Data Mining *in situ* Gene Expression Patterns at Cellular Resolution

James Carson
Baylor College of Medicine,
One Baylor Plaza, Houston TX 77030

Tao Ju
Rice University
6100 S. Main, Houston TX 77005

Christina Thaller
Baylor College of Medicine,
One Baylor Plaza, Houston TX 77030

Joe Warren
Rice University
6100 S. Main, Houston TX 77005

Musodiq Bello
The University of Houston,
4800 Calhoun Rd., Houston TX 77204

Ioannis Kakadiaris
The University of Houston,
4800 Calhoun Rd., Houston TX 77204

Wah Chiu
Baylor College of Medicine,
One Baylor Plaza, Houston TX 77030

Gregor Eichele
Baylor College of Medicine,
One Baylor Plaza, Houston TX 77030

Abstract

In the post-genomic era, large-scale efforts have begun to characterize the role of gene products. Several of these efforts aim to systematically discover the activity of all ~20,000 genes throughout functionally complex tissue specimens such as embryo and the mature brain. By applying a subdivision-based deformable model of the brain, we rapidly organize spatial gene expression data into a common coordinate system. Doing this enables powerful queries, comparisons, and associations of the data.

1. Introduction

Non-radioactive *in situ* hybridization (ISH) is a powerful technique for revealing gene expression in individual cells, the level of detail necessary for investigating how genes control cell type identity, cell differentiation, and cell-cell signaling (Fig. 1). Although the availability of robotic ISH enables the expeditious determination of expression patterns for thousands of genes in serially sectioned tissues, a large collection of ISH images is, per se, of limited benefit [1]. However, via accurate detection of expression strength and spatial normalization of expression location across different specimens, ISH images become a minable resource of annotated gene expression capable of advancing functional genomics in a mode similar to DNA sequence databases.

2. Methods

We have developed computational methods to automate ISH image annotation and applied these to over 200 genes throughout the postnatal mouse brain. First, gene expression strengths were semi-quantitatively characterized for each cell in a tissue section [2]. Atlas-based segmentation was then performed using a series of subdivision mesh maps that comprise our atlas of the postnatal mouse brain [3]. These maps were deformed to fit the tissue sections containing gene expression, and the detected expression strengths were associated with the directly overly mesh to provide a common geometric annotation of gene expression. Automated textual annotation of expression patterns took advantage of the explicitly defined boundaries of the mesh.

3. Results

Automated textual annotations of gene expression patterns were found to match accurately the annotations determined visually by expert. Spatial searches were successfully applied

to find the genes expressed in a similar fashion to custom query patterns. Gene groups identified by clustering the genes based upon their spatially normalized expression patterns suggest that cluster analysis of robotic ISH data can identify potential relationships in gene networks. Annotated gene expression patterns and query interfaces are publicly accessible at www.geneatlas.org.
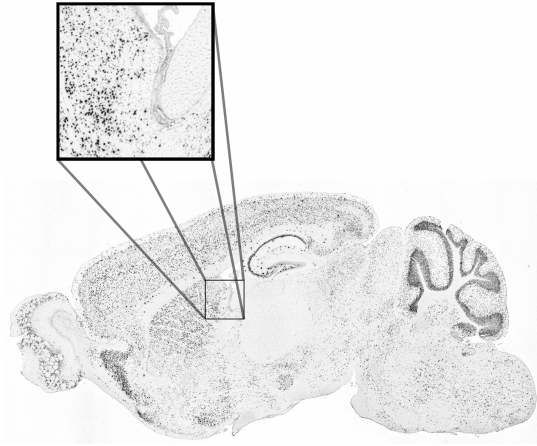
Figure 1. Example of *in situ* hybridization. Cells that express the gene of interest, in this case *Cnr1*, are darkly stained.

4. Conclusion

Automated annotation of gene expression patterns generated by robotic ISH is a vitally important development for enabling powerful knowledge discovery, especially on a genomic-scale.

5. Acknowledgments

6. References

[1] J.P. Carson, C. Thaller, and G. Eichele "A Transcriptome Atlas of the Mouse Brain at Cellular Resolution*", Current Opinions in Neurobiology*, October 2002, pp. 562-5.

[2] J.P. Carson, G. Eichele, and W. Chiu, "A Method for Automated Detection of Gene Expression Required for the Establishment of a Digital Transcriptome-wide Gene Expression Atlas", *Journal of Microscopy,* March 2005, pp. 275-81.

[3] T. Ju, J. Warren, G. Eichele, C. Thaller, W. Chiu, and J. Carson, "A Geometric Database for Gene Expression Data", *Proceedings of the First Eurographics Symposium on Geometry Processing*, Aachen, Germany, 2003, pp. 166-76.