# Development and Validation of a T7 Based Linear Amplification for Genomic DNA

# Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

# BMC Genomics

# Development and validation of a T7 based linear amplification for genomic DNA

Chih Long Liu[1,2], Stuart L Schreiber*[1] and Bradley E Bernstein*[1,3]

Address: [1]Department of Chemistry and Chemical Biology, and Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138, USA, [2]Biological and Biomedical Sciences, Division of Medical Sciences, Harvard Medical School, Boston, MA 02115, USA and [3]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA

Email: Chih Long Liu - clliu@fas.harvard.edu; Stuart L Schreiber* - sls@slsiris.harvard.edu; Bradley E Bernstein* - bernstei@slsiris.harvard.edu

* Corresponding authors

## Abstract

**Background:** Genomic maps of transcription factor binding sites and histone modification patterns provide unique insight into the nature of gene regulatory networks and chromatin structure. These systematic studies use microarrays to analyze the composition of DNA isolated by chromatin immunoprecipitation. To obtain quantities sufficient for microarray analysis, the isolated DNA must be amplified. Current protocols use PCR-based approaches to amplify in exponential fashion. However, exponential amplification protocols are highly susceptible to bias. Linear amplification strategies minimize amplification bias and have had a profound impact on mRNA expression analysis. These protocols have yet to be applied to the analysis of genomic DNA due to the lack of a suitable tag such as the polyA tail.

**Results:** We have developed a novel linear amplification protocol for genomic DNA. Terminal transferase is used to add polyT tails to the ends of DNA fragments. Tail length uniformity is ensured by including a limiting concentration of the terminating nucleotide ddCTP. Second strand synthesis using a T7-polyA primer adapter yields double stranded templates suitable for in vitro transcription (IVT). Using this approach, we are able to amplify as little as 2.5 ng of genomic DNA, while retaining the size distribution of the starting material. In contrast, we find that PCR amplification is biased towards species of greater size. Furthermore, extensive microarray-based analyses reveal that our linear amplification protocol preserves dynamic range and species representation more effectively than a commonly used PCR-based approach.

**Conclusion:** We present a T7-based linear amplification protocol for genomic DNA. Validation studies and comparisons with existing methods suggest that incorporation of this protocol will reduce amplification bias in genome mapping experiments.

## Background

Systematic, microarray-based studies are rapidly advancing the fields of biology and medicine. Most commonly, mRNA expression levels are measured on a global, often genome-wide scale. Studies employing this approach have contributed to our understanding of diverse biologi-cal processes from yeast physiology [1,2] to human cancer [3]. More recently, microarray-based approaches have been used to characterize physical properties of the genome, for example, by mapping transcription factor binding sites [4,5] and histone modifications [6,7].

Microarray-based genome mapping protocols utilize the technique of chromatin immunoprecipitation (ChIP) to isolate genomic DNA associated *in vivo* with a transcription factor of interest, or with histones exhibiting a particular post-translational modification. In ChIP, whole cells are treated with formaldehyde to covalently link DNA and associated proteins, and the resulting mix is fragmented by sonication. Protein-DNA complexes are then isolated using antibodies specific for a protein of interest (e.g., a transcription factor or a histone exhibiting a particular covalent modification). Finally, cross-links are reversed and protein degraded to yield purified DNA. Microarrays have proven particularly valuable for the unbiased analysis and identification of DNA isolated by ChIP.

Microarray-based mapping poses unique technical challenges. For example, ChIP typically yields only sub-microgram quantities of DNA. Hence, an amplification step is required to generate sufficient quantities of nucleic acid for microarray analysis. This contrasts with typical mRNA expression analyses where microgram quantities of mRNA are available for direct analysis. Furthermore, genomic DNA lacks a universal sequence tag for priming amplification and labeling steps. Again, this contrasts with expression analysis where the mRNA 3' polyA tail provides a useful priming sequence for cDNA synthesis.

These obstacles have been overcome by two related protocols, both of which are based on the polymerase chain reaction (PCR). The first protocol uses a ligation-mediated PCR strategy (LM-PCR) in which ChIP DNA is ligated to short DNA linkers, and then PCR amplified using the linker sequence for priming [4]. The second uses a random amplification strategy (R-PCR) in which template DNA is first copied using a specialized primer with degenerate 3' sequence and known 5' sequence, and then PCR amplified using the known 5' sequence for priming [5,8,9]. In both cases, the PCR steps yield quantities of nucleic acid sufficient for microarray analysis. Labeled product can be generated by incorporating fluorescent (or otherwise modified) nucleotides during PCR. LM-PCR has been used in several studies, including a systematic genome-wide location analysis of yeast transcriptional activators [10], and an unbiased search for E2F binding sites in a subset of human promoters [11]. Studies employing the R-PCR approach include genome-wide analyses of silencing enzymes [12] and histone modifications [6,7] in yeast, and a mapping of GATA-1 binding sites within the human β-globin locus [13].

There are, however, limitations of a PCR-based approach. Due to the exponential kinetics of PCR amplification, sequence-dependent and length-dependent biases may be amplified exponentially. Furthermore, theoretical considerations suggest that amplification of low abundance spe-

cies can be highly variable. Alternative amplification approaches have been developed for microarray-based mRNA analyses. A widely used protocol, developed by Eberwine and colleagues, uses T7-based linear amplification [14]. In this procedure, double stranded cDNAs are synthesized from mRNA by reverse-transcription (priming off the 3' polyA tail) and second strand synthesis. T7 promoters are incorporated during the cDNA synthesis, and T7 RNA polymerase is subsequently used to transcribe multiple RNA copies from each cDNA (this process is termed in vitro transcription or 'IVT'). Recent studies have applied this approach to microarray analysis and found that it faithfully preserves the initial representation of RNA species [15–17]. Other reports suggest that PCR-based amplification may also preserve this representation [18]. However, IVT remains more attractive from a theoretical perspective as biases should be less severe due to the linear nature of the amplification [19].

To improve the accuracy of microarray-based genome mapping protocols, we have developed a linear amplification protocol for genomic DNA. This protocol uses a terminal transferase tailing step and second strand synthesis to incorporate T7 promoters at the ends of the DNA fragments prior to IVT. We find that this approach efficiently amplifies as little as 2.5 ng of genomic DNA. Here, the protocol is presented along with a systematic evaluation of its efficacy and fidelity.

## Results

### T7-based amplification of genomic DNA

The T7-based IVT protocol for linear amplification of genomic DNA is shown in Figure 1. Since ChIP DNA is fragmented randomly and lacks a universal sequence for priming, we optimized a terminal transferase step to add polyT tails to the 3' ends of each DNA strand. We are able to generate tails with a tight distribution around an optimal length of 30 bps by including a low concentration of a terminating dideoxycytidine nucleotide (data not shown). An anchored primer adapter was designed that contains a 5' T7 promoter and a 3' polyA (18 bps in length) stretch terminating in a single C, G, or T base. This adapter is annealed to the genomic DNA fragments, and Klenow is added for second strand synthesis. In addition to extending the primer adapter, the 3'-5' exonuclease activity of Klenow should trim 3' polyTs in excess of the 18 bps complementary to the adapter, and allow fill in of the appropriate complementary bases (Fig 1). For each dsDNA species present in the starting material, these reactions generate two dsDNAs with T7 promoters at opposite ends.

The reaction products are used as template for IVT using the Eberwine method [14]. From a starting material of ChIP DNA from ~2 × $10^8$ yeast cells this protocol yields between 30 and 60 μg RNA. This yield compares favorably
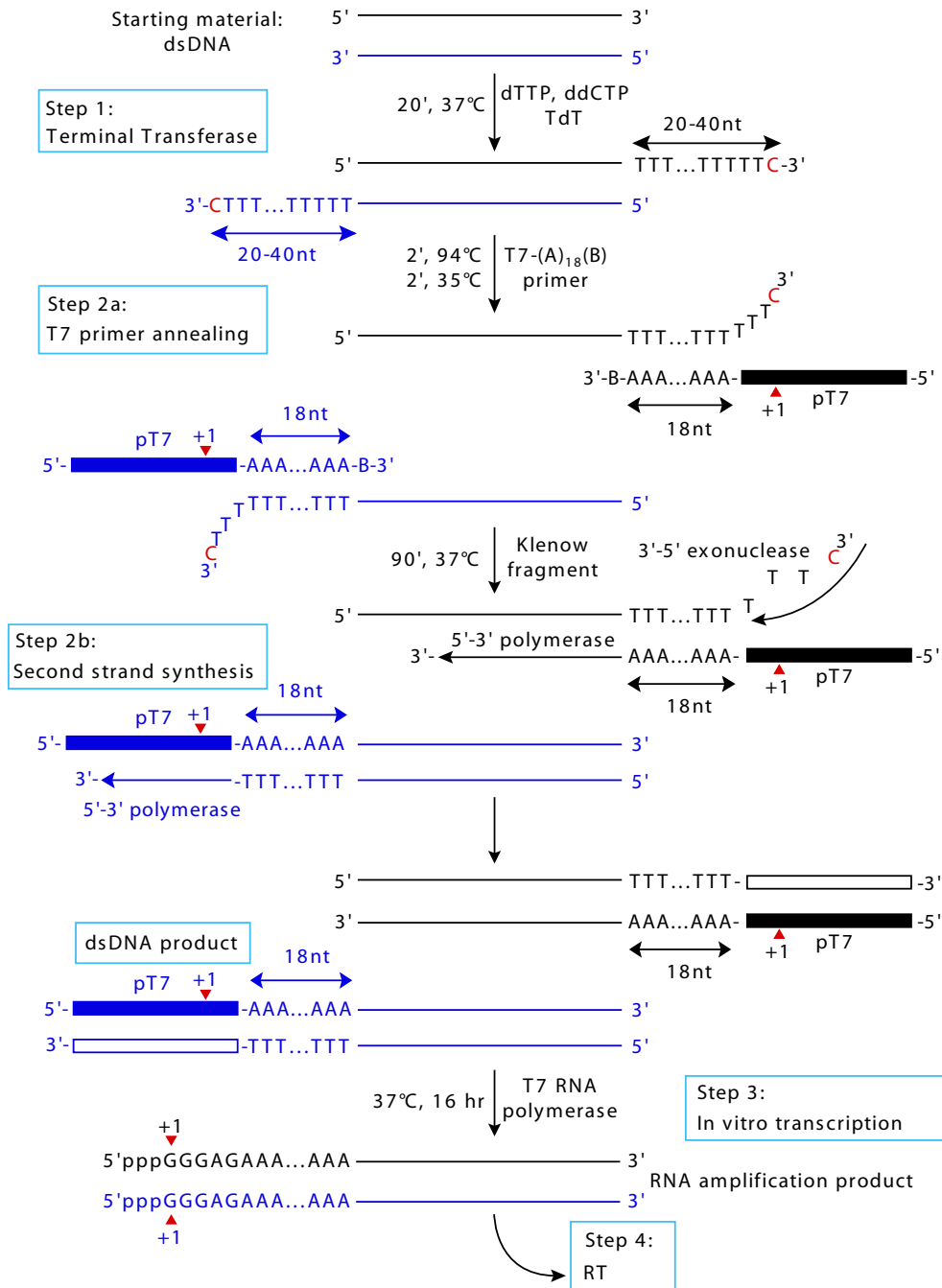
**Figure 1**
**Linear amplification scheme for genomic DNA.** Step 1: Double stranded DNA starting material (shown with one strand in black and one strand in blue) is tailed on the 3' end of each strand to generate a 20–40 bp polyT tail with a terminal dideoxycytidine base. Step 2a: A T7-(A)$_{18}$B anchored primer adaptor is annealed to the polyT tail of each template strand. Step 2b: During second strand synthesis Klenow fragment of DNA Polymerase I removes the excess bases from the tail overhang via its 3'-5' exonuclease activity, and extends from the primer to produce the second strand. This results in two double stranded DNAs identical to the original template, except that each has a T7 promoter at a different end. Step 3: The product of second strand synthesis is used as template in an in vitro transcription reaction. Step 4: To generate DNA probes for microarray analysis, amplified RNA is reverse transcribed.
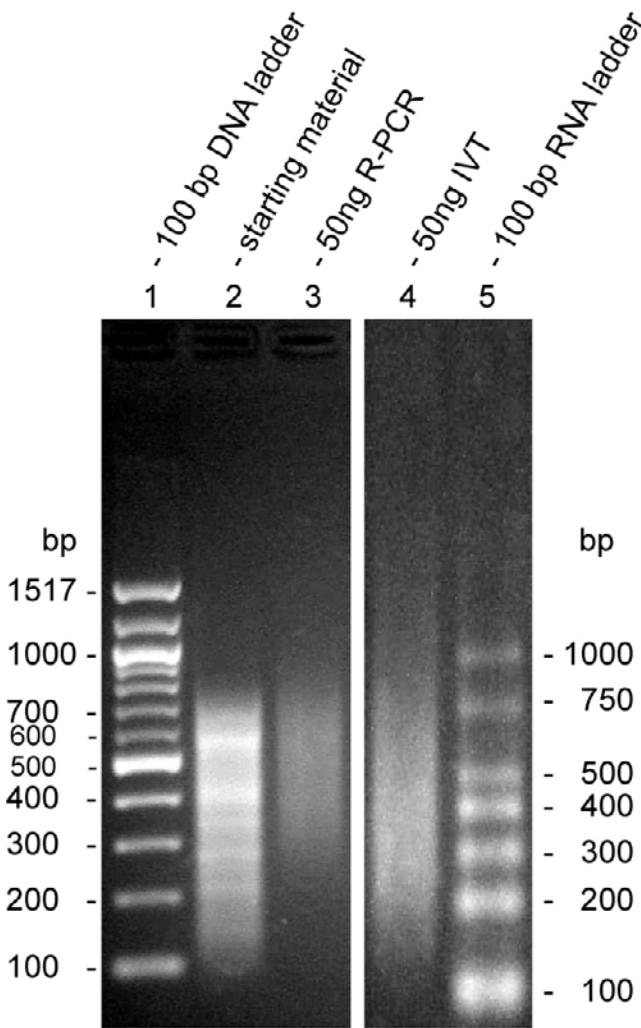
**Figure 2**
**Size distributions for starting material and IVT amplified product.** Lanes 1–3 each contain 250 ng DNA run on a 2% non-denaturing agarose gel. Lanes 4–5 each contain 500 ng RNA run on a 2% denaturing agarose gel. The denaturing gel is necessary to eliminate RNA secondary structure. Lane 1: 100 bp ladder (NEB). Lane 2: starting material (yeast genomic DNA digested with Alu I and previously gel-purified to a size range of 100–700 bp). Lane 3: amplified product generated by R-PCR from 50 ng starting material. Lane 4: amplified RNA product generated by IVT from 50 ng starting material. Lane 5: 100 bp RNA Ladder (Ambion). The R-PCR amplified product appears to significantly under-represent low molecular weight species. The IVT amplified product may slightly under-represent high molecular weight species. For clarity, the denaturing gel image was rescaled to match the ladder of the non-denaturing gel.

with R-PCR which yields up to 15 μg DNA from equivalent starting material. To more precisely evaluate the efficiency and fidelity of the linear amplification protocol, we generated a large pool of starting material by restricting yeast genomic DNA with Alu I (cuts AGCT, leaving blunt ends) and gel-purifying fragments within a size range of 100–700 bps. We subjected 50 ng of the Alu I digest to amplification by IVT and by R-PCR. The IVT yielded almost 80 μg RNA, while the R-PCR yielded 12 μg DNA (Table 2). We compared the Alu I digest with the amplified products by gel-electrophoresis (Fig 2). The IVT product retains essentially the same size distribution as the starting material. In contrast, the R-PCR product appears to under-represent lower molecular weight species.

### Limiting primer input improves yield
While analyzing the size distribution of the IVT product we noted a low molecular weight species that roughly corresponds in size to the primer adapter. IVT protocols are reportedly sensitive to excess starting concentrations of the T7 primer. Baugh and colleagues were able to improve yields when amplifying mRNA by decreasing the concentration of T7 primer [16]. When we limited the mass of T7 primer adapter to approximately five times that of the starting material with a lower limit of 50 ng (by decreasing the second strand synthesis reaction volume), this low molecular weight species disappeared, without decreasing overall yields (data not shown). In this way, we were able to efficiently amplify just 2.5 ng from the Alu I digest with a yield of 10.3 μg (Table 2).

### Evaluation of linear amplification by hybridization
We went on to evaluate the fidelity of the linear amplification protocol by microarray analysis. Restriction by Rsa I (cuts GTAC, leaving blunt ends) and gel-purification were used to generate a second population of DNA. Since Alu I and Rsa I cut at different sites, the resulting digests contain different distributions of DNA species. 50 ng of each pool were amplified by IVT. DNA probes, generated by reverse transcribing the amplified RNA, were fluorescently labeled (Cy5 for Alu I, Cy3 for Rsa I), pooled, and hybridized to microarrays containing the yeast open reading frames. After hybridization, the microarrays were washed and scanned, and Cy5/Cy3 ratios calculated. We also hybridized Alu I and Rsa I probes generated by R-PCR from 50 ng starting material, as well as probes generated by direct labeling of 1 μg starting material.

First, we assessed the reproducibility of the IVT protocol. Three Alu I / Rsa I datasets were independently generated using the IVT protocol. Each dataset contains 4,481 ratios that each reflect the relative abundance of DNA corresponding to a specific array feature (yeast ORF) in the two digests. We found the IVT protocol to be highly reproducible: correlation coefficients calculated between the

**Table 1: Mass yields generated by IVT from varying amounts of restricted DNA.**

| Input (ng) | Yield (ug) | n-fold amp* |
|---|---|---|
| 50.0 | 78.7 | 1574 |
| 25.0 | 69.8 | 2790 |
| 10.0 | 54.2 | 5420 |
| 5.0 | 25.8 | 5163 |
| 2.5 | 10.3 | 4114 |
| 50.0 PCR | 15 | 300 |

*Yield and fold amplification may be overestimated for low starting concentration due to non-specific or template independent products (see text). The yield for the PCR amplification could be improved by introducing more cycles. However, this would likely increase bias.

replicate datasets averaged 0.98. This reproducibility is maintained even when only 5 ng starting material are used (correlation of 0.97). The direct labeling and R-PCR protocols are also highly reproducible, with mean correlations of 0.93 and 0.91, respectively (Fig 3A).

Next, we assessed the fidelity of amplification by comparing the IVT and R-PCR datasets with the direct labeling dataset. We assumed the direct labeling dataset to most accurately represent the true distribution of species in the Alu I and Rsa I digests, since no amplification was required to obtain these data. The three datasets collected for each protocol were averaged and correlation coefficients calculated between the resulting composite datasets. The correlation between the IVT dataset and the direct labeling dataset is 0.96 (Fig 3A). This value increases to 0.98 when features with ratios close to 1 are excluded from the calculation. The correlation between the R-PCR dataset and the direct labeling dataset is 0.68 (Fig 3A). It rises to 0.80 when features with ratios close to 1 are excluded. Correspondingly, the highest ranking features in the direct labeling dataset overlap extensively with the highest ranking features in the IVT dataset and, to a lesser extent, with the highest ranking features in the R-PCR dataset (Fig 3B).

The high correlation between the IVT and direct labeling datasets is evident in the scatter plot shown in Fig 4. An analogous plot portrays the lower correlation between the R-PCR and direct labeling datasets. These plots also illustrate another difference between the amplification protocols: an examination of best fit lines calculated for each of the plots suggests that dynamic range is increased by IVT amplification (slope > 1), but decreased by R-PCR amplification (slope < 1). To quantify dynamic range, we determined the ratio values for data points two standard deviations above and below the mean, and calculated the range between these values by dividing the high ratio by the low ratio. For the direct labeling dataset, the high and low ratios are 3.1 and 0.31, respectively, for a range of 9.8.

The range for the IVT dataset is 17.4, while that for the R-PCR dataset is 5.8 (Table 2).

Hierarchical clustering, shown in Fig 5, was used as an additional means to identify differences between the amplification protocols [20]. Replicates for a given method cluster closely, confirming the high reproducibility observed for each method. Consistent with the correlation analysis, the individual replicates for the IVT and direct labeling protocol cluster together, while the R-PCR replicates cluster in a separate arm of the dendrogram. Most of the 4,481 features shown in the cluster diagram exhibit consistent coloring across the diagram, indicating that the direction of enrichment is consistent for all replicates. However, the cluster diagram highlights two groups of genes (purple stripes) whose ratios in the R-PCR dataset are inconsistent with their ratios in the direct labeling and IVT datasets.

**Discussion**

We have developed a T7-based linear amplification protocol for the amplification of genomic DNA. The protocol was specifically designed to amplify DNA isolated by ChIP, for use in genome mapping protocols. However, it may be useful for other studies that require amplification of small quantities of DNA in a sequence independent manner (e.g., DNA array comparative genomic hybridization [21]). Although T7-based approaches for amplification of mRNA have been described [14–17], these rely on the 3' polyA tails for priming and incorporation of the T7 promoter. Fragmented genomic DNA, such as that generated by ChIP, does not contain such a universal sequence. Our protocol overcomes this obstacle by adding 3' polyT tails to the ends of the DNA fragments for use in priming much like mRNA polyA tails. Efficient tailing by terminal transferase requires the template DNA to be free of 3' phosphate groups, and is highly dependent on whether this template contains blunt- or overhanging-ends. For DNA fragments generated by ChIP (or blunt-end restriction) these limitations are effectively overcome by prior
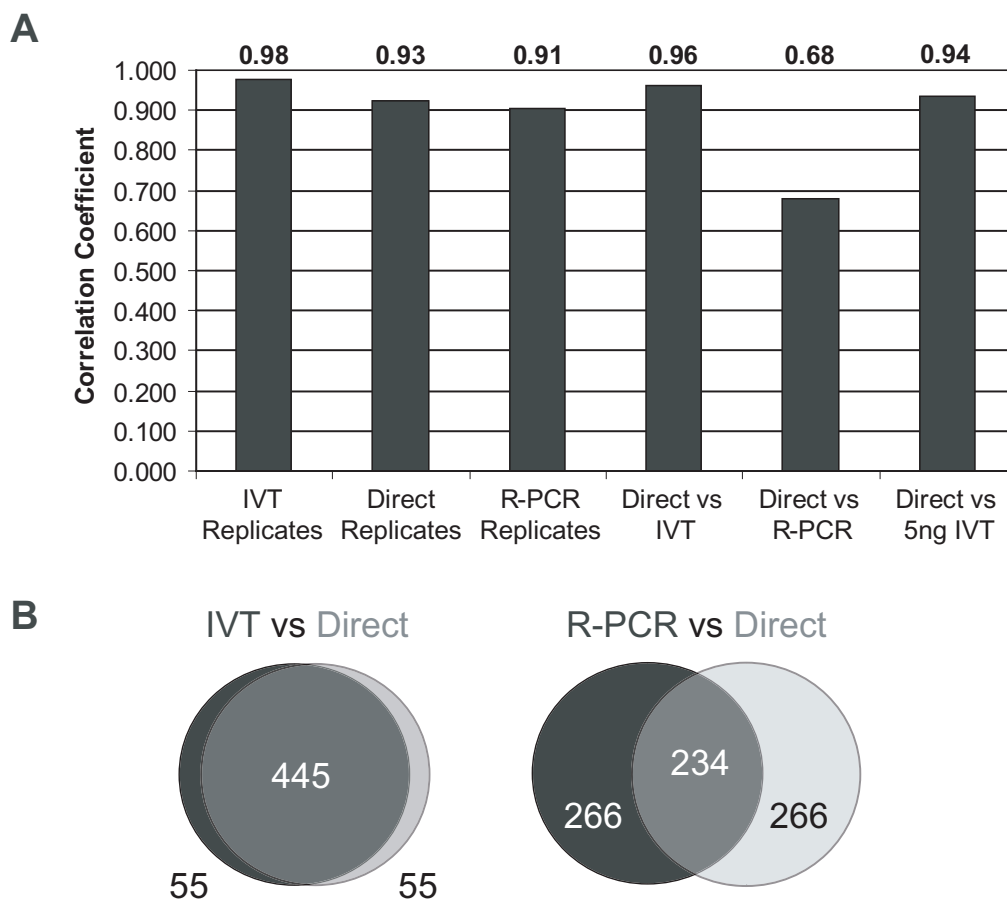
**Figure 3**
**Comparisons of microarray data collected using direct labeling, IVT or R-PCR methods.** (A) Bar graph showing correlations between replicates collected using the same protocol, and between the averaged datasets determined using different protocols. (B) Venn diagrams showing overlap between sets of features with the highest Cy5/Cy3 ratios.

phosphatase treatment, by using recombinant terminal transferase enzyme, and by spiking a small quantity of terminating nucleotide into the tailing reaction.

We compared our IVT protocol to a commonly used PCR protocol that uses a partially degenerate primer to amplify genomic DNA in a sequence independent manner (R-PCR). Initial analyses reveal specific advantages of the IVT approach. First, yields from IVT are significantly higher

than those obtained using R-PCR (Table 1), with a fold-amplification similar to that obtained in mRNA linear amplification protocols [16,17]. Since 2 µg of amplified RNA are sufficient to generate probe for a microarray experiment, RNA from a single IVT amplification can be used for several analyses. A second advantage is that the IVT protocol maintains the size distribution of the starting material more effectively than R-PCR (Fig 2).
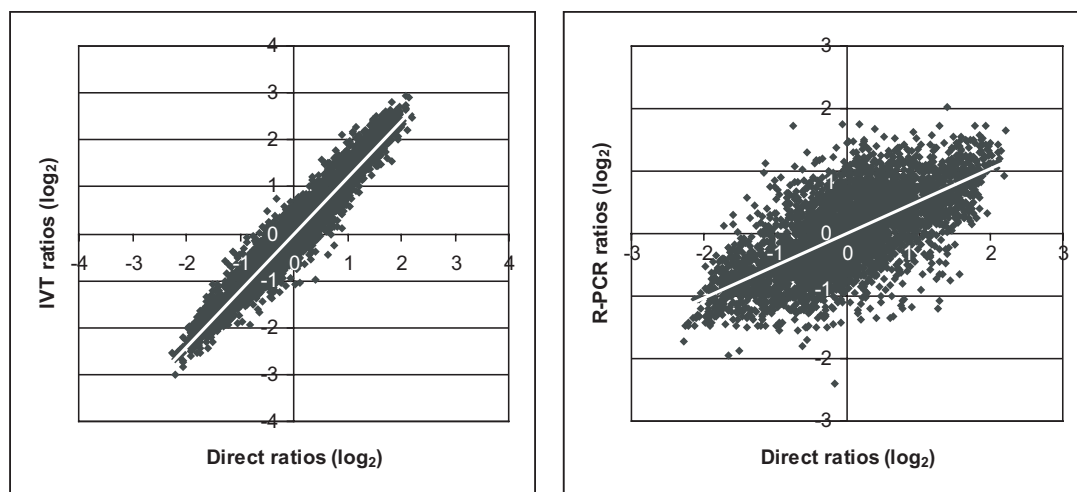
**Figure 4**
**Scatter plots of hybridization ratios.** IVT ratios (A) or R-PCR ratios (B) are plotted against direct labeling (unamplified) ratios. The tight distribution of points along the fitted line in (A) illustrates the high fidelity of the IVT amplification.

To more precisely characterize amplification bias we developed a test system involving the microarray-based analysis of two distinct populations of genomic DNA. These populations were generated by cutting genomic DNA from yeast with a different restriction enzyme and size-purifying the resulting digest. The resulting pools, an Alu I digest and a Rsa I digest, contain microgram quantities of genomic DNA, sufficient for several amplification trials. Importantly, these quantities are also sufficient for direct microarray analysis by direct labeling (this is not the case for the small quantities of DNA isolated by ChIP). Although this 'direct' analysis does require a single round of synthesis to incorporate dye labels, no amplification is involved. Hence, we used direct labeling as a 'gold standard' to validate the IVT and R-PCR methods.

We carried out a series of hybridization experiments using the different protocols to compare the Alu I and Rsa I digests. All three protocols were highly reproducible in our hands. However, we found the direct labeling data to be significantly more concordant with the IVT data than with the R-PCR data. This higher concordance is evident as higher correlations between the datasets, and greater over-

lap among high ranking features (Fig 4). Consistent with these findings, when the datasets are subjected to hierarchical clustering, the direct labeling and IVT datasets cosegregate (Fig 5). This clustering analysis also reveals discrete subsets of features whose ratios in the R-PCR dataset are inconsistent with their ratios in the direct labeling and IVT datasets. In addition to portraying species representation accurately, the IVT dataset displays a significantly expanded dynamic range. Taken together, these observations suggest that IVT has considerable advantages over R-PCR.

However, our finding that IVT is more faithful than R-PCR comes with its own caveats, and does not necessarily imply that R-PCR is an unreliable approach. First of all, it should be noted that the restriction digest analysis is an imperfect surrogate for ChIP experiments: the random fragmentation of DNA in ChIP, and the systematic cutting in the restriction analyses, each place different constraints on an amplification method. Many of the discordant ratios in the R-PCR dataset likely result from a combination of the non-random fragmentation in the Alu I / Rsa I experiment and the ineffectiveness of R-PCR in amplifying
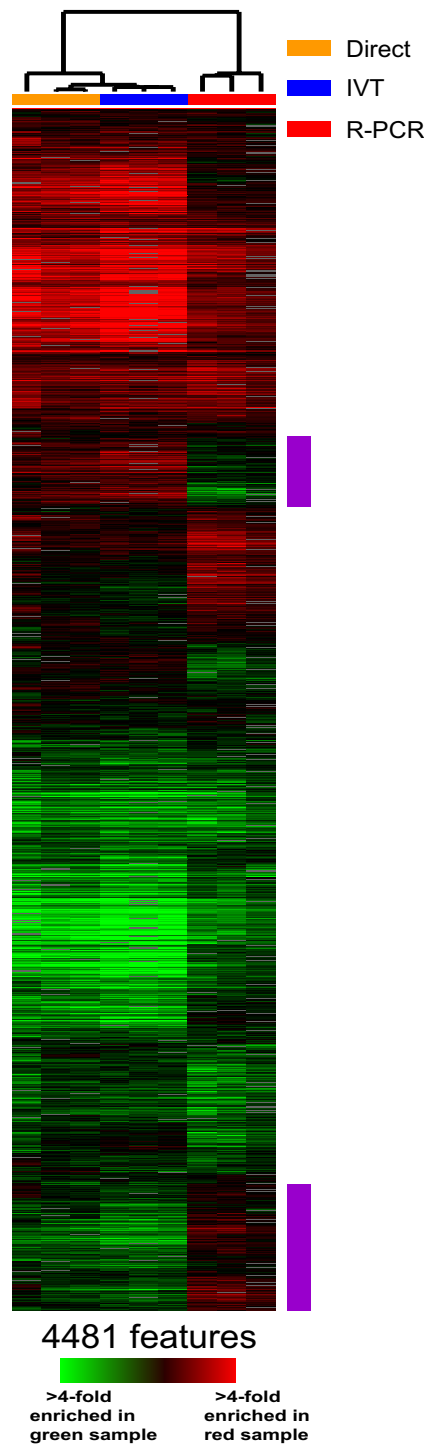
**Figure 5**
**Hierarchical clustering of replicate datasets generated by direct labeling, IVT and R-PCR.** Each thin bar represents a single datapoint. Red bars correspond to enrichment in the Cy5-labeled Alu I probe, while green bars correspond to enrichment in the Cy3-labeled Rsa I probe. The dendrograms (top) indicate clustering relationships among the sample replicates. The lengths of the branches represent the degree of similarity between the samples (shorter indicates higher similarity). Purple stripes to the right of the diagram highlight discordant areas (log ratios with opposite signs) in the R-PCR replicates relative to the direct labeling and IVT samples.

**Table 2: Dynamic range for direct labeling, R-PCR and IVT datasets.** Relative dynamic range was determined by dividing the ratio value 2-standard deviations above the mean by the ratio value 2-standard deviations below the mean. Dynamic range is compressed by R-PCR. In contrast, the IVT protocol actually increases dynamic range, relative to direct labeling.

|  | Direct-labeled | R-PCR | IVT |
|---|---|---|---|
| 2 Stdev Low | 0.31 | 0.41 | 0.23 |
| 2 Stdev High | 3.1 | 2.4 | 4.1 |
| 2 Stdev Range | 9.8 | 5.7 | 17.4 |

low molecular weight species. When these discordant genes are removed, the correlation between the R-PCR and direct labeling datasets increases to 0.77. If the analysis is further limited to features with ratios reflective of a 1.5-fold or greater intensity difference the correlation is 0.88 (compare with correlation of 0.98 for IVT versus direct). Most mapping experiments are primarily interested in these enriched or depleted features. Our initial experiences analyzing ChIP DNA by IVT suggest that results obtained using either amplification method are compatible, with overall correlations around 0.80 (C.L.L., S.L.S., B.E.B., unpublished). These analyses suggest that IVT and R-PCR are both valid approaches for microarray-based genome mapping. However, we expect that the greatly increased dynamic range and improved fidelity obtained through IVT will facilitate the identification of additional biological trends within these data.

## Conclusion

In conclusion, we have presented a novel linear amplification protocol for genomic DNA. We find that this protocol generates higher yields and better maintains the size distribution of the starting material than an alternative R-PCR approach. Extensive microarray-based analyses suggest that improvements in yield, dynamic range and fidelity render linear amplification a better option for genome mapping studies.

## Methods

### Preparation of genomic DNA from yeast

Genomic DNA used in amplification protocols was obtained either by ChIP or by restriction digests. ChIP DNA was fragmented by sonication and isolated using antibody against di-methyl-H3 K4 as described previously [6]. Restricted genomic DNA was prepared as follows: yeast genomic DNA isolated by bead lysis, phenol/chloroform extraction, and ethanol precipitation, was restricted either with Alu I or with Rsa I (New England BioLabs (NEB)). Digested products underwent electrophoresis on a 2% agarose gel. Restriction fragments in the 100–700 bp size range were excised from the gel and purified using the QIAquick Gel Extraction Kit (Qiagen).

### CIP treatment

Calf intestinal alkaline phosphatase (CIP) (NEB) was used to remove 3' phosphate groups from DNA samples prior to IVT. Up to 500 ng DNA were incubated with 2.5 U enzyme in a 10 μL volume with the supplied buffer at 37°C for 1 hour. The reaction was cleaned up with the MinElute Reaction Cleanup Kit (Qiagen) per manufacturer instructions except that the elution volume was increased to 20 μL.

### Poly-dT tailing with terminal transferase

PolyT tails were generated using terminal transferase (TdT) as follows. Up to 50 ng of CIP-treated template DNA were incubated for 20 minutes at 37°C in a 10 μL solution containing 20 U TdT (NEB), 0.2 M potassium cacodylate, 25 mM Tris-HCl pH 6.6, 0.25 mg/ml BSA, 0.75 mM $CoCl_2$, 4.6 μM dTTP and 0.4 μM ddCTP. The reaction was halted by the addition of 2 μL of 0.5 M EDTA pH 8.0, and product isolated with the MinElute Reaction Cleanup Kit (Qiagen), increasing the elution volume to 20 μL.

### Second strand synthesis and incorporation of T7 promoter

Second strand synthesis and incorporation of the T7 promoter sequence was carried out as follows: the 20 μL tailing reaction product was mixed with 0.6 μL of 25 μM T7-$A_{18}$B primer (5'-GCATTAGCGGCCGCGAAATTAATACGACTCACTATAGGGAG(A)$_{18}$ [B], where B refers to C, G or T), 5 μL 10X EcoPol buffer (100 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$, 75 mM dTT), 2 μL 5.0 mM dNTPs, and 20.4 μL nuclease-free water. In experiments with 10–50 ng starting material, the end primer concentration was kept at 300 nM, while the reaction volume was scaled down to maintain an end concentration of 1 ng/ul starting material. For starting amounts less than 10 ng, the volume was kept at 10 μL. If necessary, volume reduction of the eluate from the TdT tailing was performed in a vacuum centrifuge on medium heat. Samples were incubated at 94°C for 2 minutes to denature, ramped down at -1 C°/sec to 35°C, held at 35°C for 2 minutes to anneal, ramped down at -0.5 C°/sec to 25°C and held while Klenow enzyme was added (NEB) to an end concentration of 0.2 U/μL. The sample was then incubated at 37°C for 90 minutes for extension. The reaction was halted by addition of

5 µL 0.5 M EDTA pH 8.0 and product isolated with the MinElute Reaction Cleanup Kit (Qiagen), increasing the elution volume to 20 µL.

### *In vitro transcription*
Prior to in vitro transcription, samples were concentrated in a vacuum centrifuge at medium heat to 8 µL volume. The in vitro transcription was performed with the T7 Megascript Kit (Ambion) per manufacturer's instructions, except that the 37°C incubation was increased to 16 hours. The samples were purified with the RNeasy Mini Kit (Qiagen) per manufacturer's RNA cleanup protocol, except with an additional 500 µL wash with buffer RPE. RNA was quantified by absorbance at 260 nm, and visualized on a denaturing 1.25X MOPS-EDTA-Sodium Acetate gel.

### *Reverse transcription and labeling*
For each sample to be analyzed by microarray, 4 µg amplified RNA were primed with 5 µg random hexamers and 5 µg oligo dT and reverse-transcribed, incorporating aa-dUTP, as described at http://www.microarrays.org. Note: the presence/absence of the oligo dT primer does not affect the RT and labeling efficiency. RT was halted with 10 µL 0.5 M EDTA, and RNA hydrolyzed by the addition of 10 µL 1 N NaOH and incubation at 65°C for 15 minutes. Reaction cleanup and labeling with monofunctional reactive Cy5 (Alu I digest) or Cy3 (Rsa I digest) dye was carried out as described in [22] and at http://www.microarrays.org.

### *R-PCR*
R-PCR samples were generated from 50 ng restricted DNA using the Round A/ Round B protocol described in [9,12]. Briefly, Sequenase (Amersham) was used for two cycles of synthesis using a degenerate primer (5'-GTTTCCCAGT-CACGATCNNNNNNNNN-3' ('round A'). The resulting end-tagged DNAs were subjected to 30 cycles of PCR using the following primer: 5'-GTTTCCCAGTCACGATC-3'. Amino-allyl dUTP was incorporated during the PCR step as described in [22], and at http://www.microarrays.org. Amplified DNA was purified with a QIAquick PCR purification kit (Qiagen). For microarray analysis, 7 µg of each probe were fluorescently labeled by incubation with monofunctional reactive Cy5 (Alu I digest) or Cy3 (Rsa I digest) dye as described [22].

### *Direct labeling*
Unamplified DNA samples for labeling were generated from 1 µg restricted DNA using Klenow fragment and random primers, and incorporating amino-allyl dUTP, as described at http://www.microarrays.org.

### *Microarrays, hybridization, and scanning*
Yeast ORFs were amplified from a set of 6,218 plasmids using universal primers as described [1]. This set of ORFs was printed on separate slides, hydrated, and snap-dried as described [1]. Mixed Cy5/Cy3-labeled probes were hybridized to microarrays for 12–15 h at 60°C. After hybridization, microarrays were washed as described [1] and scanned with a GenePix 4000B scanner with GENEPIX PRO 4.0 software (Axon Instruments). Microarray data are available at http://www.schreiber.chem.harvard.edu

### *Microarray data processing*
The Cy5/Cy3 ratio for each array element was calculated using GENEPIX PRO 4.0, $\log_2$ transformed, and normalized using the default computed normalization method used by the Stanford Microarray Database [23]. We applied a high stringency data selection, keeping only microarray elements for which at least 80% of the measurements within a set of experiments had fluorescence intensity in both channels at least 3-fold over background intensity. 4,481 array elements passed these selection criteria.

### *Comparison of amplification methods*
IVT, R-PCR, and direct labeling procedures and hybridizations were each carried out in triplicate. To assess reproducibility of each method, the correlation coefficients were calculated for all possible combinations of pairs within each set of replicates, and then averaged. Next, a single averaged dataset was generated for each method by determining the geometric mean for each array element from a given set of replicates. To assess the degree of similarity among the different methods, correlation coefficients were calculated in pair-wise comparisons of these averaged datasets. An additional correlation coefficient was calculated between averaged datasets using only those features whose ratios in the direct labeling dataset reflect at least a 1.5-fold change, either up or down.

Hierarchical clustering of the non-averaged replicate datasets was carried out using the Cluster program, with average linkage using Pearson correlation as the similarity metric [20]. Clusters were visualized using TreeView [20]. To assess the degree of overlap among the different methods, the number of features among the top 500 $\log_2$ Cy5/Cy3 ratios was determined for each pair-wise comparison and plotted in Venn Diagrams, with the overlap region scaled to the number of features. To assess the dynamic range characteristics of the three methods, the range of the data for each method was taken as an interval extending 2 standard deviations on either side of the mean across the entire array.

## Author's contributions

C.L.L. implemented the linear amplification protocol, carried out the comparative analyses, and did the bioinformatics. S.L.S. coordinated the project and provided support. B.B. conceived of the approach and guided the study. C.L.L. and B.B. drafted the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

## References

1. DeRisi JL, Iyer VR and Brown PO **Exploring the metabolic and genetic control of gene expression on a genomic scale** *Science* 1997, **278:**680-6
2. Wodicka L, Dong H, Mittmann M, Ho MH and Lockhart DJ **Genome-wide expression monitoring in Saccharomyces cerevisiae** *Nat Biotechnol* 1997, **15:**1359-67
3. Ramaswamy S and Golub TR **DNA microarrays in clinical oncology** *J Clin Oncol* 2002, **20:**1932-41
4. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N and Kanin E **Genome-wide location and function of DNA binding proteins** *Science* 2000, **290:**2306-9
5. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M and Brown PO **Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF** *Nature* 2001, **409:**533-8
6. Bernstein BE, Humphrey EL, Erlich RL, Schneider R, Bouman P, Liu JS, Kouzarides T and Schreiber SL **Methylation of histone H3 Lys 4 in coding regions of active genes** *Proc Natl Acad Sci U S A* 2002, **99:**8695-700
7. Robyr D, Suka Y, Xenarios I, Kurdistani SK, Wang A, Suka N and Grunstein M **Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases** *Cell* 2002, **109:**437-46
8. Bohlander SK, Espinosa R 3rd, Le Beau MM, Rowley JD and Diaz MO **A method for the rapid sequence-independent amplification of microdissected chromosomal material** *Genomics* 1992, **13:**1322-4
9. Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO and Petes TD **Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae** *Proc Natl Acad Sci U S A* 2000, **97:**11383-90
10. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM and Simon I **Transcriptional regulatory networks in Saccharomyces cerevisiae** *Science* 2002, **298:**799-804
11. Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA and Dynlacht BD **E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints** *Genes Dev* 2002, **16:**245-56
12. Lieb JD, Liu X, Botstein D and Brown PO **Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association** *Nat Genet* 2001, **28:**327-34
13. Horak CE, Mahajan MC, Luscombe NM, Gerstein M, Weissman SM and Snyder M **GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis** *Proc Natl Acad Sci U S A* 2002, **99:**2924-9
14. Phillips J and Eberwine JH **Antisense RNA Amplification: A Linear Amplification Method for Analyzing the mRNA Population from Single Living Cells** *Methods* 1996, **10:**283-8
15. Wang E, Miller LD, Ohnmacht GA, Liu ET and Marincola FM **High-fidelity mRNA amplification for gene profiling** *Nat Biotechnol* 2000, **18:**457-9
16. Baugh LR, Hill AA, Brown EL and Hunter CP **Quantitative analysis of mRNA amplification by in vitro transcription** *Nucleic Acids Res* 2001, **29:**E29
17. Zhao H, Hastie T, Whitfield ML, Borresen-Dale AL and Jeffrey SS **Optimization and evaluation of T7 based RNA linear amplification protocols for cDNA microarray analysis** *BMC Genomics* 2002, **3:**31
18. Iscove NN, Barbara M, Gu M, Gibson M, Modi C and Winegarden N **Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA** *Nat Biotechnol* 2002, **20:**940-3
19. Shaw CA **Theoretical consideration of amplification strategies** *Neurochem Res* 2002, **27:**1123-31
20. Eisen MB, Spellman PT, Brown PO and Botstein D **Cluster analysis and display of genome-wide expression patterns** *Proc Natl Acad Sci U S A* 1998, **95:**14863-8
21. Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, Law S, Myambo K, Palmer J, Ylstra B, Yue JP, Gray JW, Jain AN, Pinkel D and Albertson DG **Assembly of microarrays for genome-wide measurement of DNA copy number** *Nat Genet* 2001, **29:**263-4
22. Carroll AS, Bishop AC, DeRisi JL, Shokat KM and O'Shea EK **Chemical inhibition of the Pho85 cyclin-dependent kinase reveals a role in the environmental stress response** *Proc Natl Acad Sci U S A* 2001, **98:**12578-83
23. Gollub J, Ball CA, Binkley G, Demeter J, Finkelstein DB, Hebert JM, Hernandez-Boussard T, Jin H, Kaloper M and Matese JC **The Stanford Microarray Database: data access and quality assessment tools** *Nucleic Acids Res* 2003, **31:**94-6