BMC Bioinformatics

Practicality of identifying mitochondria variants from exome and RNAseq data

Pan Zhang[1], David C Samuels[2], Brian Lehmann[3], Thomas Stricker[4], Jennifer Pietenpol[3], Yu Shyr[1], Yan Guo[1*]

From 14th Annual UT-KBRIN Bioinformatics Summit 2015
Buchanan, TN, USA. 20-22 March 2015

Background

The rapid progress in high throughput sequencing technology has significantly enriched our capability to study mitochondria genomes. Other than performing mitochondria targeted sequencing, an increasingly popular alternative approach is to utilize the off-target reads from exome sequencing to infer mitochondria genomic variants including SNP and heteroplasmy[1-9]. However, the effectiveness and practicality of such an approach has not been tested. Recently, RNAseq data has also been suggested as good source for alternative data mining[10,11], but whether mitochondria variants are minable has not been studied.

Materials and methods

We designed a specific study using targeted mitochondria sequencing data as a gold standard to evaluate the practicality of SNP and heteroplasmy detection using exome sequencing and RNAseq data. Five breast cancer cell lines were sequenced for mitochondria targeted sequencing, exome sequencing, and RNAseq. Furthermore, we examined three mitochondria alignment strategies: 1) align all reads directly to the mitochondria genome; 2) align all reads to the nuclear genome and mitochondria genome simultaneously; 3) align all reads to the nuclear genome first, then used the unaligned reads to align to the mitochondria genome.

Results

Our analyses found that exome sequencing can accurately detect mitochondria SNPs and can detect a portion of the true heteroplasmies with a reasonable false discovery rate. RNAseq data on the other hand had a lower detection rate

of SNP but higher detection rate for heteroplasmy. However, the higher false discovery rate makes RNAseq a less ideal source for studying mitochondria compared to exome sequencing data. Furthermore, we found that aligning all reads directly to the mitochondria genome reference or aligning all reads to the nuclear genome and mitochondria genome references simultaneously produced the best results.

Conclusions

Exome sequencing and RNAseq data can be potentially mined for mitochondria variants. Overall, exome sequencing provides less false discovery than RNAseq for mitochondria variant detection, making it a more desirable choice. In conclusion, our study provides important guidelines for future studies that intend to use exome sequencing or RNAseq data to infer mitochondria SNP and heteroplasmy.

Authors' details
[1]Center for Quantitative Sciences, Vanderbilt University, Nashville TN 37027, USA. [2]Center for Human Genetics Research, Vanderbilt University, Nashville TN 37037, USA. [3]Department of Biochemistry, Vanderbilt University, Nashville TN 37232, USA. [4]Department of Pathology, Vanderbilt University, Nashville TN 37027, USA.

References
1. Samuels DC, Han L, Li J, Quanghu S, Clark TA, Shyr Y, Guo Y: **Finding the lost treasures in exome sequencing data.** *Trends Genet* 2013, **29**(10):593-599.
2. Ye F, Samuels DC, Clark T, Guo Y: **High-throughput sequencing in mitochondrial DNA research.** *Mitochondrion* 2014, **17**:157-163.
3. Picardi E, Pesole G: **Mitochondrial genomes gleaned from human whole-exome sequencing.** *Nature Methods* 2012, **9**(6):523-524.
4. Guo Y, Li J, Li CI, Shyr Y, Samuels DC: **MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis.** *Bioinformatics* 2013, **29**(9):1210-1211.
5. Dinwiddie DL, Smith LD, Miller NA, Atherton AM, Farrow EG, Strenk ME, Soden SE, Saunders CJ, Kingsmore SF: **Diagnosis of mitochondrial**

* Correspondence: yan.guo@vanderbilt.edu
[1]Center for Quantitative Sciences, Vanderbilt University, Nashville TN 37027, USA
Full list of author information is available at the end of the article

disorders by concomitant next-generation sequencing of the exome and mitochondrial genome. *Genomics* 2013, **102**(3):148-156.

6. Falk MJ, Pierce EA, Consugar M, Xie MH, Guadalupe M, Hardy O, Rappaport EF, Wallace DC, LeProust E, Gai XW: **Mitochondrial Disease Genetic Diagnostics: Optimized Whole-Exome Analysis for All MitoCarta Nuclear Genes and the Mitochondrial Genome.** *Discov Med* 2012, **79**:389-U140.

7. Nemeth AH, Kwasniewska AC, Lise S, Schnekenberg RP, Becker EBE, Bera KD, Shanks ME, Gregory L, Buck D, Cader MZ, Talbot K, De Silva R, Fletcher N, Hastings R, Jayawant S, Morrison PJ, Worth P, Taylor M, Tolmie J, O'Regan M, Consortium UA, Valentine R, Packham E, Evans J, Seller A, Ragoussis J: **Next generation sequencing for molecular diagnosis of neurological disorders using ataxias as a model.** *Brain* 2013, **136**:3106-3118.

8. Sevini F, Giuliani C, Vianello D, Giampieri E, Santoro A, Biondi F, Garagnani P, Passarino G, Luiselli D, Capri M, Franceschi C, Salvioli S: **mtDNA mutations in human aging and longevity: controversies and new perspectives opened by high-throughput technologies.** *Exp Gerontol* 2014, **56**:234-244.

9. McMahon S, LaFramboise T: **Mutational patterns in the breast cancer mitochondrial genome, with clinical correlates.** *Carcinogenesis* 2014, **35**(5):1046-1054.

10. Han L, Vickers KC, Samuels DC, Guo Y: **Alternative applications for distinct RNA sequencing strategies.** *Brief Bioinform* 2014, **16**(4):629-639.

11. Vickers KC, Roteta LA, Hucheson-Dilks H, Han L, Guo Y: **Mining diverse small RNA species in the deep transcriptome.** *Trends Biochem Sci* 2015, **40**(1):4-7.