

## Poster H-27

### Predicting Nuclear Localization



#### Authors:

John Hawkins (*University of Queensland*)

Mikael Boden (*University of Queensland*)

**Short Abstract:** Predicting nuclear localization of proteins is complicated due to the diversity of targeting signals. Many nuclear proteins shuttle in and out of the nucleus. However, current nuclear localization predictors use exclusively nuclear localized proteins in their training sets. We present the first complete model for identifying nuclear localized proteins.

#### Long Abstract:

Nuclear localization of proteins is perhaps the most complicated of localization processes, due to the massive diversity of targeting signals. Furthermore, many nuclear proteins spend their lives shuttling in and out of the cytoplasm, a process which makes nuclear localization a crucial element in the dynamic life of the cell. In spite of this, nearly all nuclear localization predictors use only exclusively nuclear localized proteins in their training sets, meaning that all shuttling proteins are excluded from the development of the model. Furthermore, many models rely on sequence homology for their predictions, thus the models are trained to recognize overall sequence similarity rather than recognizing the presence of the targeting signals used by the cell. This means that these models will not perform well on truly novel data, in other words will have little capacity for genuine generalisation. We present a model focused exclusively on the task of identifying nuclearly localized proteins, on the basis of sequence features, regardless of whether this is a permanent or temporary assignment.

**Results:** The resulting classifier has an accuracy of 0.76 and a Matthews' Correlation Coefficient of 0.52. We demonstrate that these results are the best current estimate of generalisation on this problem, previous studies having used only partial datasets, often with considerable redundancy.

**Availability:** The NucProwler prediction service is being integrated with the PProwler protein localization predictor available at: <http://pprowler.itee.uq.edu.au>