

# Assessing Differential Expression Measurements by Highly Parallel Pyrosequencing and DNA Microarrays: A Comparative Study

Joaquín Ariño,<sup>1\*</sup> Antonio Casamayor,<sup>1</sup> Julián Perez Pérez,<sup>2,3</sup> Laia Pedrola,<sup>3</sup> Miguel Álvarez-Tejado,<sup>4</sup> Martina Marbà,<sup>5,6</sup> Javier Santoyo,<sup>5,7</sup> and Joaquín Dopazo<sup>5,6,7\*</sup>

## Abstract

To explore the feasibility of pyrosequencing for quantitative differential gene expression analysis we have performed a comparative study of the results of the sequencing experiments to those obtained by a conventional DNA microarray platform. A conclusion from our analysis is that, over a threshold of 35 normalized reads per gene, the measurements of gene expression display a good correlation with the references. The observed concordance between pyrosequencing and DNA microarray platforms beyond the threshold was of 0.8, measured as a Pearson's correlation coefficient. In differential gene expression the initial aim is the quantification of the differences among transcripts when comparing experimental conditions. Thus, even in a scenario of low coverage the concordance in the measurements is quite acceptable. On the other hand, the comparatively longer read size obtained by pyrosequencing allows detecting unconventional splicing forms.

## Introduction

RNA SEQUENCING (RNA-seq) is becoming one of the most popular applications of next generation sequencing technologies (Cloonan et al., 2008; Marioni et al., 2008; Mortazavi et al., 2008; Wang et al., 2009). Despite the fact that RNA-seq is a relatively new methodology, it has already been applied to the study of the transcriptional complexities of different organisms, including yeast (Nagalakshmi et al., 2008), mice (Mortazavi et al., 2008), *Arabidopsis* (Eveland et al., 2008), and humans (Sultan et al., 2008). In addition, to provide a quantitative estimation of the mRNA levels, RNA-seq also offers extra information on novel splice junctions, novel transcripts, alternate transcription start sites, rare transcripts, etc., constituting a first-order tool for analysis and discovery in transcriptomics.

Recent articles have compared short-read technologies to microarray references, finding a good correspondence when a high number of reads is available (Bloom et al., 2009; t'Hoen et al., 2008). Here we extend, for the first time, this comparison to long reads to check to what extent the results of differential

gene expression comparisons obtained by pyrosequencing are consistent with those ones obtained with gene expression microarrays. To this end, we have chosen a well-established perturbation model, the exposure of yeast cells to an abrupt increase in external pH. Previous DNA microarray-based analyses of the effect of this perturbation on gene expression remodeling have shown that it results in several hundred of genes being induced or repressed (Lamb and Mitchell, 2003; Lamb et al., 2001; Platara et al., 2006; Ruiz et al., 2008; Serrano et al., 2002, 2006; Viladevall et al., 2004), as a result of the modulation of diverse signaling pathways (for a recent review, see Arino, 2010). In this work we present a comparison of the transcriptional changes detected after short-term (10 min) exposure to high pH stress (pH=8.0) for both wild-type and *snf1* mutant strains. *snf1* encodes a Ser/Thr protein kinase, homolog to the mammalian AMP-activated protein kinase, which is required for regulating transcription of glucose-repressed genes in yeast and becomes activated in response to diverse stress situations (Hedbacker and Carlson, 2008). The biological implications of Snf1 in high pH response

<sup>1</sup>Departament de Bioquímica y Biología Molecular and Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain.

<sup>2</sup>Secugen SL, Madrid, Spain.

<sup>3</sup>Lifesequencing SL, Valencia, Spain.

<sup>4</sup>Roche Applied Science, Sant Cugat del Vallès, Spain.

<sup>5</sup>Department of Bioinformatics and Genomics, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain.

<sup>6</sup>Functional Genomics Node (INB) at CIPF, Valencia, Spain.

<sup>7</sup>CIBER de Enfermedades Raras (CIBERER), Valencia, Spain.

\*These authors should be considered coauthors.

will be described elsewhere (Casamayor, Ruiz, Serrano, Platara, Ferrer-Dalmau, Barreto and Ariño, manuscript in preparation). Our results show that there is a more than reasonable agreement between both technologies if a threshold for the number of reads for gene is used. In addition, despite the low coverage, still uncommon splicing events can be found.

## Material and Methods

### RNA preparation

Cultures (50 mL) of wild-type (BY4741) and the *snf1::kanMX* derivative were grown on YPD until  $OD_{660nm} \approx 0.6-0.8$  and split into two aliquots. One set of aliquots received 16 mM KCl (nonstressed cells) or 16 mM KOH (stressed cells, pH=8.0 approx.). After 10 min cells were collected by rapid filtration through 0.45  $\mu$ m MetriCel membrane filters (Pall Corporation, Port Washington, NY) and dried cells kept at  $-80^{\circ}\text{C}$ . Total RNA was extracted by using the RiboPure™-Yeast kit (Ambion, ref. AM1926), with two steps of elution from the column (50  $\mu$ L each). Samples were treated with DNase to eliminate genomic DNA traces. RNA quality was assessed by denaturing 0.8% agarose gel electrophoresis, and RNA quantification was performed by measuring  $A_{260}$  in a BioPhotometer (Eppendorf, Westbury, NY).

### DNA microarray hybridizations

Eight micrograms of total RNA were employed for cDNA synthesis and labeling, using the indirect labeling kit (CyScribe Post-Labeling kit, GE-Amersham Biosciences, Piscataway, NJ), in conjunction with Cy3-dUTP and Cy5-dUTP fluorescent nucleotides. The cDNA obtained was dried and resuspended in the hybridization buffer. DNA amount and labeling was evaluated with a Nanodrop spectrophotometer (Nanodrop Technologies, Inc., Wilmington, DE). Fluorescently labeled cDNAs were combined and hybridized to yeast genomic microchips constructed in our laboratory by arraying 6014 different PCR-amplified open reading frames from *Saccharomyces cerevisiae* (Alberola et al., 2004; Viladevall et al., 2004).

Prehybridization, hybridization, and washing conditions were essentially as described previously (Hegde et al., 2000).

### Microarray data analysis

Slides were scanned with a ScanArray 4000 apparatus (Packard BioChips Technologies, New York, NY) and the output analyzed using GenePix Pro 6.0 software. For each experiment two independent biological replicates were performed, each by duplicate (dyes were swapped to avoid dye-specific bias) and data were combined (four microarrays/experiment). All the microarray data in a MIAME-compliant format have been deposited in GEO database and will be available under series record GSE25697.

Raw data were normalized using the Limma method (Smyth, 2005) with dye swap, as implemented in the Babelomics package (Medina et al., 2010). Fold changes were calculated using Babelomics as well.

### Preparation of cDNA for sequencing

For first-strand cDNA synthesis the RevertAid™ H Minus First Strand cDNA Synthesis Kit (Fermentas, Hanover, MD) was used. Two independent reactions were carried out (20  $\mu$ g

of total RNA each) for each experimental condition. Priming was achieved by using the degenerate oligonucleotide 5'-GAGCTAGTTCGGAG(T)<sub>16</sub>VN-3' (V stands for any nucleotide but T and N for any nucleotide), which contains a GsuI recognition site. Extension was achieved by incubation with RevertAid™ H Minus M-MuLV Reverse Transcriptase in the presence of dATP, dTTP, dGTP, and 5-methyl-dCTP (0.25 mM each, final concentration) at  $42^{\circ}\text{C}$  for 60 min. Finally, the RNA/DNA hybrid was precipitated with ammonium acetate and ethanol by incubation at  $-80^{\circ}\text{C}$  for 5 min and centrifugation for 20 min at  $4^{\circ}\text{C}$  at maximum speed in a microfuge. Pellets were rinsed with cold 70% ethanol, dried briefly and resuspended in 40  $\mu$ L of nuclease-free water.

Second-strand cDNA synthesis was accomplished using the Second Strand cDNA Synthesis (Fermentas) following the manufacturer's directions, except that the starting first strand material was 40  $\mu$ L and all amounts were doubled accordingly. The synthesis reaction was terminated by addition of 10  $\mu$ L of 0.5 M EDTA, pH 8.0. The volume was adjusted to 200  $\mu$ L with deionized water and the nucleic acid precipitated by addition of 20  $\mu$ L of 3 M sodium acetate, pH 5.2 solution, 1.5  $\mu$ L glycogen (20 g/L, Roche, Indianapolis, IN) and 600  $\mu$ L absolute ethanol. The mixture was kept for 30 min at  $-80^{\circ}\text{C}$ , then centrifuged for 30 min at maximum speed at  $4^{\circ}\text{C}$  in a microfuge. Double-strand cDNA pellets were washed with 70% ethanol and resuspended in 45  $\mu$ L water. At this step both independent reactions were combined and stored at  $-20^{\circ}\text{C}$  until processed. Removal of the polyA tail was carried out by incubating 75  $\mu$ L of each cDNA sample with 8.6  $\mu$ L of  $10\times$  Tango buffer (Fermentas) and 2.5  $\mu$ L of GsuI (5 U/ $\mu$ L) for 5 h at  $30^{\circ}\text{C}$ . At the end of the digestion GsuI was inactivated by incubation for 20 min at  $65^{\circ}\text{C}$ . Samples were stored at  $-80^{\circ}\text{C}$  until use. This procedure yielded cDNA samples with a concentration ranging from 0.87 to 1.31 mg/mL.

### Pyrosequencing

Gene expression was analyzed using 454 pyrosequencing data generated by sequencing of cDNA synthesized from four total RNA samples from two different *S. cerevisiae* strains at the Universitat Autònoma de Barcelona. Raw data (.sff files) from the GS FLX sequencer were provided by Lifesequencing S.L. (Valencia, Spain). This represent half sequencing run of cDNA from each sample, summing up a total of 502,959 reads. Approximately 5  $\mu$ g of the cDNA-preparation in the size range 250–600 bp were used to generate a sstDNA shotgun library according to the protocol supplied by the manufacturer. Briefly, purified cDNA fragments were nebulized and concentrated with AMPure PCR purification beads to remove fragments below 200 bp. These fragments were hybridized to DNA capture beads and each cDNA fragment was individually amplified by emulsion-based clonal amplification polymerase chain reaction (PCR) (emPCR). The DNA capture beads containing amplified DNA were then deposited in individual wells of a PicoTiter plate. After titration, the optimums DNA copies per bead were used for the main sequencing run. After emulsion PCR and subsequent bead recovery, DNA beads were loaded onto one PicoTiterPlate and were subjected to sequencing.

### Sequence data analysis

Four experiments were incorporated in the same run by means of multiplexing using different tag sequences.

TABLE 1. NUMBER OF READS, FILTERING, AND MAPPING OF THE TWO REPLICATES

	Replicate 1				Replicate 2			
	WT	WT_pH	snf1	Snf1_pH	WT	WT_pH	snf1	snf1_pH
Total reads	25,609	43,559	37,296	43,308	28,322	47,040	40,230	47,327
Filtered duplicated reads	4,350	7,106	6,065	6,774	3,003	5,146	4,035	4,442
Non-redundant reads	21,259	36,453	31,231	36,534	25,319	41,894	36,195	42,885
Reads with good quality	20,964	35,740	30,709	35,971	24,326	40,235	34,683	41,412
Number of reads with unique mapping on Ensembl transcripts	7,768	12,450	11,911	13,980	9,119	13,961	14,152	16,188
Number of reads with unique mapping on transcripts as in (Nagalakshmi et al., 2008)	17,525	29,152	25,693	29,004	20,391	32,505	29,111	33,592
Number of reads with no transcript mapping	1,503	2,397	2,398	3,034	1,683	2,987	2,478	3,489
Number of reads with multiple mapping on transcripts	860	1,476	1,146	1,524	942	1,597	1,354	1,580
Number of transcripts with reads	3,564	4,172	4,384	4,770	3,862	4,344	4,552	4,955
Number of undetected transcripts	3,560	2,952	2,740	2,354	3,262	2,780	2,572	2,169

A custom Perl script was made to deconvolute the multiplexing and to sort the sequences into the four different experiments that were carried out. The script looks for the tag sequence on the 5' end of the sequences, prune it, and assign the sequence to the corresponding experiment.

All the remaining reads were mapped to the yeast genome using the BWA-SW program (Li and Durbin, 2010). Reads mapping in more than one genomic location were discarded. The genomic coordinates for the genes were taken from the Ensembl (Flicek et al., 2008). Gene coordinates were further corrected with more recent estimations based on experimental results (Nagalakshmi et al., 2008)

Table 1 shows the original number of reads and the reads at each step of filtering and mapping. Both technical replicates were joined for the subsequent step of quantification.

The quantification of the relative expression of the mRNAs corresponding to the genes is measured as the number of

reads mapping onto it. This value must also be corrected by the gene length. Thus, absolute number of reads mapped are normalized as reads per kilobase per million of bases (RPKM) (Mortazavi et al., 2008).

For each of the four matrices of counts we have applied the RPKM normalization (Mortazavi et al., 2008) and later we have calculated the fold changes ( $\log_2$ -ratios) of WT\_pH vs WT and snf1\_pH vs snf1. In this way sequencing results are comparable to  $\log_2$ -ratios obtained from the microarray platform.

#### Microarray to sequence comparison

The fold change values obtained from the microarrays were compared to the corresponding fold changes obtained from the corresponding sequencing experiments using a simple Pearson's coefficient of correlation.

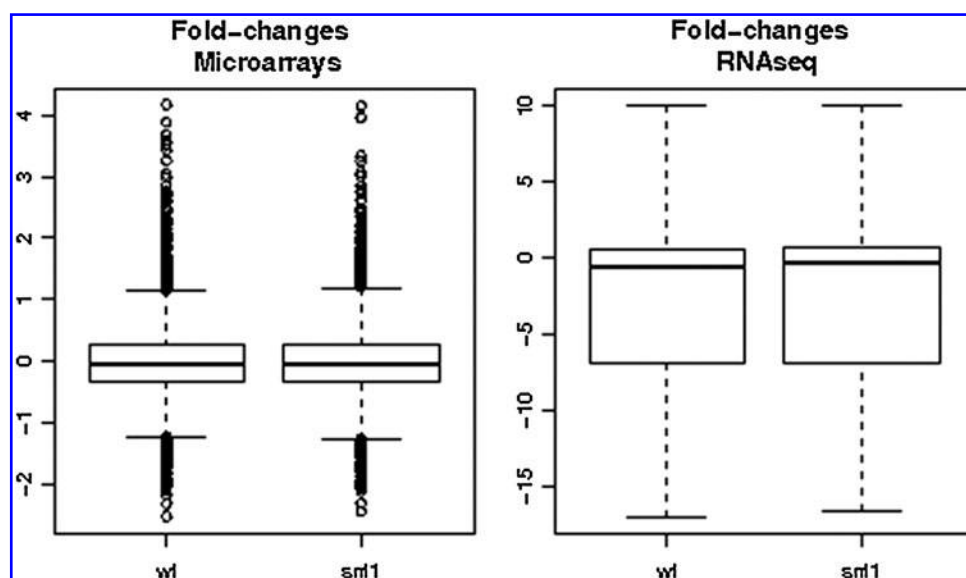


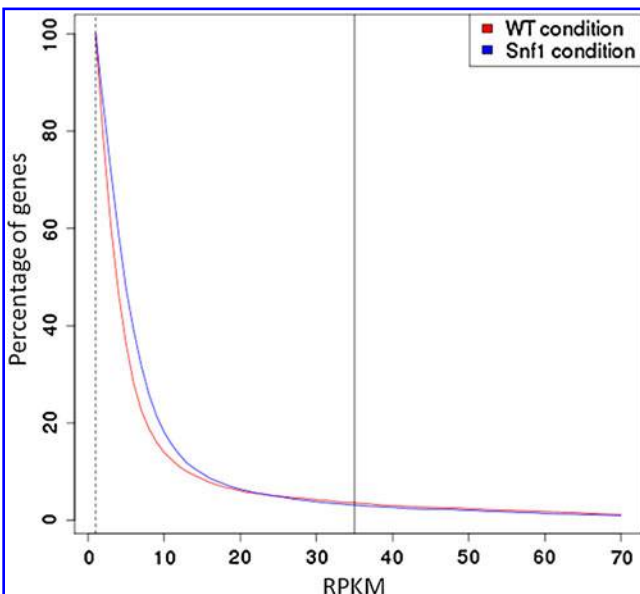
FIG. 1. Boxplots of the comparisons. Comparisons of transcriptional changes detected by DNA microarray and pyrosequencing experiments. The changes in expression detected after exposure of wild-type (WT) and *snf1* cells (*snf1*) to alkaline pH are shown for each type of experiment as boxplots.

## Results and Discussion

### Data generation and preprocessing

Four *S. cerevisiae* cultures were analyzed: (1) a control, wild-type strain (WT), (2) the control strain under alkaline stress (a shift to pH 8.0 for 10 min; WT\_pH), (3) a mutant *snf1* strain (*snf1*), and (4) the *snf1* strain under alkaline stress (*snf1*\_pH). The RNA was extracted and the corresponding cDNA prepared as described in methods. The cDNA corresponding to the four conditions studied was subjected to competitive hybridization in two-color microarrays representing the comparisons WT versus WT\_pH and *snf1* versus *snf1*\_pH (see Methods section). On the other hand, two replicates of the four experimental conditions were sequenced as described in the Methods section. A number of total reads (merging both replicates) between 40,000 and 80,000 were obtained for each experimental condition analyzed. Table 1 shows details on the number of reads obtained and the remaining number of reads after the filtering and mapping steps.

Both, microarray and RNA-seq data were normalized as described in methods. Two comparisons (WT vs. WT\_pH and *snf1* vs. *snf1*\_pH) were carried out and analyzed with the two technologies. Log ratios of the fold change values were obtained in the two comparisons. Figure 1 shows the boxplots of the comparisons. It is apparent that the distribution of values is different when microarrays are compared to RNA-seq results. Although microarrays produced almost symmetrical boxplots, RNA-seq results seem to be slightly biased toward low values of log-ratios. Table 1 shows the process of filtering and mapping the reads onto the transcripts. The number of mapped reads when the transcript definitions provided by Ensembl (Flicek et al., 2008) are used is significantly lower



**FIG. 2.** Percentage of genes that show a normalized number of reads per transcript (RPKM value) represented in the x-axis. The vertical line corresponds to the threshold of RPKM value where the correlation between the microarray experiment and the sequencing experiment is high (see text). In this particular case, the number of genes corresponding to the percentage at the RPKM value is of around 200 (actually 208 genes for the WT cells and 189 genes for the *snf1* cells).

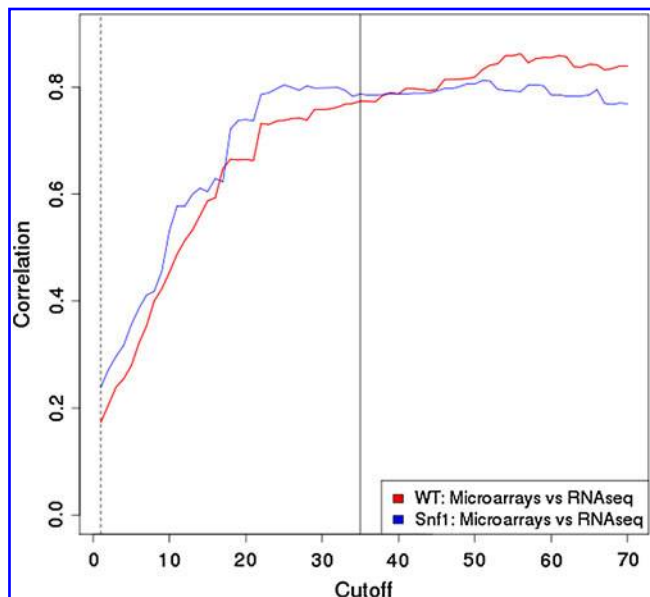
than that obtained when more accurate definitions, based on experimental values (Nagalakshmi et al., 2008), are used. Such more accurate definitions include the 5' and 3' UTR, missing in many genes, which increases the regions for mapping in the transcripts. This explains the corresponding increase in the number of reads mapping in the transcripts.

With small variations among the experiments, over an 80% of the total genes of the *S. cerevisiae* genome is detected as apparently actively transcribing by, at least, one read. This value is very similar to the number of genes detected as transcriptionally active by DNA microarray analysis (80.2 and 73.9% for wild-type and *snf1* cells, respectively). Figure 2 shows the distribution of reads per gene in the gene spectrum of *S. cerevisiae* obtained in the sequencing experiments. Due to the low coverage a substantial number of the genes are represented by only a few reads.

### DNA microarray and pyrosequencing data correlations

Beyond the simple detection, when the aim is the quantification of the relative expression of the mRNAs corresponding to the genes, a minimum number of reads is required for each gene, which must also be corrected by the gene length (Mortazavi et al., 2008).

The correlations between microarrays and RNA-seq data were studied. The most extreme thresholds of coverage were used given that this factor is known to have a strong effect on the reliability of the measurements. Figure 3 shows how correlation, computed as the Pearson correlation coefficient, changes as a function of the threshold of normalized counts (in RPKM) for including a gene in the comparison. The higher the coverage of the genes considered, the better the correlation until a plateau is reached. Figure 3 documents how the trends obtained in the two genetic backgrounds studied (WT and *snf1*) are quite similar and that the coefficient of correlation



**FIG. 3.** Pearson's correlation coefficient computed between the microarray and the pyrosequencing comparisons as a function of the threshold in normalized number of reads per transcript (RPKM). Red line corresponds to the WT comparison and blue line to the *Snf1* comparison.

stabilizes approximately at a value of 0.8 over 35 RPKM per transcript. Higher coverage did not result in more correlated measurements, suggesting that the limit of concordance between both technologies is around this value.

Obviously, the more stringent is the threshold the fewer genes can be used in the comparisons. Figure 2 shows the number of genes remaining after the threshold. Some recent reports suggest that, in mammalian genomes, several million reads would be required to obtain accurate quantification of >95% of expressed transcripts (Blencowe et al., 2009), but as yet there has not been a detailed analysis on how sequencing coverage affects differential expression calls yet (Oshlack et al., 2010). Knowledge on the relationship between sequencing depth, feature detection, and differential expression is needed for experimental design purposes and for understanding the characteristics of the analysis results.

#### Comparisons of differentially expressed genes

In the optimal threshold of 35 RPKM the trends observed for the microarrays and the RNA-seq were quite similar. The concordance in the number of genes differentially expressed

TABLE 2. INTERSECTIONS BETWEEN THE 20, 50, AND 100 GENES MORE DIFFERENTIALLY EXPRESSED BY RNA-SEQ AND MICROARRAY

	20	50	100
WT microarray/ WT RNAseq	12/20 (60%)	16/50 (32%)	48/100 (48%)
<i>snf1</i> microarray/ <i>snf1</i> RNAseq	12/20 (60%)	22/50 (44%)	67/100 (67%)

when both platforms are compared is reasonably high, reaching 60% among the 20 most differentially expressed genes (see Table 2). This is a high concordance value, even for comparisons between platforms based on the same technology (Ioannidis et al., 2009; Shi et al., 2006).

Figure 4 shows all the possible pairwise comparisons at the optimal threshold. RNA-seq produces more extreme values in the most differentially expressed genes given that the scales of measurement in the hybridization and in the sequencing counts are different. However, overall the results are quite

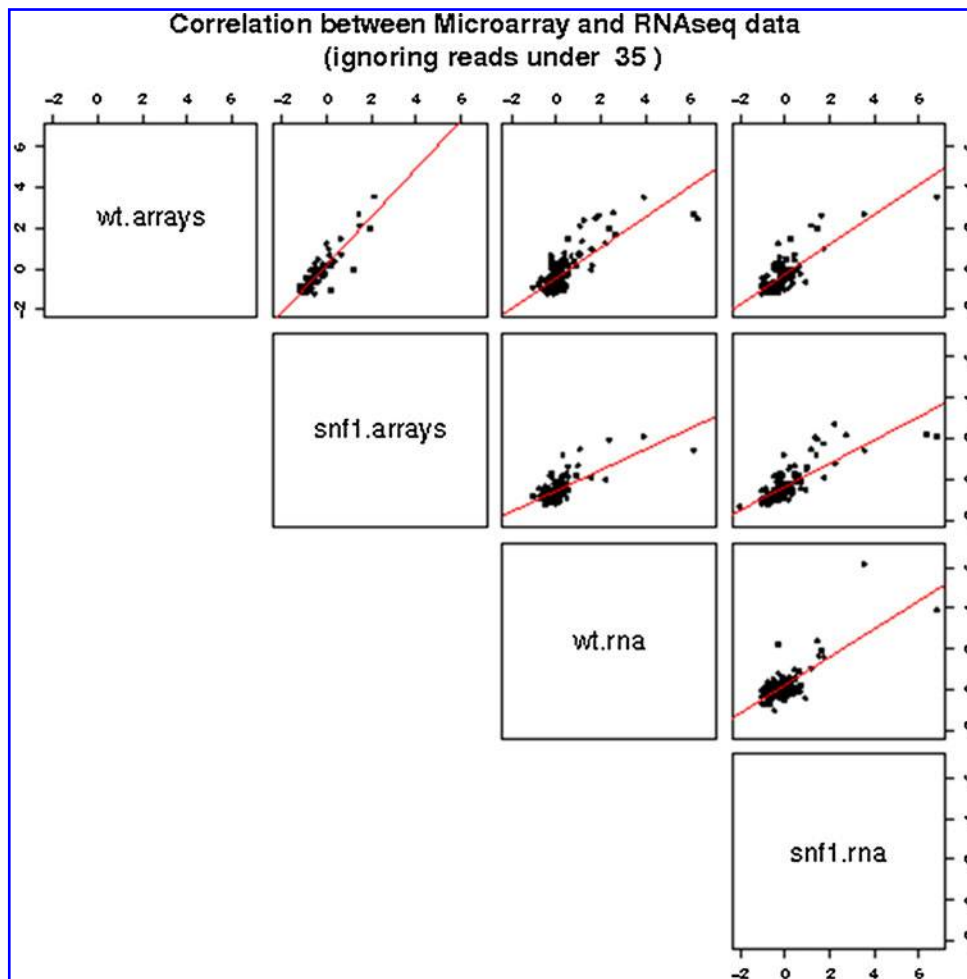


FIG. 4. Comparisons of log-ratio values between microarrays and RNAseq filtering only genes over the optimal threshold of 35 RPKM. The diagonal contains the name of the variable that is represented in the x-axis for the row and in the y-axis for the column, respectively. Thus, the plot in the first row and the second column corresponds to the comparison between the arrays, the plot in the third row and the fourth column corresponds to the comparison between the two sequencing experiments, and the rest of plots correspond to the remaining crosscomparisons.

similar despite the differences in the methodologies used to measure mRNA abundances.

#### Detection of unconventional transcripts

Even with a low degree of coverage, the comparatively longer read size provided by pyrosequencing potentially allows for the detection of new events of splicing. Upon the application of Top-hat cuff links (Trapnell et al., 2010) to all the reads, including those that did not map into known transcripts, we have found six new splicing sites that originate new transcripts, not reported either in Ensembl or in SGD. Genes YGL031C (ribosomal protein L30 of the large 60S ribosomal subunit) and YER131W, YGR027C, YLR333C, YLR367W, and YLR388W (protein components of the small 40S ribosomal subunit) were found to display a new transcript form. All of them were 5' heterogeneity length variants. However, a comprehensive survey of the literature revealed that such unconventional splice variants had already been described elsewhere (Juneau et al., 2007; Miura et al., 2006). Only new splice variants supported by, at least, a 10× coverage were considered. It is expected that an increase of the coverage will reveal new unconventional transcripts.

#### Conclusions

The main aim of differential gene expression experiments is the quantification of the differences among transcripts when comparing conditions. For this purpose we have shown that only a few reads per gene can account for a reasonably accurate evaluation of changes in its transcriptional pattern. In addition, it also allows detection of a number of transcriptionally active genes quite similar to of the number observed in the microarray analysis. Therefore, our work demonstrates that, even in a scenario of low coverage, the concordance among DNA microarrays, which has been the reference technology for transcriptional studies and the pyrosequencing technologies, is quite acceptable. Furthermore, uncommon splicing variants can be found even at low levels of coverage. Obviously, the detection of low-abundance variants will be critically dependent on higher coverage measurements. On the other hand, low coverage implies a lower number of genes over this minimum number of reads required for detecting changes in their transcriptional statuses.

An interesting conclusion from this study is the fact that using the standard annotation of yeast the results obtained are clearly suboptimal. The reason for this is that many reads map beyond the standard gene boundaries. The use of more detailed annotations of gene and transcript boundaries, derived from more recent works (Nagalakshmi et al., 2008) produces clearly improved results.

#### Acknowledgments

The authors acknowledge the help of Raquel Serrano at the initial steps of the microarray analysis. Work was supported by grants BFU2008-04188-C03-01, GEN2006-27748-C2-1-E/SYS, and EUI2009-04147 (SysMO ERA-NET) to J.A.; BFU2009-11593 to A.C. (Ministry of Science and Innovation, Spain), BIO2008-04212 (Ministry of Science and Innovation), and PROMETEO/2010/001 (GVA-FEDER) to J.D. J.A. is recipient of an "Ajut de Suport a les Activitats dels Grups de Recerca" (2009SGR-1091, Generalitat de Catalunya). The National In-

stitute of Bioinformatics (www.inab.org) and the CIBER de Enfermedades Raras (CIBERER) are initiatives of the ISCIII.

#### Author Disclosure Statement

J.P.P. is an employee of Secugen SL; L.P. is an employee of Lifesequencing SL; M.A.T. is an employee of Roche Applied Science.

#### References

- Alberola, T.M., Garcia-Martinez, J., Antunez, O., Viladevall, L., Barcelo, A., Arino, J., et al. (2004). A new set of DNA microchips for the yeast *Saccharomyces cerevisiae*: features and uses. *Int Microbiol* 7, 199–206.
- Arino, J. (2010). Integrative responses to high pH stress in *S. cerevisiae*. *Omics* 14, 517–523.
- Blencowe, B.J., Ahmad, S., and Lee, L.J. (2009). Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes. *Genes Dev* 23, 1379–1386.
- Bloom, J.S., Khan, Z., Kruglyak, L., Singh, M., and Caudy, A.A. (2009). Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics* 10, 221.
- Cloonan, N., Forrest, A.R., Kolle, G., Gardiner, B.B., Faulkner, G.J., Brown, M.K., et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5, 613–619.
- Eveland, A.L., McCarty, D.R., and Koch, K.E. (2008). Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. *Plant Physiol* 146, 32–44.
- Flicek, P., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., et al. (2008). Ensembl 2008. *Nucleic Acids Res* 36, D707–D714.
- Hedbacker, K., and Carlson, M. (2008). SNF1/AMPK pathways in yeast. *Front Biosci* 13, 2408–2420.
- Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., et al. (2000). A concise guide to cDNA microarray analysis. *Biotechniques* 29, 548–550, 552–544, 556 passim.
- Ioannidis, J.P., Allison, D.B., Ball, C.A., Coulibaly, I., Cui, X., Culhane, A.C., et al. (2009). Repeatability of published microarray gene expression analyses. *Nat Genet* 41, 149–155.
- Juneau, K., Palm, C., Miranda, M., and Davis, R.W., 2007. High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc Natl Acad Sci USA* 104, 1522–1527.
- Lamb, T.M., and Mitchell, A.P. (2003). The transcription factor Rim101p governs ion tolerance and cell differentiation by direct repression of the regulatory genes NRG1 and SMP1 in *Saccharomyces cerevisiae*. *Mol Cell Biol* 23, 677–686.
- Lamb, T.M., Xu, W., Diamond, A., and Mitchell, A.P. (2001). Alkaline response genes of *Saccharomyces cerevisiae* and their relationship to the RIM101 pathway. *J Biol Chem* 276, 1850–1856.
- Li, H., Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18, 1509–1517.
- Medina, I., Carbonell, J., Pulido, L., Madeira, S.C., Goetz, S., Conesa, A., et al. (2010). Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic Acids Res* 38(Suppl), W210–W213.

- Miura, F., Kawaguchi, N., Sese, J., Toyoda, A., Hattori, M., Morishita, S., et al. (2006). A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci USA* 103, 17846–17851.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5, 621–628.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344–1349.
- Oshlack, A., Robinson, M.D., and Young, M.D. (2010). From RNA-seq reads to differential expression results. *Genome Biol* 11, 220.
- Platara, M., Ruiz, A., Serrano, R., Palomino, A., Moreno, F., and Arino, J. (2006). The transcriptional response of the yeast Na(+)-ATPase ENA1 gene to alkaline stress involves three main signaling pathways. *J Biol Chem* 281, 36632–36642.
- Ruiz, A., Serrano, R., and Arino, J. (2008). Direct regulation of genes involved in glucose utilization by the calcium/calci-neurin pathway. *J Biol Chem* 283, 13923–13933.
- Serrano, R., Martin, H., Casamayor, A., and Arino, J. (2006). Signaling alkaline pH stress in the yeast *Saccharomyces cerevisiae* through the Wsc1 cell surface sensor and the Sit2 MAPK pathway. *J Biol Chem* 281, 39785–39795.
- Serrano, R., Ruiz, A., Bernal, D., Chambers, J.R., and Arino, J. (2002). The transcriptional response to alkaline pH in *Saccharomyces cerevisiae*: evidence for calcium-mediated signaling. *Mol Microbiol* 46, 1319–1333.
- Shi, L., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C., et al. (2006). The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24, 1151–1161.
- Smyth, G. (2005). Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber, eds. (Springer, New York), pp. 397–420.
- Sultan, M., Schulz, M.H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., et al. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321, 956–960.
- T’Hoen, P.A.C., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H., de Menezes, R.X., et al. (2008). Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res* 36, e141.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28, 511–515.
- Viladevall, L., Serrano, R., Ruiz, A., Domenech, G., Giraldo, J., Barcelo, A., et al. (2004). Characterization of the calcium-mediated response to alkaline stress in *Saccharomyces cerevisiae*. *J Biol Chem* 279, 43614–43624.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10, 57–63.

Address correspondence to:

Joaquín Dopazo  
Department of Bioinformatics and Genomics  
Centro de Investigación Príncipe Felipe (CIPF)  
Valencia, Spain 46013

E-mail: jdopazo@cipf.es

