ChipQC: Microarray Artifact Visualization Tool

Peter A. Henning[1#], Paul K. Tan[2#], Tung Yu Chu[3], David A. Stiles[3], David Wheeler[4], Pushkar Mukewar[5], Margaret C. Cam[3]*, May D. Wang[1]*

Keywords: Microarray, Quality Control, Error Detection, Web Server

1 Background

Recent advances in biotechnology have led to the development of several high-throughput procedures, such as DNA microarrays, have dramatically increased the amount of information obtained in a relatively short time span[1]. Although experiments using these new techniques are highly automated, downstream analytical methods still lag behind in terms of automation and efficiency[2]. Microarray studies typically investigate gene expression profiling[1, 3], which can be greatly skewed by the slightest amount of contamination. The fabrication or hybridization processes would easily introduce areas of high local variability, which can only be accurately detected when several replicates are analyzed together[4].

ChipQC is a novel web-based software tool that is joined researched and developed by MIBLab at Biomedical Engineering Department of Georgia Tech and Emory University and the Microarray Core Facility at National Institute of Diabetes and Digestive and Kidney Diseases. It is designed to perform standard error analysis and statistical techniques on multiple array sets (technical or biological replicates). Although chip quality control is often overlooked in microarray experiments, it the reproducibility and reduction of systematic error afforded by extensive quality control that will allow microarray systems to migrate from an experimental research tool to a clinical diagnostic device.

2 System Development

ChipQC has evolved from a simple Unix program used to investigating small sections of Affymetrix GeneChips® to a web-based comprehensive microarray analysis tool that can employed on any of six supported chip types, both commercial and "homemade". ChipQC include calculating the coefficient of variation, standard deviation, mean intensity, fold change, and statistical significance. ChipQC, using a heat map scheme, graphically represents the error analysis and various metrics of each gene at its proper chip coordinates, thus mimicking the analyzed chip's configuration. Use of this visualization tool revealed localized areas of high variability pattern in some arrays consistent with an edge effect, which persisted even after lowess or linear normalization had been applied, demonstrating the need to select and flag specific spots which would potentially yield falsely

[1] Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, 313 Ferst Dr., Atlanta, Georgia, USA. Email: maywang@bme.gatech.edu

[2] John A. Burns School of Medicine, University of Hawaii, Manoa, Hawaii, USA. Work done while at NIDDK Microarray Core Facility.

[3] Microarray Core Facility, National Institute of Diabetes Digestive and Kidney Disorders (NIDDK), National Institutes of Health, Building 8, Bethesda, Maryland, USA. Email: MaggieC@intra.niddk.nih.gov

[4] National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Building 38, Bethesda, Maryland, USA.

[5] School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA.

[#] These authors contributed equally.

[*] Authors whom correspondence should be addressed.

positive or negative gene expression results. The latest version of the software still contains the orginial Perl core but has been enhanced by C and Java programs, which put common normalization schemes and the ability to create histrograms just a click away. The heatmap image generation is made possible by the Boutell GD Perl module (www.boutell.com). There are six key steps contained in the ChipQC workflow (Figure 1).

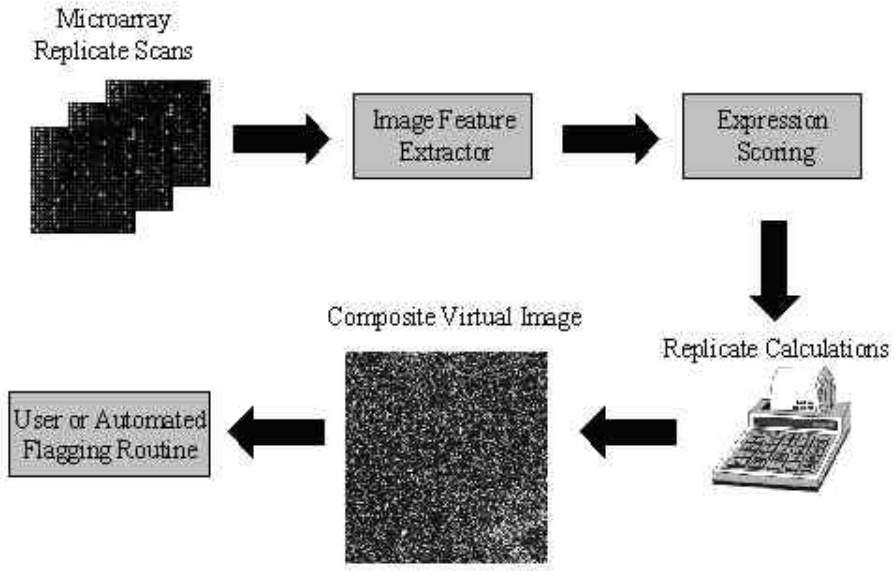

Figure 1 ChipQC System Diagram

3 Future Work

The composite image has been used to identify systematic errors presented such as edge effect. While we are in the process of reporting such discovery in full detail to the technical community, we have been working to add other features such as more normalization options, statistical options, and more supported chip types to the existing software. The hope is to also pursue more substantial improvements by implementing the latest image processing methods to enhance image feature extraction, expression scoring, and possibly the automated flagging routine.

References

1. Schena, M., et al., *Quantitative Monitoring of Gene-Expression Patterns with a Complementary-DNA Microarray*. Science, 1995. **270**(5235): p. 467-470.
2. Holloway, A.J., et al., *Options available - from start to finish - for obtaining data from DNA microarrays.* Nature Genetics, 2002. **32**: p. 481-489.
3. Lockhart, D.J., et al., *Expression monitoring by hybridization to high-density oligonucleotide arrays.* Nature Biotechnology, 1996. **14**(13): p. 1675-1680.
4. Slonim, D.K., *From patterns to pathways: gene expression data analysis comes of age.* Nature Genetics, 2002. **32**: p. 502-508.