KG Ouwens[1,2], B Tolhuis[2], JJ Hottenga[1], R Jansen[3], MA Swertz[4], H Karten[2], The Genome of the Netherlands Consortium, BJWH Penninx[3], DI Boomsma[1]

Testing and quantification of performance of NGS variance calling methods

[1]Department of Biological Psychology, VU University Amsterdam, [2]Genalice BV Harderwijk,
[3]Department of Psychiatry, VU Medical Center, Amsterdam
[4]Genetics Department, University Medical Centre Groningen

Introduction

De novo mutations are important factors contributing to the etiology of rare diseases, and particularly difficult to detect. To estimate the number and location of rare events, it is crucial to have non-biased tools and metrics to estimate specificity and sensitivity of variance calling methods. Various variant calling techniques make different assumptions (or utilize *best practices)* of which the benefit may depend on sequencing coverage depth. Investigating monozygotic twin pairs, sequenced on multiple platforms with various read depths, will offer valuable insight in the performance and culprits of variant caller methods.

Data

Whole-genome, whole-blood DNA sequencing of two pairs of monozygotic twins is available. One twin pair is aged 40, the other aged 100. Sequencing has been done three times: using the Illumina platform, by Complete Genomics, and as part of the GoNL project. For more details, see Table 1.

	Illumina	Complete Genomics	GoNL protocol at BGI
depth	~20X	45-87X	14-15X
read size	100 bp		91bp
coverage ≥20X	94.14%	95.2%	
coverage ≥40X	66.69%		
variants detected	~4.1M	~3.9M	

Table 1

Method

Variant calling will be performed on both MZ twin pairs with data from all three platforms, using various approaches (including GATK and Platypus) and compared with software developed by Genalice. Results will be compared with Golden Standard SNPs (where applicable), by Ti/Tv ratio and with variant calling benchmarking toolkits.

Results from this project will be used as input for a larger project with 11 quartets with MZ twins. Using the best methods and parameters will aid in reliable detection of number as well as timing of *de novo* mutations.

References

Boomsma, D. I. et al., (2014). The Genome of the Netherlands: design, and project goals. *European Journal of Human Genetics*,*22*(2), 221-227.

Drmanac, R. et al., (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science*,*327*(5961), 78-81.

Ye, K et al., (2013). Aging as accelerated accumulation of somatic variants: whole-genome sequencing of centenarian and middle-aged monozygotic twin pairs. *Twin Research and Human Genetics*, *16*(06), 1026-1032.