ADNI Alzheimer's Disease Neuroimaging Initiative

Assembly of 809 whole mitochondrial genomes with clinical, imaging, and fluid biomarker phenotyping: the Alzheimer's Disease Neuroimaging Initiative

Perry G. Ridge[1], Mark E. Wadsworth[1], Justin B. Miller[1], Andrew J. Saykin[2], Robert C. Green[3], for the Alzheimer's Disease Neuroimaging Initiative, John S.K. Kauwe[1,4]

[1]Department of Biology, Brigham Young University, Provo, UT

[2]Radiology and Imaging Sciences, Medical and Molecular Genetics and the Indiana Alzheimer's Disease Center, Indiana University School of Medicine, Indianapolis, IN

[3]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Partners HealthCare Personalized Medicine, the Broad Institute and Harvard Medical School, Boston, MA

[4]Department of Neuroscience, Brigham Young University, Provo, UT

Contents

Abstract

Mitochondrial genetics are an important, but largely neglected area of research in Alzheimer's disease. A major impediment is a lack of datasets. To help address this impediment we prepared a dataset that consists of 809 complete and annotated mitochondrial genomes from samples from the Alzheimer's Disease Neuroimaging Initiative (ADNI). These whole mitochondrial genomes include rich phenotyping, such as clinical, fluid biomarker, and imaging data, all of which is available through the ADNI website. Genomes are cleaned, annotated, and prepared for analysis. These data provide an important resource for investigating the impact of mitochondrial genetic variation on risk for Alzheimer's disease and other phenotypes that have been measured in the ADNI samples.

Methods

Genome Assembly and Variant Detection

ADNI mapped the whole genome sequences and called variants using the Burrows-Wheeler Aligner (BWA) [1] for mapping and standard best practices from the Genome Analysis Toolkit (GATK) [2, 3] for variant detection (for details see http://adni.loni.usc.edu/data-samples/genetic-data/wgs/). However, these steps needed to be redone for two reasons. First, original mappings were to Hg19, which includes a version of the mitochondrial genome, but not the standard rCRS (NC_01292) that is typically used for mitochondrial genetics. Second, the standard GATK pipeline is designed for diploid genomes.

Using the original mappings, we extracted all reads that mapped to the mitochondrial genome, or were unmapped, with SAMTools [4] and mapped them to NC_01292 using BWA. Next, we performed local realignments around indels and base recalibration with GATK to refine the mappings. Finally, we used FreeBayes (-p 1 –F 0.6, and removed variants with quality less than 20) [5] to joint-call variants and converted the resulting VCF file to fasta with vcf2fasta (vcflib, https://github.com/vcflib/vcflib).

Genome Annotation

We annotated mitochondrial variants and haplotypes for each sample. We downloaded 9228 mitochondrial DNA coding and RNA sequence variants and 2792 control region variants from MITOMAP [6]. For each variant present in the datasets downloaded from MITOMAP we added complete information (i.e. frequency, source, locus names, etc.) to the "Info" column in the VCF file and added the corresponding annotation information to the header lines. For each variant that was reported by multiple studies in MITOMAP we included all studies in the annotation.

Next, we annotated mitochondrial haplotypes with Phy-Mer [7]. Phy-Mer reports the five most likely mitochondrial haplotypes and a score, where 1 is a perfect score. For each of the samples we selected the top hit. All samples had scores >0.99 except for one that had a score of 0.988.

Dataset Information

This methods document applies to the following dataset(s) available from the ADNI repository:

Dataset Name	Date Submitted
ADNI Whole Mitochondrial Genomes	1 February 2017

References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PubMed Central PMCID: PMC2705234.
2. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43(5):491-8. Epub 2011/04/12. doi: ng.806 [pii] 10.1038/ng.806. PubMed PMID: 21478889; PubMed Central PMCID: PMC3083463.
3. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010;20(9):1297-303. Epub 2010/07/21. doi: gr.107524.110 [pii] 10.1101/gr.107524.110. PubMed PMID: 20644199; PubMed Central PMCID: PMC2928508.
4. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID: PMC2723002.
5. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:12073907. 2012.
6. Lott MT, Leipzig JN, Derbeneva O, Xie HM, Chalkia D, Sarmady M, et al. mtDNA Variation and Analysis Using Mitomap and Mitomaster. Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]. 2013;44:1 23 1-6. doi: 10.1002/0471250953.bi0123s44. PubMed PMID: 25489354; PubMed Central PMCID: PMC4257604.
7. Navarro-Gomez D, Leipzig J, Shen L, Lott M, Stassen AP, Wallace DC, et al. Phy-Mer: a novel alignment-free and reference-independent mitochondrial haplogroup classifier. Bioinformatics. 2015;31(8):1310-2. doi: 10.1093/bioinformatics/btu825. PubMed PMID: 25505086; PubMed Central PMCID: PMC4393525.

About the Authors

This document was prepared by Dr. Perry Ridge, Brigham Young University, Department of Biology. For more information please contact Drs. Perry Ridge or John "Keoni" Kauwe by email at perry.ridge@byu.edu or kauwe@byu.edu, respectively.