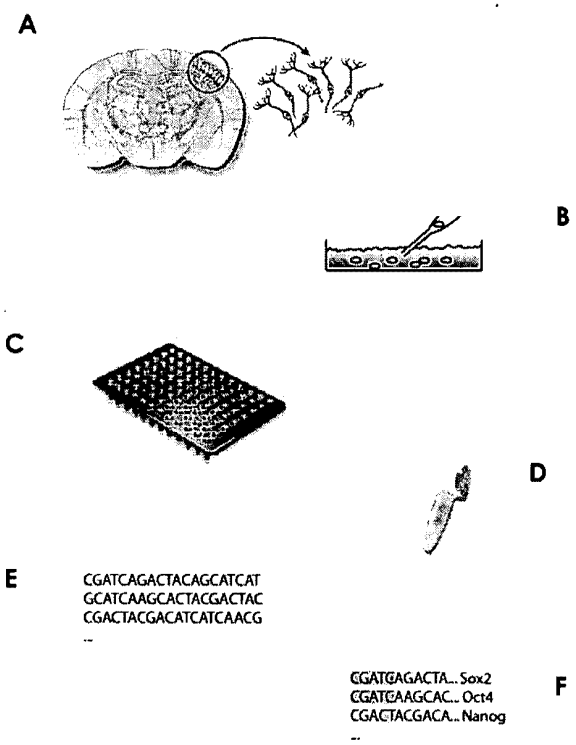


Figure 1

PARSE EX1015.028

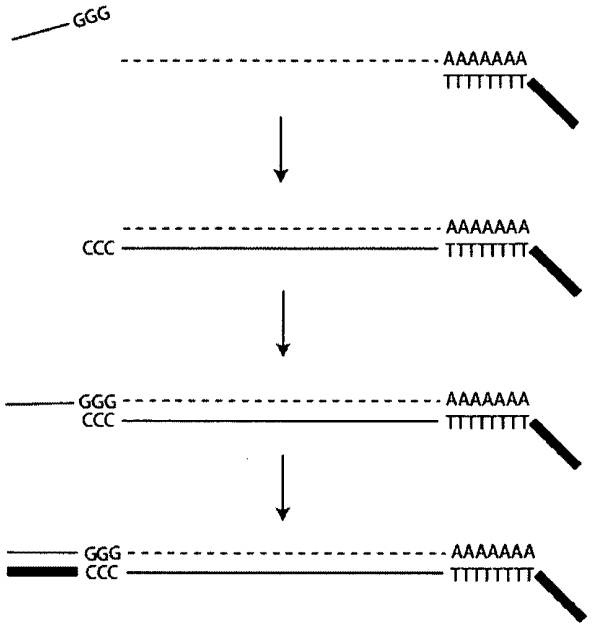

GGG
AAAAAAA
TTTTTTTT
AAAAAAA
CCC
TTTTTTTT
GGG
AAAAAAA
CCC
TTTTTTTT
GGG
AAAAAAA
CCC
TTTTTTTT

Figure 2

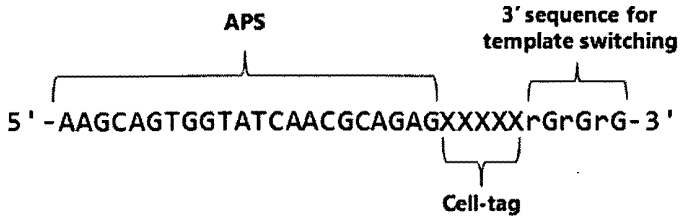


Figure 3

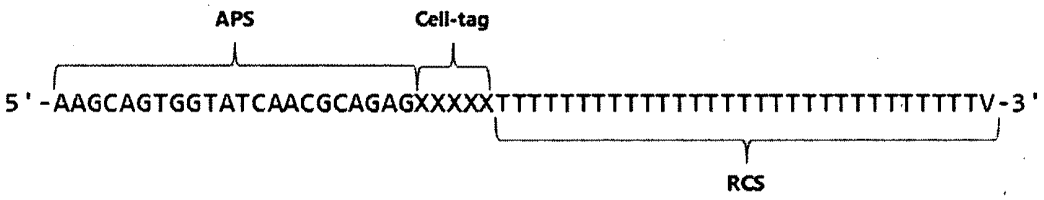


Figure 4

PARSE EX1015.030

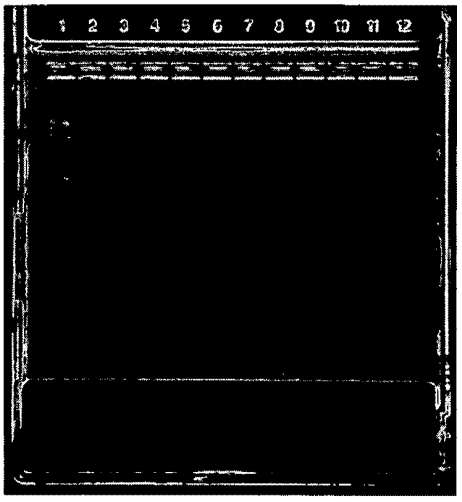


Figure 5

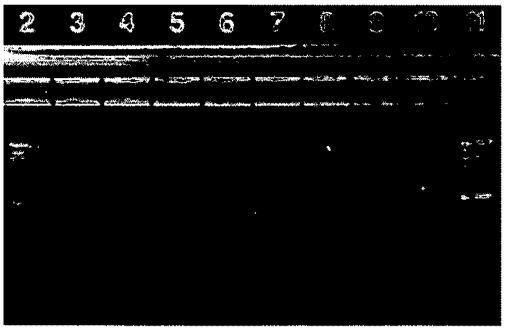


Figure 6

PARSE EX1015.031

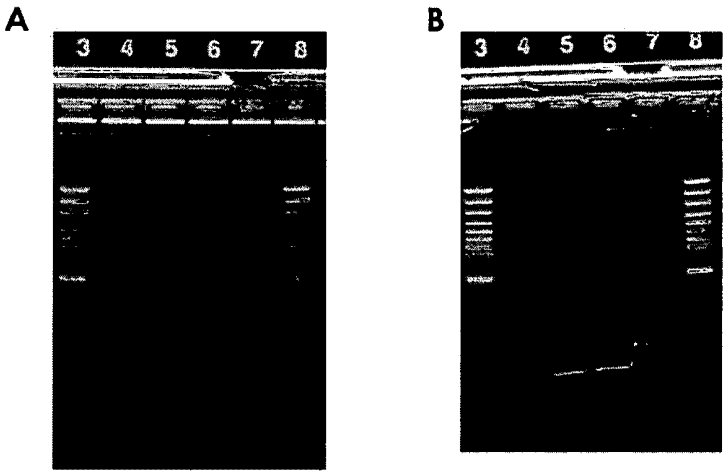


Figure 7

PARSE EX1015.032

*GTGGCTCTGATGCATGCTCGAGCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTT*CCACTAC
GCCTCCGCTTTC
CTCTCTATGGGCAGTCGGTGATCTCAGAAGCGCGATCCAAGACATGGGCGCTACTTGACTGTGGCTGC
CGTGTTCAGGGGC
CGCATG*AAGAGAATGAGGAACCCGGGGCAG**AAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGT***
GGATCCGAGCTCG
GTACCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCC
ACACAACATACG

Figure 8

PARSE EX1015.033

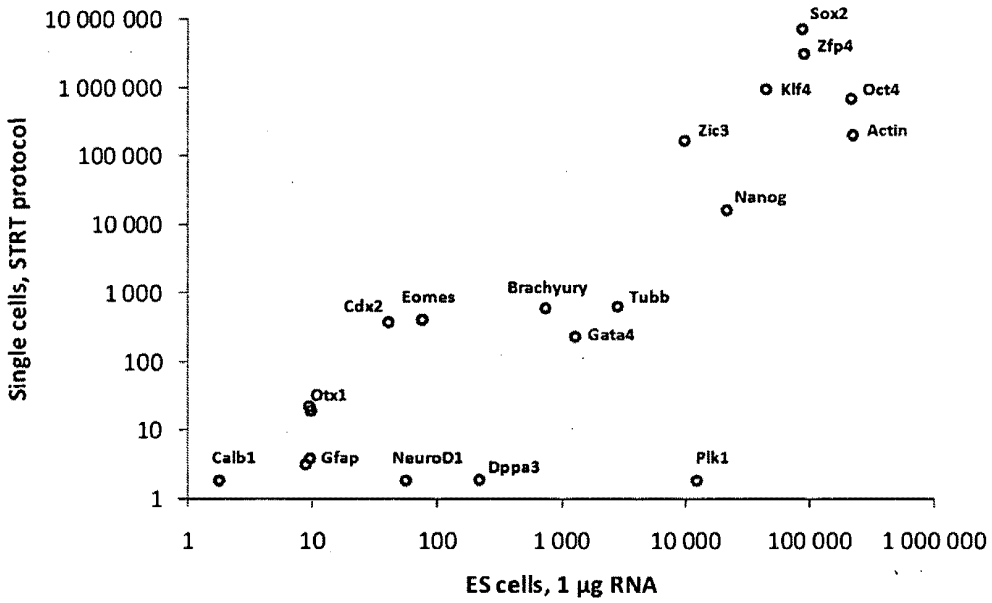


Figure 9

PARSE EX1015.034