# Segway: semi-automated genome annotation

Eric Roberts[1], Michael M. Hoffman[1,2,3]

[1] Princess Margaret Cancer Centre, Toronto, ON, Canada.

[2] Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada.

[3] Department of Computer Science, University of Toronto, Toronto, ON, Canada. Email: michael.hoffman@utoronto.ca

**Project Website**: http://segway.hoffmanlab.org/
**Source Code**: https://bitbucket.com/hoffmanlab/segway/
**License**: GNU General Public License version 2

The free Segway software package contains a method for analyzing multiple tracks of functional genomics data. Our method uses a dynamic Bayesian network (DBN) model, which enables it to analyze the entire genome at 1-bp resolution even in the face of heterogeneous patterns of missing data. This method is the first application of DBN techniques to genome-scale data and the first genomic segmentation method designed for use with the maximum resolution data available from ChIP-seq experiments without downsampling. Our software has extensive documentation and was designed from the outset with external users in mind. Segway annotations for the human epigenome are now built-in to the Ensembl and UCSC Genome Browsers.

We have continued development of Segway and Segway annotations since our initial publication (Hoffman et al. 2012 *Nat Methods* 9:473). We have switched to open development in a Mercurial repository on Bitbucket. Improvements in deployment of Segway and the underlying Graphical Models Toolkit (GMTK) and Genomedata functional data storage system mean user installation is now substantially easier. We use the drone.io continuous integration system and a series of new regression tests to automatically ensure a high level of software quality.

We have added a number of new features to Segway. It now includes support for running in a local mode without a cluster, in addition to the Grid Engine, Platform Load Sharing Facility (LSF), and Portable Batch System (PBS) cluster management systems. Segway now supports a concatenated segmentation mode for analyzing multiple grouped datasets. We continue to develop other new features such as a hierarchical segmentation mode.