Genomics Data 5 (2015) 223–227

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

# An integrative approach to analyze microarray datasets for prioritization of genes relevant to lens biology and disease

Deepti Anand [a], Smriti A. Agrawal [a], Archana D. Siddam [a], Hozumi Motohashi [b], Masayuki Yamamoto [c], Salil A. Lachke [a,d,*]

[a] Department of Biological Sciences, University of Delaware, Newark, DE, USA
[b] Department of Gene Expression Regulation, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan
[c] Department of Medical Biochemistry, Graduate School of Medicine, Tohoku University, Sendai, Japan
[d] Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, USA

ARTICLE INFO

*Article history:*
Received 10 June 2015
Received in revised form 11 June 2015
Accepted 11 June 2015
Available online 14 June 2015

*Keywords:*
Cataract
Lens
Microarrays
*iSyTE*
Gene expression

ABSTRACT

Microarray-based profiling represents an effective method to analyze cellular or tissue-specific gene expression on the genome-level. However, in comparative analyses between control and mutant samples, microarrays often identify a large number of differentially expressed genes, in turn making it challenging to isolate the select "high-priority candidates" that are most relevant to an observed mutant phenotype. Here, we describe an integrative approach for mouse mutant lens microarray gene expression analysis using publically accessible systems-level information such as wild-type mouse lens expression data in *iSyTE* (integrated Systems Tool for Eye gene discovery), protein–protein interaction data in public databases, gene ontology enrichment data, and transcription factor binding profile data. This strategy, when applied to small Maf *Mafg*−/−:*Mafk*+/− mouse lens microarray datasets (deposited in NCBI Gene Expression Omnibus database with accession number GSE65500) in Agrawal et al. 2015 [1], led to the effective prioritization of candidate genes linked to lens defects in these mutants. Indeed, from the original list of genes that are differentially expressed at ±1.5-fold and $p < 0.05$ in *Mafg*−/−:*Mafk*+/− mutant lenses, this analysis led to the identification of thirty-six high-priority candidates, in turn reducing the number of genes for further study by approximately 1/3 of the total. Moreover, eight of these genes are linked to mammalian cataract in the published literature, validating the efficacy of this approach. Additionally, these high-priority candidates contribute valuable information for the assembly of a gene regulatory network in the lens. In sum, the pipeline outlined in this report represents an effective approach for initial as well as downstream microarray expression data analysis to identify genes important for lens biology and cataracts. We anticipate that this integrative strategy can be extended to prioritize phenotypically relevant candidate genes from microarray data in other cells and tissues.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

| Specifications | |
|---|---|
| Organism/cell line/tissue | Mus musculus; 60 day post-natal lens tissue |
| Sex | N/A |
| Sequencer or array type | MouseWG-6 v2.0 Expression BeadChip arrays (Illumina) |
| Data format | Raw |
| Experimental factors | Mafg−/−:Mafk+/− (test) vs. Mafg+/−:Mafk+/− (control) (Both on mixed background, with contributions from the 129Sv/J, C57BL/6J, and ICR strains) |
| Experimental features | Identification of differentially expressed genes in the lens of mice lacking two copies of Mafg and one copy of Mafk that exhibit lens defects with age compared to lens of mice lacking one copy of Mafg and one copy of Mafk that are normal |
| Consent | N/A |
| Sample source location | Newark, Delaware, USA |

## 1. Direct link to deposited data

Deposited data can be found at: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE65500.

* Corresponding author at: Department of Biological Sciences, University of Delaware, Newark, DE 19716, USA.
*E-mail address:* salil@udel.edu (S.A. Lachke).

http://dx.doi.org/10.1016/j.gdata.2015.06.017
2213-5960/© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 2. Experimental design, materials and methods

### 2.1. Experimental design

The eye gene discovery tool *iSyTE* [2] predicted that the small Maf transcription factor genes *Mafg* and *Mafk* are highly enriched and expressed, respectively, in the mouse lens, and may potentially function in its development or homeostasis. Indeed, we find that *Mafg*−/−:*Mafk*+/− compound mouse mutants exhibit lens defects including cataracts [1]. To investigate the molecular changes associated with this phenotype, we performed microarray-based gene expression profiling on lens tissue obtained from *Mafg*−/−:*Mafk*+/− ("test" animals that exhibit lens defects) and *Mafg*+/−:*Mafk*+/− ("control" animals that do not exhibit lens defects) mice. These microarray datasets were analyzed to identify differentially expressed genes (DEGs) that were subjected to further integrated analysis using a workflow pipeline that incorporates wild-type lens expression data from the *iSyTE* database (http://bioinformatics.udel.edu/Research/iSyTE), chromatin immunoprecipitation sequencing (ChIP-Seq) data on Mafg and its co-regulatory protein Nrf2, protein–protein interactions from open resource databases, as well as gene ontology enrichment and *cis*-regulatory motif analyses (Fig. 1). Together, this approach led to the prioritization of genes most relevant to the phenotype, and to the assembly of a small Maf gene regulatory network in the lens. Below, we describe these analytical procedures in detail.

### 2.2. Microarrays

Lenses were collected on post-natal day (P) 60 from *Mafg*−/−:*Mafk*+/− (test) and *Mafg*+/−:*Mafk*+/− (control) mice for microarray-based gene expression profiling. Selection of the stage for genome-wide expression profiling of mutant tissue represents an important step, because this data influences the proper interpretation of the phenotype as well as the directions of future experiments. For example, microarray profiling performed at a stage when the phenotype is already evident, and therefore well manifested, may provide differential expression data that is potentially "contaminated" with secondary gene expression changes. This presents a challenge to resolve primary expression changes from such secondary expression changes, in turn making it difficult to correlate genetic perturbations to the observed phenotype. To address this issue, the age of the animals in this analysis was selected as P60 (2 months) because at this stage, *Mafg*−/−:*Mafk*+/− compound mutants do not exhibit any overt abnormalities in the lens (Fig. 2A–B'), which are observed in these mutants at later stages (Fig. 2C–D'). Thus, analysis of lenses at P60 increases the likelihood of detecting changes in mutant gene expression that occur prior to the onset of the overt phenotype and therefore represent primary alterations. Total RNA was isolated from both test and control lenses in biological duplicates using the RNeasy Mini Kit (Qiagen). Microarrays were performed on MouseWG-6 v2.0 Expression BeadChip arrays (Illumina) using manufacturer recommended hybridization conditions. Microarray chips were scanned using the Illumina BeadArray reader.

### 2.3. Microarray data quality assessment

To assess the quality of microarray expression data, we used inbuilt 'R' functions and performed Principle Component Analysis (PCA), derived boxplots, and analyzed histograms of the array intensities of raw unprocessed and normalized processed (see below) data (Fig. 3). We observed high level of background noise in raw mutant and control datasets in all quality plots. PCA plot from processed data showed clear separation of control and mutant datasets (Fig. 3A–B). Similar results were obtained with histograms (Fig. 3C–D) and boxplots (Fig. 3E–F) analysis of raw and processed data.

### 2.4. Microarray data preprocessing and analysis

Microarray data processing and analysis was performed under the 'R' statistical environment (http://www.r-project.org/) using *lumi* package for Illumina microarray data, available through Bioconductor (www.bioconductor.org). The workflow for microarray data analysis is described in detail (Fig. 1). As first step, raw output files from Illumina Bead Studio toolkit were read using *lumi* [3] to generate a *lumiBatch* object through *lumiR.batch* function. These imported files were then preprocessed using *lumi* built-in methods, *bgAdjust* which corrects raw Illumina probe intensities followed by *rankinvariant* for normalizing these corrected intensities. The choice of *rankinvariant* was made because only a small number of genes were expected to be differentially expressed, and Illumina recommends this normalizing method for such experiments. After pre-processing, presence–absence calls were generated using built-in function of *lumi*, and probe sets with detection *p*-value of ≤0.05 in at least two samples were considered as significantly present and used for further downstream analysis. Finally, probe set-level experiment was converted to gene-level by selecting a single

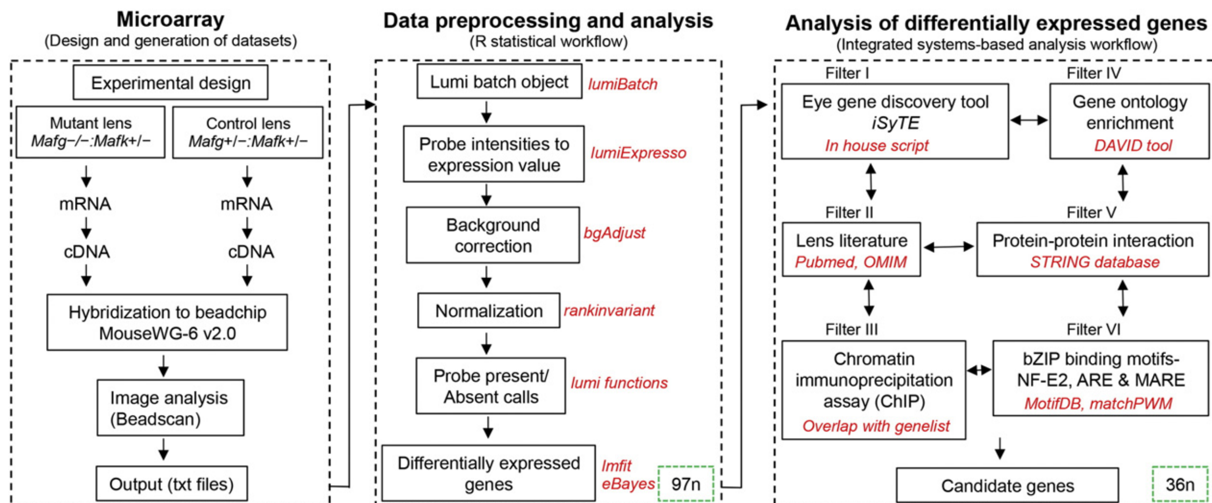


**Fig. 1.** Workflow of microarray design, data pre-processing and analysis of differentially expressed genes.

Download English Version:

https://daneshyari.com/en/article/2821870

Download Persian Version:

https://daneshyari.com/article/2821870

Daneshyari.com