



Category: Bioinformatics

Statistical model based gender prediction for targeted NGS clinical panels

Palani Kannan Kandavel and Vivek Gopalan

MedGenome Laboratory Pvt. Ltd., Narayana Health city, Bangalore, Karnataka, INDIA

Abstract

In the NGS based clinical diagnosis, the gender information plays key roles such as Mendelian inheritance study, clerical or sampling error identification, quality control over clinical samples (e.g., transfusion performed blood samples). For this purpose, the gender determination using the sequenced NGS targeted panel data is crucial need irrespective of gender information from Test Requisition Form (TRF) in clinical diagnosis. But, sample impurities and its quantity, coverage over sex chromosomes in targeted panels and depth over sex chromosomes are limiting factors for the targeted NGS panels. Three common approaches such as genotype (Hom/Het) composition in ChrX, read depth proportion between ChrX vs ChrY and ChrX vs selected autosomal chromosome are being used for gender determination. As an improvement from MedGenome's current gender prediction (based on "genotype composition in ChrX") approach, we have developed a new statistical model based approach for targeted panels which can be used to predict gender with increased sensitivity especially for captured NGS panels such as Whole Exome and Clinical Exome panels. In this new approach, instead of considering a single approach for gender prediction, we included three approaches together with built statistical reference model.

The model has been built with a set of 700 control clinical dataset (sequenced in NGS targeted panels) as training data. The clinical data such as confirmed gender from clinician on TRF (as response variable), genotype proportion in ChrX, depth ratio between ChrX and ChrY and depth ratio between ChrX and Chr17 (as predictor variables) have been used as training criteria. A multi-regression model has been developed for each targeted panel in MedGenome. The prediction score of being male/female (with respect to the built reference model) gives the confidence on the gender of clinical dataset.

The reference test dataset are being used to test the model. The sensitivity on predicting the gender has been increased from the current "genotype composition in ChrX" based approach. In addition, the prediction score given by the model can be used to evaluate the quality of clinical dataset. The higher prediction score towards its respective gender indicates the higher quality of sequenced data.

Citation: Kandavel, P.K. and Gopalan, V. Statistical model based gender prediction for targeted NGS clinical panels [Abstract]. In: Abstracts of the NGBT conference; Oct 02-04, 2017; Bhubaneswar, Odisha, India: *Can J biotech*, Volume 1, Special Issue (Supplement), Page 242.
<https://doi.org/10.24870/cjb.2017-a227>