Amplification of Nanogram Amounts of Total RNA by the SMART-Based PCR Method for High-Density Oligonucleotide Microarrays, Wenli Zhou,* Ronald V. Abruzzese, Irina Polejaeva, Sara Davis, Scott Davis, and Wan Ji (ViaGen Inc., Austin, TX; * address correspondence to this author at: ViaGen Inc., 12357-A Riata Trace Pkwy., Suite 100, Austin, TX 78727; fax 512-401-5919, e-mail wenli.zhou@viagen.com)

The SMART (switching mechanism at the 5′ end of RNA templates of reverse transcriptase) method (1, 2), in combination with PCR, has been used to amplify minute samples obtained from sources such as laser capture microdissections and biopsies for gene expression analysis using oligonucleotide microarray gene chip technology (3–7). Amplifications >10^5-fold can be achieved in a short time. However, reports in the literature have rarely discussed the feasibility and reproducibility of PCR-based methods when the amount of starting total RNA in a sample is only a very few nanograms or less, which is often the case for clinical samples.

We developed a modified SMART-based PCR protocol with which reliable and robust gene expression data can be obtained from 1 ng of starting total RNA. We also investigated sources of array data variations.

The modified protocol (see Supplement 1 in the Data Supplement that accompanies the online version of this Technical Brief at http://www.clinchem.org/content/vol51/issue12) has three major changes compared with the published standard procedure (5): the volume of the reverse transcription reaction is reduced to 5 µL; 50 ng of poly(dG/dC) carrier (Sigma) is added to the reverse transcription reaction; and PowerScript reverse transcriptase (BD Biosciences Clontech) is used instead of Superscript II reverse transcriptase (Invitrogen). When we used the modified protocol to reverse-transcribe and amplify 10, 1, 0.1, and 0.01 ng of human heart or liver total RNA (BD Biosciences Clontech), we obtained ample cDNA from one round of 23, 26, 29, and 32 PCR cycles, respectively (Fig. 1A). The addition of poly(dG/dC) facilitates the reverse transcription of longer transcripts as indicated by larger ranges of cDNA sizes and clearer band patterns, which are important for getting consistent good array data, but it does not adversely interfere with the reactions. Use of Superscript II instead of PowerScript in the reverse transcription step under otherwise similar conditions failed to yield visible amounts of cDNA in the agarose gel when starting total RNA was ≤10 ng, and a second round of PCR had to be performed to yield sufficient cDNA for synthesizing adequate amounts of cRNA for arrays. The number of genes detected on the arrays, however, was considerably lower than that obtained with PowerScript reverse transcriptase.

The cDNA obtained with the modified protocol from 1, 10, and 100 ng of human total heart or liver RNA was in vitro–transcribed into complementary RNA (cRNA), which was then labeled with Bio-11-CTP and Bio-16-UTP (Enzo Biochem) by use of MAXiScriptr™ reagents (Ambion). The cRNA was then hybridized to the human GeneChip H133A (Affymetrix) as described previously (5). The numbers of genes detected as present (~30%) were comparable among arrays prepared with various starting amounts of total RNA. We determined the reproducibility of the expression signals detected from these minute samples by running 2 replicate assays. Shown in Fig. 1B are the M–A plots of replicate arrays (M represents the log ratio of 2 replicates, and A represents the averaged logarithmic intensity). At high signal intensities (averaged raw signal intensity >1024 or logarithmic intensity >10), except for a few outliers, the data in all 3 plots were scattered around the line M = 0, indicating very good reproducibility even when starting the amount of total RNA was 1 ng. However, at lower signal intensities, particularly when the averaged raw signal intensity was <256 or the logarithmic intensity was <8, the data in the plots were scattered farther from the M = 0 line. The smaller the amount of starting total RNA used, the greater was the observed variation. When replicate arrays were prepared from 100 ng of starting total RNA, ~1.2% of the examined probe sets exhibited more than 2-fold differences in expression values between 2 replicates; this number increased to 4.2% for arrays prepared from 10 ng of starting total RNA and 18.2% for 1 ng of starting total RNA.

We then conducted quadruple replicate assays for arrays prepared from 1 ng of starting total RNA. The 4 replicates were randomly divided into 2 groups of duplicate samples. Mean expression intensities from 2 arrays were calculated for each probe set in each group. The variation was greatly reduced, with 7.6% of probe sets exhibiting greater than 2-fold differences between duplicate groups. For the following studies, therefore, the mean of 2 replicates was used when the starting total RNA was 1 ng.

To investigate whether the modified protocol can faithfully amplify massive genes from 2 samples under different conditions so that the expression ratio for each gene in the 2 samples will be preserved and correlate to that obtained from linear amplification, we obtained independent gene expression ratios for human heart and liver, using both the linear amplification scheme with Message-Amp™ reagents (Ambion) and the modified protocol developed for nanograms of total RNA, with 2 μg each of heart or liver total RNA as starting material. Both methods generated strong comparable hybridization signals in the gene chips (see Supplement 2 in the online Data Supplement). The Pearson correlation coefficients for the gene expression ratios in human heart and liver obtained with the 2 methods were 0.95, 0.94, and 0.92 when the amounts of starting total RNA for the PCR-based amplification were 100, 10, and 1 ng, respectively. Despite the amplification preference associated with each method (3–5), we achieved good correlation between expression ratios obtained from both methods for the majority of transcripts.

The fact that this modified PCR-based protocol reproducibly preserves expression ratios can be further demonstrated by comparing gene expression ratios obtained
under different amplification conditions, i.e., various amounts of starting total RNA with various numbers of PCR cycles. Expression arrays were prepared from 100 ng of total RNA amplified for 20 PCR cycles, 10 ng of total RNA amplified for 23 PCR cycles, or 1 ng of total RNA amplified for 26 PCR cycles. The correlation coefficients ($r$) for the 100- and 10-ng, 100- and 1-ng, and 10- and 1-ng array pairs were 0.97, 0.95, and 0.92, respectively. The preservation of expression ratios over such a wide range of PCR cycles and template concentrations can be explained theoretically (see Supplement 3 in the online Data Supplement) and makes more convenient the study of clinical samples, in which the amount of starting material may be very different.

During preparation of cRNA for chip hybridization from nanograms of total RNA, reverse transcription and PCR are the only steps involving minute amounts of reactants. The fact that the array data became more variable when the amount of starting total RNA was $<10$ ng indicates that this variation is introduced primarily in these 2 early reactions. To further understand which of the 2 reactions, reverse transcription or PCR, generates the larger variation, we performed 3 sets of replicate assays. In the first assay, we reverse-transcribed 100 ng of total heart RNA. We then used 1/100th of the first-strand cDNA, which was quantitatively equivalent to the cDNA reverse-transcribed from 1 ng of total starting RNA, for PCR amplification and cRNA synthesis. In the second assay, we reverse-transcribed 100 ng of total heart RNA and used all the first-strand cDNA for PCR amplification and cRNA synthesis. The third assay was the same as the second assay except that the starting total RNA used was 1 ng. Replicate cRNA samples for each set of arrays were then hybridized to gene chips, and the Pearson correlation coefficients were 0.87, 0.96, and 0.88, respectively, for the 3 sets of replicate arrays (see Supplement 4 in the online Data Supplement). If we regard 1 ng as minute, the second set of arrays ($r = 0.96$) does not have any minute amounts of reactants. In the first set of arrays, only the PCR step involved minute amounts of reactants. In the third set of arrays, both reverse transcription and PCR used minute amounts of reactants, but the Pearson correlation coefficients of these 2 sets of arrays were basically the same, 0.87 and 0.88, suggesting that the PCR step.

![Agarose gel image of cDNA prepared according to the modified protocol](image)

**Fig. 1.** Agarose gel image of cDNA prepared according to the modified protocol (A), and MA plots of replicate human heart arrays (B).

(A), total RNA (0.01–10 ng) was reverse-transcribed, and cDNA was amplified either with (+ polydG/dC) or without (− polydG/dC) addition of poly(dG/dC) carrier, or with SuperScript II (SS II) instead of PowerScript reverse transcriptase. One tenth of the amplified cDNA product was electrophoresed on the agarose gel. (B), $M$ represents the log ratio of 2 replicates; $A$ represents the averaged logarithmic intensity. Genes that were called present in both replicates are included in the plots. Replicate arrays were made from 100, 10, and 1 ng of total RNA. Dashed lines represent 2-fold differences.
contributes the most to the variation between arrays prepared from 1 ng of starting total RNA. Our findings are consistent with mathematical simulations of PCR amplification illustrating that when transcripts with small copy numbers are amplified, the amplified gene population distribution is broad and has a large standard deviation (8, 9).

In conclusion, the modified protocol greatly increases the sensitivity of array detection when as little as 1 ng of total RNA is amplified. Highly reproducible array data were obtained from different quantities of starting RNA that were amplified with different numbers of PCR cycles. The variation associated with minute samples can be effectively reduced with multiple replicates. The findings that when a minute sample is reverse-transcribed and amplified the sensitivity is determined mainly at the step of reverse transcription and that the reproducibility is largely controlled by the PCR reaction could guide the development of more powerful enzymes and methods.

We thank Dr. Kelcey Walker for editing the manuscript.

References

DOI: 10.1373/clinchem.2005.056721

Identification of Novel Mutations in Patients with Cof- fin–Lowry Syndrome by a Denaturing HPLC-Based Assay, Michele Falco,1 Corrado Romano,2 Antonino Alberti,2 Donatella Greco,2 Carmela Scuderi,3 Emanuela Avola,2 Pinella Failla,2 Serena Belli,4 John L. Tolmie,5 Silvestra Amata,1 and Marco Fichera1*

Coffin–Lowry syndrome (CLS; MIM #303600) is characterized by learning difficulties and dysmorphic traits in male patients and in some female carriers of the X-chromosome–linked gene mutation. The dysmorphic traits, skeletal abnormalities, and other clinical findings have been described (1). Mutations of the RSK2 gene (also called RPS6KA3, MIM *300075), mapping to Xp22.2, are found in the disease (2). The gene encodes a 740-amino acid protein member of the 90-kDa ribosomal S6 serine/threonine kinase family (3). In humans, the RSK family includes 4 growth factor–regulated members (RSK1 to -4) produced in all examined tissues and regions of the brain (4). The highly conserved feature of all these proteins is the presence of 2 nonidentical kinase catalytic domains. The N-terminal kinase domain (amino acids 68–323 in RSK2) is responsible for phosphotransferase activity toward substrates, whereas the C-terminal kinase domain (amino acids 422–675 in RSK2) is necessary for enzymatic activation of the N-terminal domain. RSKs are activated through direct phosphorylation by the mitogen-activated protein kinases/extracellular signal-regulated kinases (MAPK/ERK) in response to insulin and growth factors, oncogenic events, and ultraviolet irradiation. RSKs have been implicated in the stimulation of cell proliferation and differentiation, in the cellular stress response, and in apoptosis (5). To date, only a few RSK2-specific physiologic substrates have been described: the transcription factor CREB, histone H3, STAT3 (6), ATF4 (7), and p53 (8).

According to the most recent data (http://alsace.ustrasbg.fr/chimbo/diag/coffin/index.html), mutations in the RSK2 gene occur within all 22 exons, except exon 2. Approximately one third of the RSK2 gene mutations are missense changes, the other two thirds lead to premature translation termination. The proportion of de novo mutations is unusually high compared with other X-linked disorders, with most cases (80%) being sporadic (1). The 2 largest studies on the molecular diagnosis of CLS used single-strand conformational polymorphism analysis (3,9). The large size of the transcript (a 2223-bp open reading frame) and the distribution of mutations all along its length require a cost-effective screening technique. We propose a highly sensitive and rapid approach to mutation detection based on denaturing HPLC (DHPLC) (10) to screen for RSK2 mutations.

We studied 16 (10 male, 6 female) unrelated Italian and British individuals. Among them, 9 showed a typical CLS phenotype, whereas the remaining 7 exhibited only some of the characteristic features of the syndrome. As described below, a CLS phenotype was also observed in mothers of 3 of the typical CLS patients, but none of the CLS patients had other affected relatives. Genomic DNA was isolated from leukocytes in peripheral blood by salting out procedures after receipt of informed written