BMC Bioinformatics

Variant detection and the Autism sequencing project

Orion Buske[1*], Misko Dzamba[1], Justin Foong[2], Lynette Lau[2], Marc Fiume[1], Christian Marshall[2], Susan Walker[2], Aparna Prasad[2], Michael Brudno[1,2,3]

Background

Early detection of autism can improve the quality of life of affected individuals [1]. Qualitative screening methods continue to improve, but still suffer from low sensitivity despite increasing specificity [2,3]. In collaboration with the Hospital for Sick Children, we are sequencing the exomes of 1 000 individuals with autism in order to discover genetic variants associated with the disorder. Discovery of associated variants can lead to earlier diagnosis and treatment.

Materials and methods

We will present our current sequencing and analysis pipeline, from SureSelect exome capture and SOLiD sequencing through Sanger validation of predicted harmful variants, along with tools we have developed for color-space-aware alignment, variant detection, and visualization of next-generation sequencing data.

Color-space sequencing provides a tradeoff between enhanced ability to distinguish Single Nucleotide Variants (SNVs) from sequencing errors at the price of a higher sequencing error rate versus traditional letter-space sequencing. This technology has the potential to provide higher accuracy at lower cost, but opens new computational challenges that need to be addressed.

Results and conclusions

We have sequenced over 70 individuals so far at approximately 30x mean coverage, have found and validated several interesting non-synonymous Single Nucleotide Variants (SNVs), and have identified a number of potential de novo non-synonymous mutations. After filtering, we are identifying an average of over 17 000 non-synonymous SNVs per individual, of which over 11 000 are novel to dbSNP. We also find that support from both strands is more informative than total depth of coverage for predicting SNVs from high-throughput sequencing data. This is of considerable importance in exome capture data, since only a small region at each probe captures sequence from both strands.

Author details
[1]Department of Computer Science, University of Toronto, Toronto, Canada. [2]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada. [3]Donnelly Centre, University of Toronto, Toronto, Canada.

References
1. Lord C, McGee JP: **Committee on educational interventions for children with autism, national research council: conclusions and recommendations.** In *Educating Children with Autism*. Washington, DC;Press NA 2001:211-230.
2. Baird G, Charman T, Baron-Cohen S, Cox A, Swettenham J, Wheelwright S, Drew A: **A screening instrument for autism at 18 months of age: a 6-year follow-up study.** *J Am Acad Child Adolesc Psychiatry* 2000, **39(6)**:694-702.
3. Robins DL, Fein D, Barton ML, Green JA: **The modified checklist for autism in toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders.** *J Autism Dev Disord* 2001, **31(2)**:131-144.

[1]Department of Computer Science, University of Toronto, Toronto, Canada
Full list of author information is available at the end of the article

BioMed Central