Quantitative pattern analysis of a large tissue collection across multiple experimental variables

Macura, T., Johnston, J., Cordero, J., Orlov, N., Shamir, L., Goldberg, I. G.

Image Informatics and Computational Biology Unit, Laboratory of Genetics, NIA/NIH. 333 Cassell Dr., Baltimore, MD, 21224.

The Atlas of Gene Expression in Mouse Aging Project (AGEMAP) has generated a dataset of hematoxylin and eosin stained mouse tissues that we are analyzing with automated image classifiers to study how tissue structural decline is correlated with aging. In AGEMAP, 48 male and female mice on ad-libitum and caloric restriction diets were sacrificed at multiple ages (1, 6, 16, and 24 months). Mouse organs were harvested, sectioned, and put on slides according to a standardized protocol. Imaging the slides, we collected ~5000 images of mouse livers, skeletal muscles, and kidney tubules. A unique feature of this dataset is that its images can be compared along several crosscutting axes (age, gender, diet, and tissue). By training image classifiers along each axis, we were able to quantitatively assess the effects of age, gender, and diet on tissue morphology. Thus for each organ we constructed four gender, two diet, and two aging classifiers; twenty-four classifiers in total. Beginning with 18GB of raw pixels, it took approximately 2000 processor hours to generate 125GB of intermediary pixels and extract 90 million image descriptor features used for classification. Ultimately the classification results were summarized in just a dozen figures and tables. Here we describe the informatics tools, data management, and data analysis for this complex biological dataset. Preliminary findings indicate that our image classifiers are able to detect and quantify aging related tissue morphology changes. We are on-track to scaling this project by imaging additional organs, including brain, lungs, pancreas, thymus, and heart. The resulting 25,000 image dataset will be analyzed with the informatics techniques described here.