Detecting Somatic Mutation In Tumour/Normal Paired Sample Sequence Data

Keywords: Cancer – Somatic Mutations – Paired Data Sequencing – Machine Learning – Probabilistic Graphical Models –

Abstract: Next generation sequencing is playing an increasingly important role in understanding the biology of cancer. Among the most import problems that can be solved using genomic sequence data, is the identification of somatic mutations in primary or metastatic tumours. Current experimental designs for identifying somatic mutations typically use a sample of healthy (normal) tissue as a control. Sequence data from tumour samples can then be compared against this background to identify somatic mutations while screening out germline polymorphisms. Independent analysis of tumour and normal samples to detect SNVs ignores the significant correlation present. We have developed a probabilistic graphical model to jointly analyse both samples which allows us to borrow statistical strength to improve accuracy. We compare our joint modeling approach to an independent analysis using the previously published SNVMix software. We observe an improvement in both sensitivity and specificity using synthetic and real data with validated ground truth mutations.

Authors:

- Andrew Roth andrewjlroth@gmail.com Canada University of British Columbia
- Samuel Aparicio saparicio@bccrc.ca Canada BC Cancer Agency
- Sohrab Shah sshah@bccrc.ca Canada BC Cancer Agency